Analyzing XCMG and JPM Extremes

By: Mitch Verhelle (mbv27)

Project Description

The methods used in this project were learned from Cornell University's ORIE 4565 Extreme Values in Finance course, taught by Professor Gennady Samorodnitsky in Spring 2024.

For the project, I select two stocks from Yahoo Finance, XCMG and JPM. I then analyze the extreme value distribution shape parameter of these returns to provide insights into extreme value modeling for these returns. This allows me to draw some suggestions about how to approach hypothetically investing in either company.

The returns are taken as daily returns. Code is written in R, and the data-scraping automatically computes the daily returns from adjusted closing prices. Sometimes log returns are used, so all code is specified at the end of each section for transparency.

Contents

1 Company Selection

I chose to analyze **XCMG** (Xuzhou Construction Machinery Group Co., Ltd.) and **JPM** (JPMorgan Chase & Co.) returns. XCMG is a construction equipment manufacturing company based in China, operating in the **Heavy Equipment** industry. JPM is a finance company headquartered in New York City, operating in the **Financial Services** industry.

I selected XCMG because a company that I interned for gets most of their revenue from financing equipment with XCMG.

I chose JPM because I am interested in working in the financial services industry, and from what I've observed in the news, JPM consistently performs at the top of its competition.

I expect to see differences in the extreme value distribution for both stocks as they are a part of very different industries.

1.1 Closing Prices Data

I programmed everything in R Studio. This uses R programming language with various packages.

To set up the keys to fetch data from Yahoo Finance from last 5 years, I wrote:

```r
library(quantmod)
library(lubridate)

end_date <- as.Date("2024-04-30")
start_date <- as.Date(end_date) - years(5)

# Yahoo Finance
getSymbols(Symbols = c("000425.SZ", "JPM"), from = start_date,
        to = end_date, auto.assign = TRUE)
```

Listing 1: Setup

1.2 Daily Returns

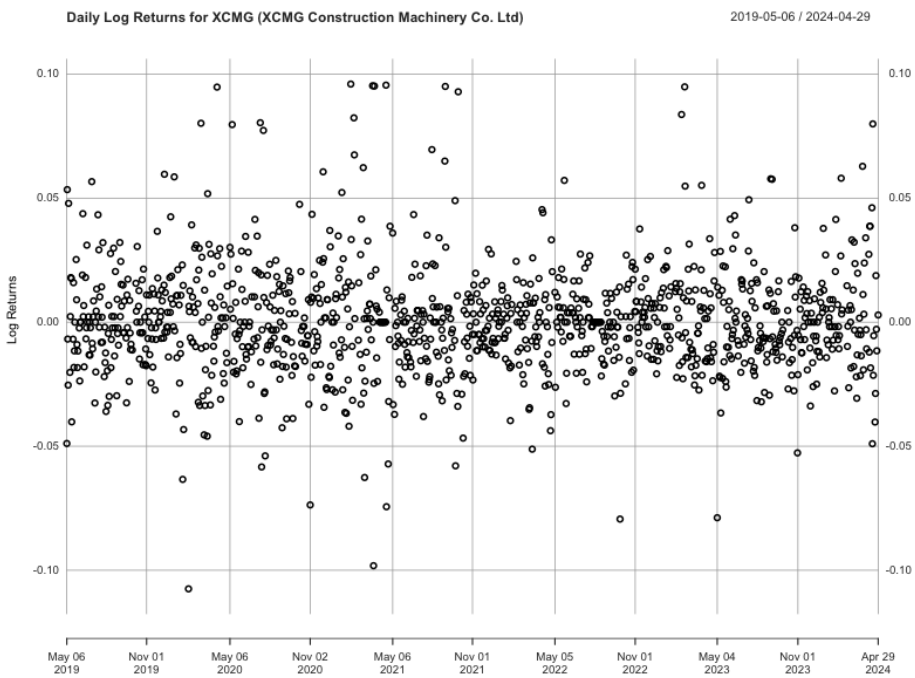

Figure 1: Daily Log Returns for XCMG

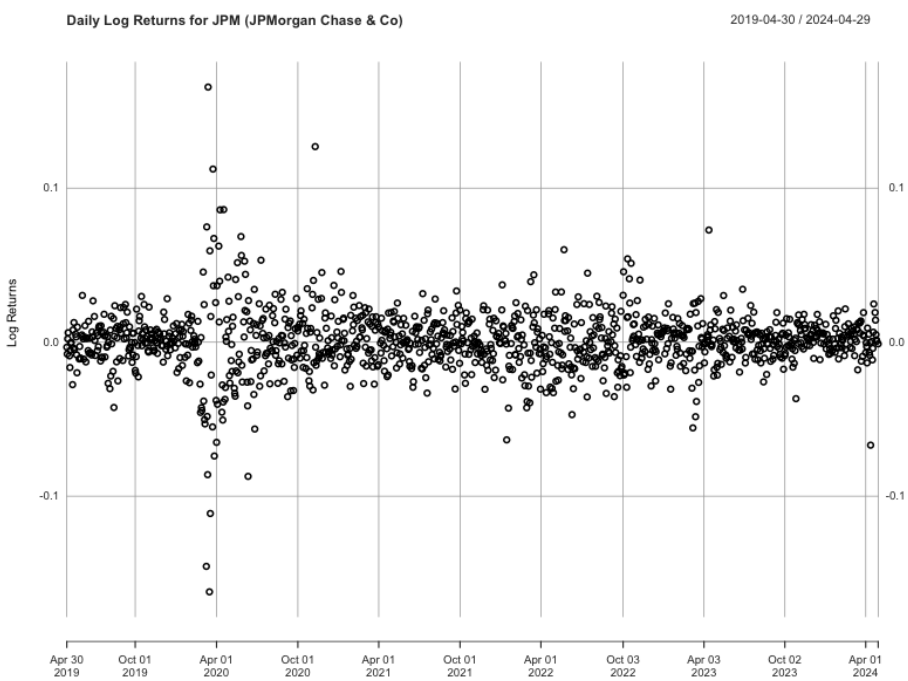

Figure 2: Daily Log Returns for JPM

XCMG Extremes appear uncorrelated and frequent.

JPM Extremes appear to spike during the incidence of COVID-19 and seem to be smaller than those of XCMG, possibly indicating inter-extreme dependence.
To plot these figures, I used the daily log-return data from past 5 years for both stocks:

```
XCMG_returns <- dailyReturn(`000425.SZ`, type='log')
JPM_returns <- dailyReturn(JPM, type='log')
```

Listing 2: Fetch Returns from Yahoo Finance

Some of these returns may be **NA** or $\pm\infty$. I create cleaned data whenever necessary, however for some cases, R will overlook these missing values.

The figures above show the plots of these daily log-returns for XCMG and JPM respectively.

2 Empirical Mean Excess Plots

Empirical Mean Excess plots are plots for various thresholds, u, such that we simulate plotting the mean excess of returns X. Given order statistics $X_{(i,n)}$, the empirical mean excess plot is obtained by plotting thresholds $u = X_{(k,n)}$ by

$$\mathbb{E}\left[X - u \mid X > u\right] \approx \frac{1}{k}\sum_{i=1}^{k-1} X_{(i,n)} - X_{(k,n)}$$

Empirical Mean Excess Equation [2, 4]

If we assume the extreme values under a centering and scaling converge in distribution to some Y, one can rewrite this Y in terms of a rescaled Generalized Pareto Distribution (GPD). In doing so, we observe that the mean excess should be roughly linear when it converges to this distribution.

In the following, I identify parts of the plot where it starts to look linear. I identified these points by experimenting with changing the bounds of the mean excess plots to ensure it was as linear as possible.

It's important to find the first point where the plot appears to become linear because this will allow us to use as many order statistics as possible when deriving the estimates for the shape parameter ξ in Section 3.

2.1 XCMG

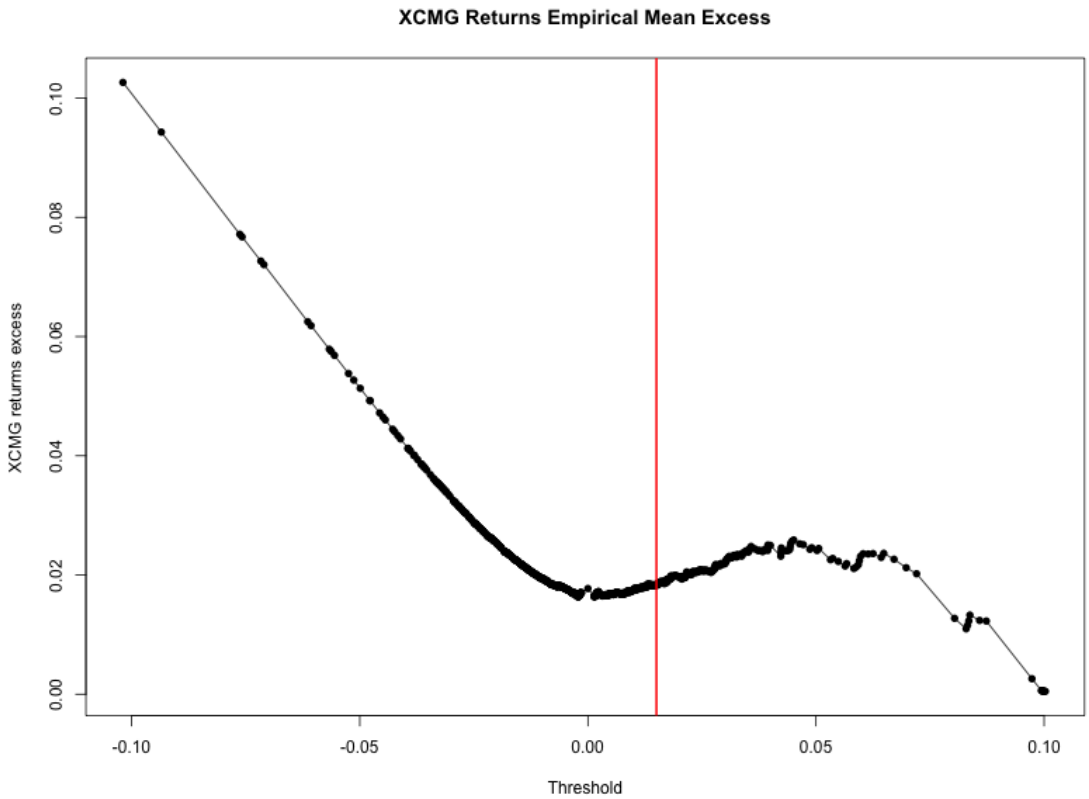

Figure 3: XCMG Empirical Mean Excess Plot

At the threshold value $u = 0.01502507$, we observe a roughly linear relationship in the empirical excess plot for excess values $X - u \mid X > u$.

The excess values greater than u appear to slope upward before becoming highly volatile. The volatile parts aside, this upward trend indicates that XCMG returns are heavy tailed.

Note, the above is all returns from XCMG. Since it's customary to analyze gains separately from returns, I do that in the following sub-subsections.

2.1.1 XCMG Gains

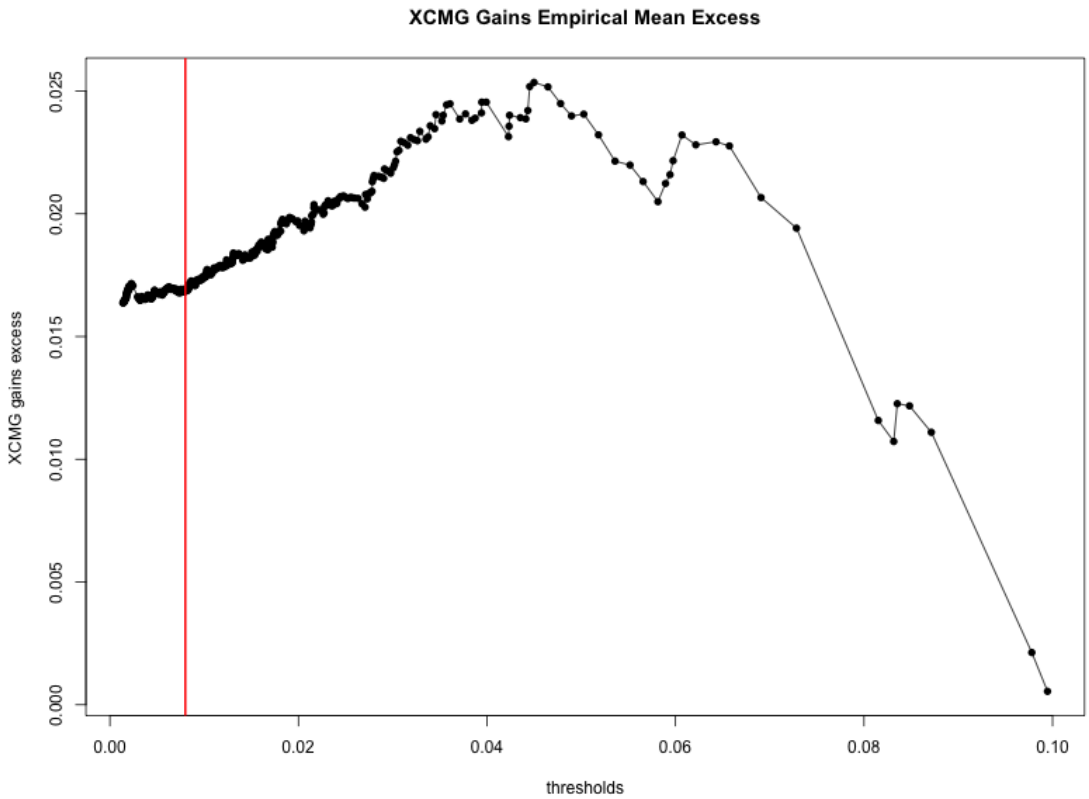

Figure 4: XCMG Gains Empirical Mean Excess Plot

The XCMG gains appear to linearly increase at the threshold $u = 0.008000031$.

the upward trend indicates that the XCMG gains are heavy-tailed.

We use this threshold later when calculating estimates for shape parameter ξ.

2.1.2 XCMG Losses

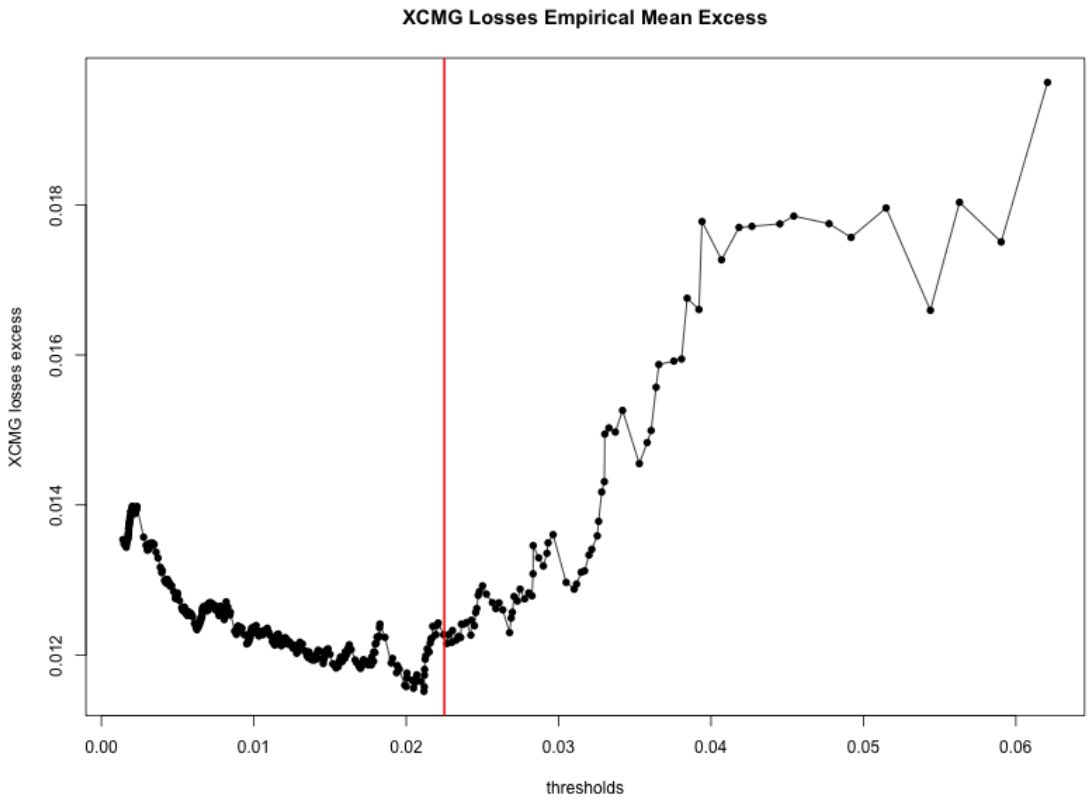

Figure 5: XCMG Losses Empirical Mean Excess Plot

XCMG losses appear to linearly decrease and then increase. They increase linearly starting at $u = 0.02254426$.

The trend indicates that XCMG losses are heavy-tailed after threshold u, which implies that this is where the extremes of the distribution are most stable.

We will use this threshold later when calculating estimates for shape parameter ξ, so it's important that it is my best guess as to when the excess returns start linearly increasing. Clearly, the excess returns become highly volatile beyond 0.032, and one could argue that they are volatile beyond 0.018. But the best guess for where the plot starts to become linearly increasing is around the red line at 0.0225, hence my choice of this threshold.

2.2 JPM

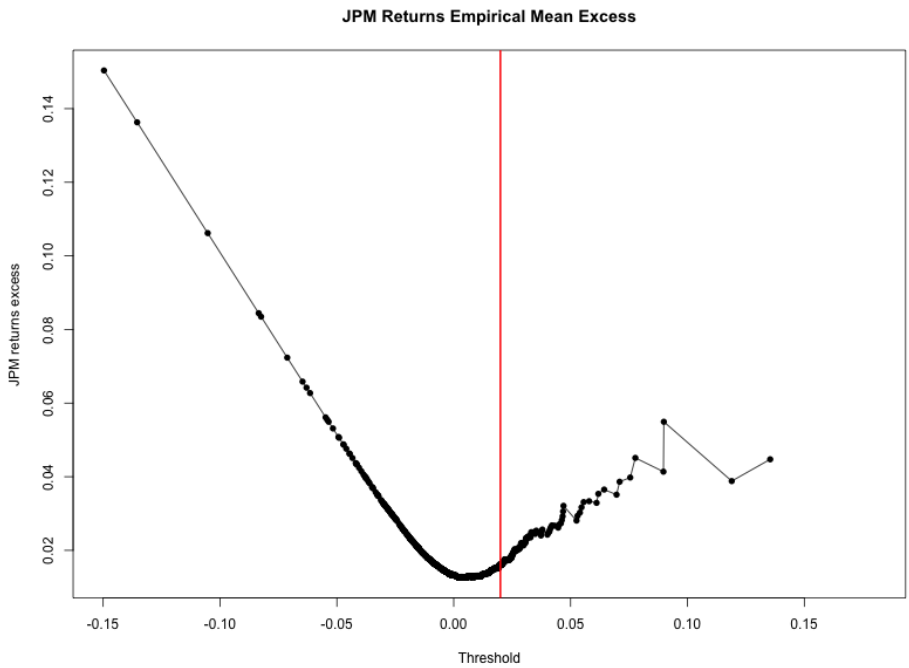

Figure 6: JPM Empirical Mean Excess Plot

At the threshold value of $u = 0.01995859$, the excess values for all JPM returns become roughly linear.

The JPM return excess values appear to slope upwards, linearly, indicating heavier tails. Moreover, they slope up faster than those in the XCMG return empirical mean excess plot, suggesting JPM returns have heavier tails than XCMG returns.

We next look at the gains separately from losses.

2.2.1 JPM Gains

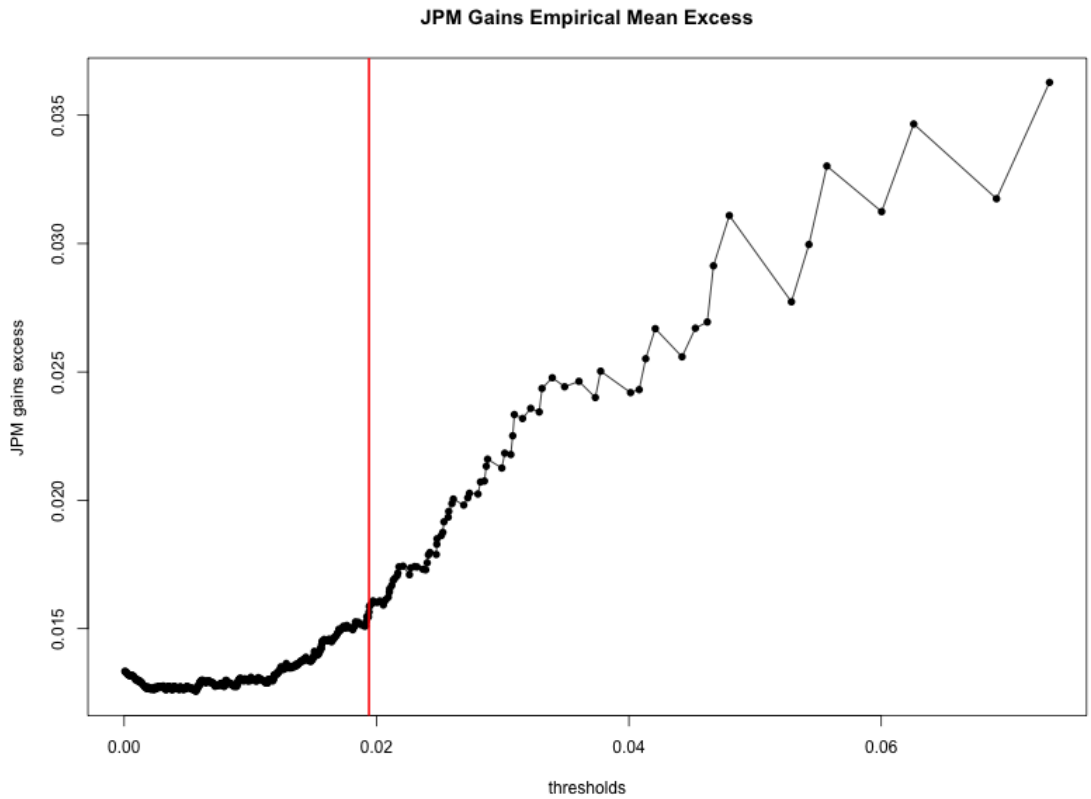

Figure 7: JPM Gains Empirical Mean Excess Plot

At $u = 0.01939756$, the JPM gains linearly increase.

This indicates that the returns are heavy tailed, and that we can use this threshold of where the returns become roughly linear as our threshold in later methods of approximating shape parameter ξ.

2.2.2 JPM Losses

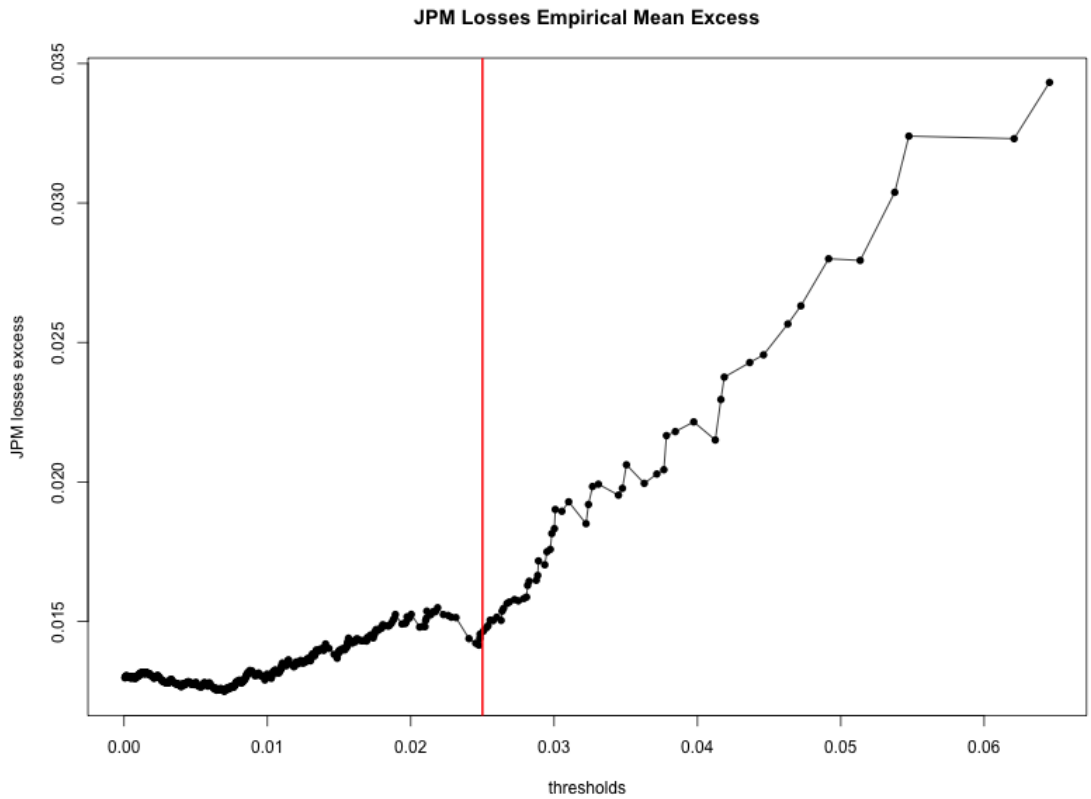

Figure 8: JPM Losses Empirical Mean Excess Plot

At $u = 0.02501967$, the JPM losses linearly increase, slightly less so than the gains.

This reveals the heavy tail nature of JPM losses. It's important to note the volatility involved. One could argue that the extremes start earlier than the threshold I selected, however, when I replotted this graph, zoomed in for smaller thresholds, I found that the threshold I provided best encapsulates where the extremes become roughly linear. In other places, they seem to be concave up.

This is the threshold I will use in later attempts to approximate shape parameter ξ for JPM losses.

2.3 Code

```r
## Setup:
library(quantmod)
library(lubridate)
getSymbols(Symbols = c("000425.SZ", "JPM"), from = as.Date(today()) - years(5), to = today(),
           auto.assign = TRUE)
XCMG_returns <- dailyReturn(`000425.SZ`)
JPM_returns <- dailyReturn(JPM)
# Split to gains and losses
XCMG_gains <- na.omit(ifelse(XCMG_returns > 0, XCMG_returns, NA))
XCMG_losses <- na.omit(ifelse(XCMG_returns < 0, abs(XCMG_returns), NA))
JPM_gains <- na.omit(ifelse(JPM_returns > 0, JPM_returns, NA))
JPM_losses <- na.omit(ifelse(JPM_returns < 0, abs(JPM_returns), NA))
rm(list=setdiff(ls(), c("XCMG_gains", "XCMG_losses", "JPM_gains", "JPM_losses")))

## Empirical Mean Excess plots
library(ismev)

mean_excess <- function(data, threshold) {
  excess <- data[data > threshold] - threshold
  mean(excess)
}

eme_plot <- function(returns, y_value=0, y_lbl="", plot_title="", file_name="NA.png") {
  thresholds <- sort(unique(returns))
  mean_excess_values <- sapply(thresholds, mean_excess, data=returns)
  closest_index <- which.min(abs(thresholds - y_value))
  print(thresholds[closest_index])
  png(file.path("~/Project", file_name), width=800, height=600)
  plot(thresholds, mean_excess_values, type='o', col='black', pch=16,
       main=plot_title, xlab="Threshold", ylab=y_lbl)
  abline(v = thresholds[closest_index], col = "red", lwd = 2)  # Vertical line at the threshold
  dev.off()
}

eme_plot(XCMG_returns,
         y_value=0.015, #From Mean Excess plot
         y_lbl="XCMG returms excess",
         plot_title="XCMG Returns Empirical Mean Excess",
```

```
         file_name="XCMG_eme.png")
eme_plot(JPM_returns,
         y_value=0.02, #From Mean Excess plot
         y_lbl="JPM returns excess",
         plot_title="JPM Returns Empirical Mean Excess",
         file_name="JPM_eme.png")
eme_plot(XCMG_gains,
         y_value=0.008, #From Mean Excess plot
         y_lbl="XCMG gains excess",
         plot_title="XCMG Gains Empirical Mean Excess",
         file_name="XCMG_gains_eme.png")
eme_plot(XCMG_losses,
         y_value=0.0225, #From Mean Excess plot
         y_lbl="XCMG losses excess",
         plot_title="XCMG Losses Empirical Mean Excess",
         file_name="XCMG_losses_eme.png")
eme_plot(JPM_gains,
         y_value=0.0194, #From Mean Excess plot
         y_lbl="JPM gains excess",
         plot_title="JPM Gains Empirical Mean Excess",
         file_name="JPM_gains_eme.png")
eme_plot(JPM_losses,
         y_value=0.025, #From Mean Excess plot
         y_lbl="JPM losses excess",
         plot_title="JPM Losses Empirical Mean Excess",
         file_name="JPM_losses_eme.png")
```

Listing 3: Empirical Mean Excess Plots

3 Shape Parameter Estimation

In order to get a sense of how large these extremes may be, we need to estimate the shape parameter of the respective extreme value distributions.

I employ three different methods to estimate the shape parameters of the extreme value distributions for XCMG and JPM returns: Hill's Estimator, Pickand's Estimator, and the Probability Weighted Moments method for estimating the parameters of the generalized extreme value distribution.

The parameter estimates are performed for shape parameter ξ for each stock's returns split by gains and losses as well as combined. This is for more robust parameter selection, as we expect one of the parameters of the gains and losses to be close to that of all returns. This quality should reassure us that the method is a good estimate for ξ.
The methods for computing each are mostly applied through R packages, as mentioned in the Code sections after each subsection.

3.1 Hill Estimator

$$\hat{\xi}_{k,n}^{(H)} = \frac{1}{k}\sum_{j=1}^{k} \log X_{(j,n)} - \log X_{(k,n}$$

Hill Estimator Equation [2, 4]

The Hill Estimator, implemented in the R package, evir, incorporates k large log-order statistics to estimate ξ, the shape parameter of the distribution of the financial return extremes.

This method is useful because it incorporates multiple observations and allows us to visualize the shape parameter. **We expect that as we vary the threshold, the shape parameter (determined on the y-axis) will converge towards the true shape parameter**.

We must be wary of using too many order statistics, or too few order statistics, as we prefer this method to analyze enough extreme values to output a reasonable guess for ξ.

One limitation as we will soon see in JPM's hill estimator is that this method works only when the shape parameter ξ is positive. We resort to the other methods of shape parameter estimation in the case that ξ may be 0 or negative.

3.1.1 XCMG

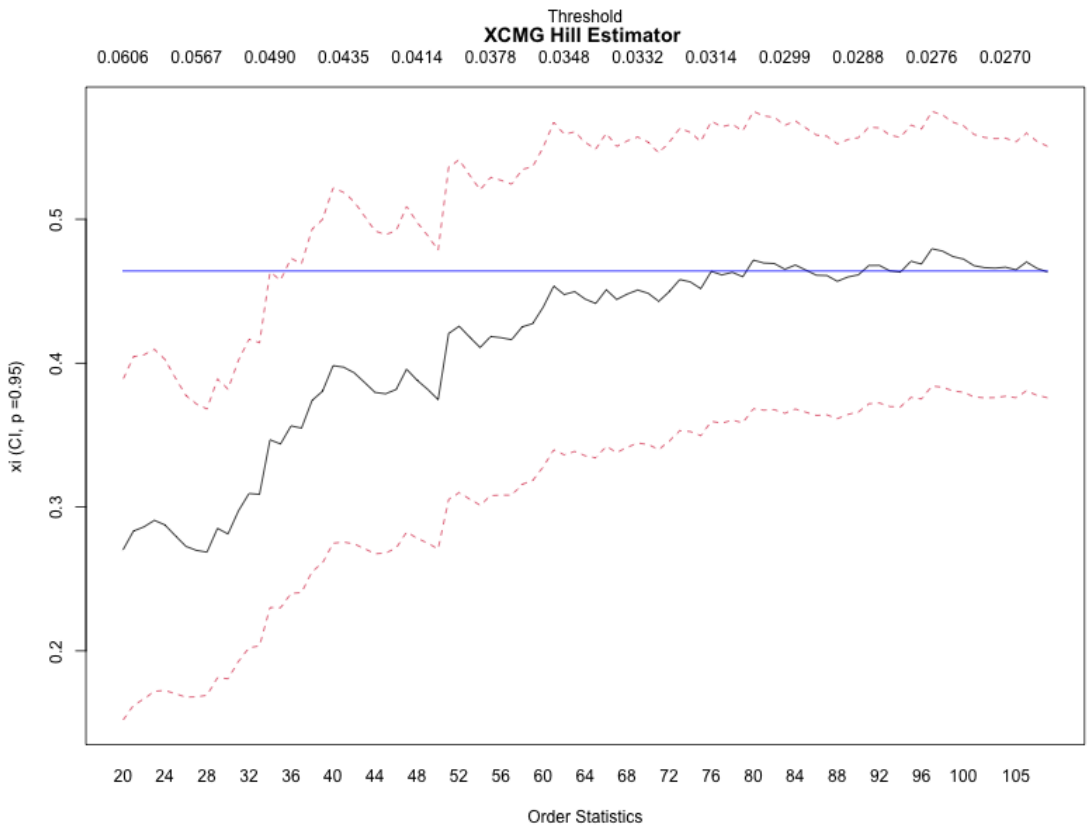

Figure 9: XCMG Hill Estimator

The Hill estimator appears to converge to 0.464

$$\hat{\xi}_{k,n}^{(H)} \approx 0.464$$

The plot seems to hover around $\hat{\xi}_{k,n}^{(H)} \in (.52, .38)$ for most previous thresholds, indicating that the Hill estimator may be a good fit for the shape parameter of XCMG extreme returns. We expect either the gains or losses to be contributing more to this estimate. We analyze this next.

3.1.2 Gains

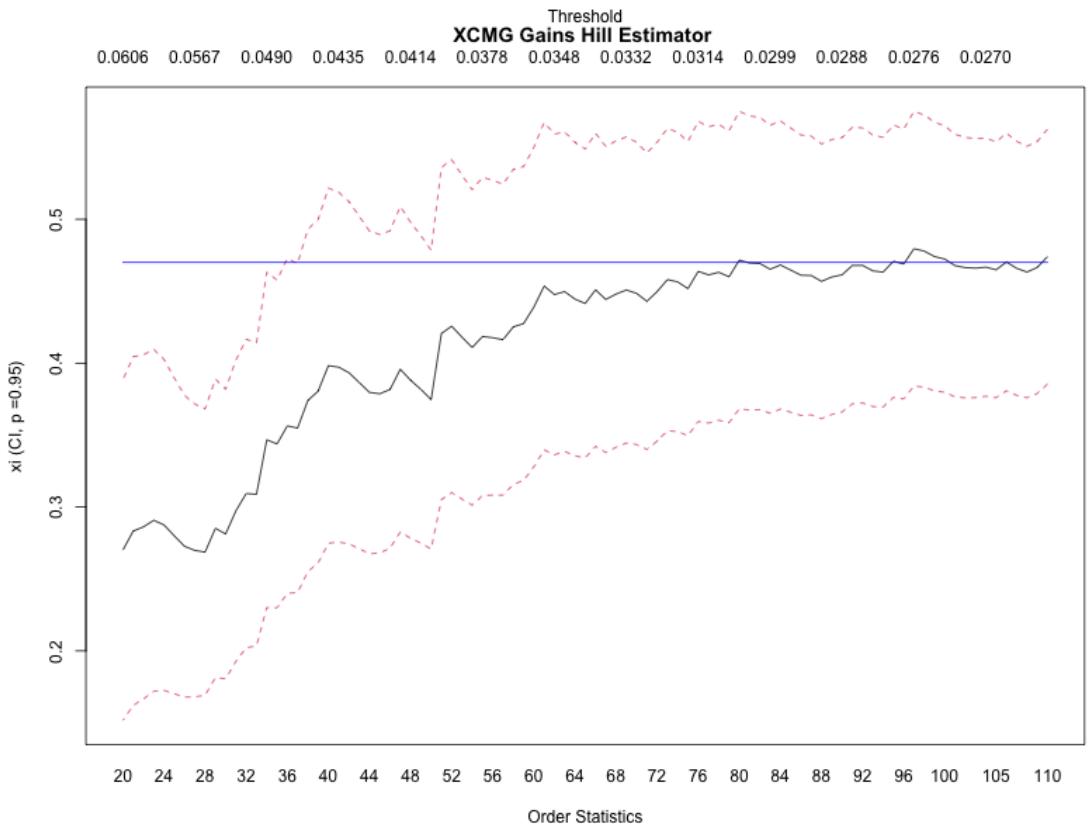

Figure 10: XCMG Gains Hill Estimator

The XCMG gains hill estimate is about 0.47.
This is close to the shape parameter of the overall XCMG returns. We expect that the gains will be either similar in value or less than 0.464, the estimate of the overall returns. This expectation is upheld, as 0.47 is very close to 0.464.

3.1.3 Losses

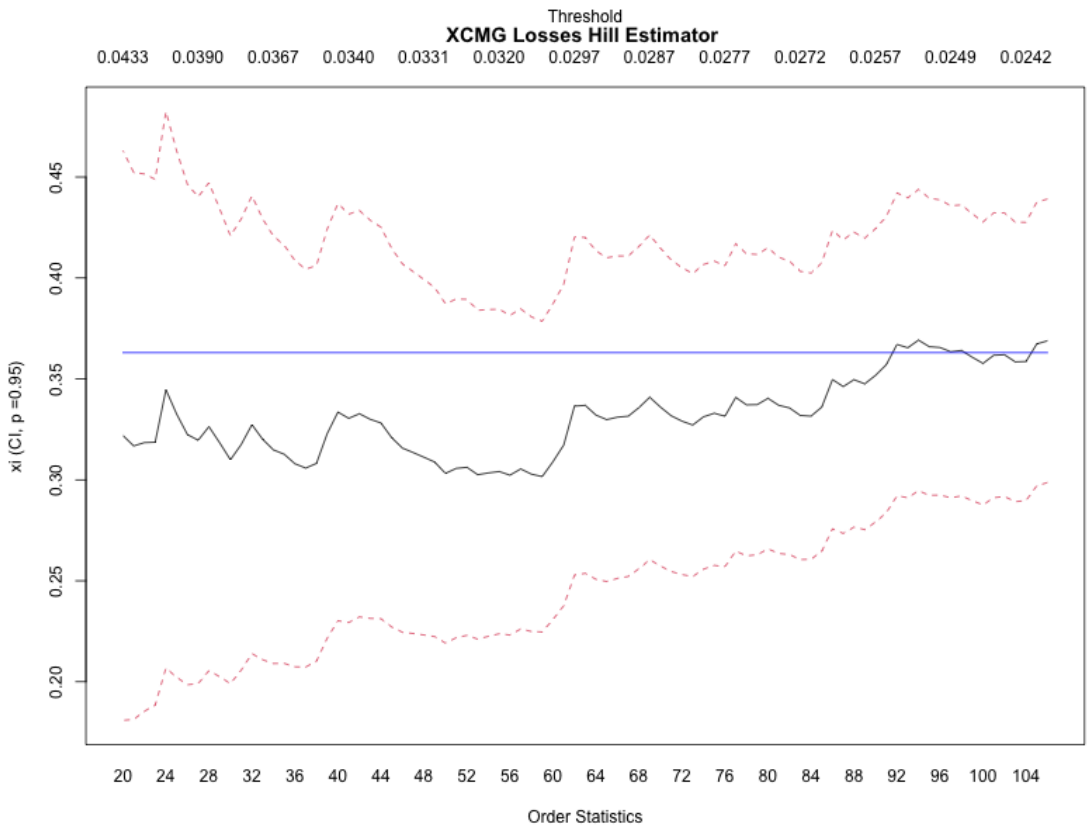

Figure 11: XCMG Losses Hill Estimator

The XCMG losses hill estimate is about 0.363.
The shape parameter of 0.363 is quite a bit smaller than that of the gains and the overall returns. This means the shape parameter of the overall distribution of returns for XCMG is governed by the extremes of XCMG gains, not losses.

3.1.4 JPM

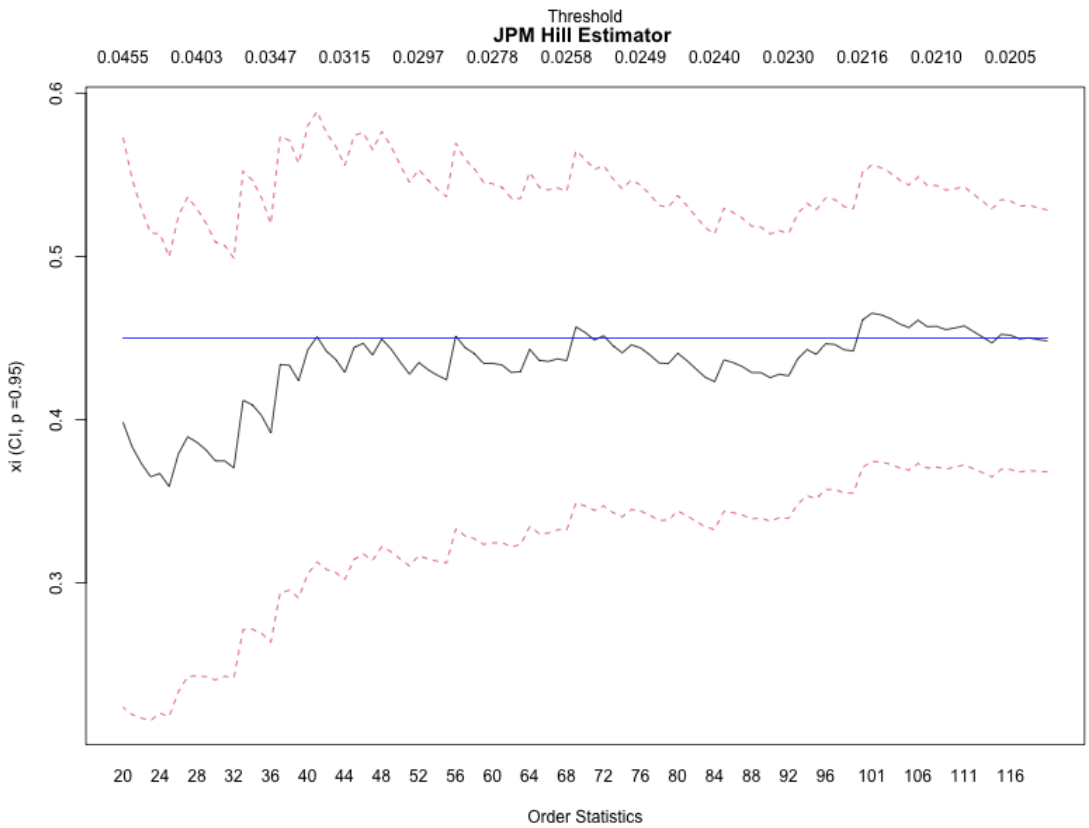

Figure 12: JPM Hill Estimator

The JPM Hill estimator looks like it approaches 0.45.

$$\hat{\xi}_{k,n}^{(H)} \approx 0.45$$

This indicates that the absolute returns have an estimated shape parameter of 0.45. We expect either the gains or losses to be contributing more to this estimate. We analyze this next.

3.1.5 Gains

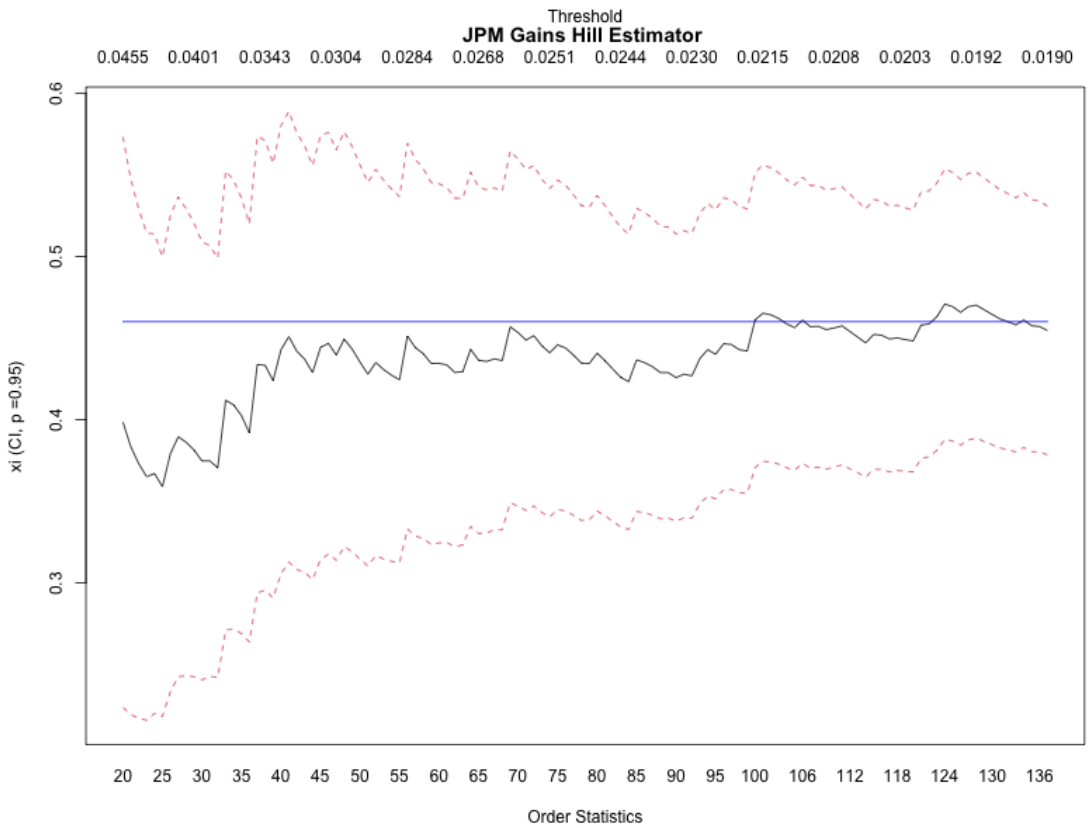

Figure 13: JPM Gains Hill Estimator

The JPM gains hill estimate is about 0.46.
This is close to the shape parameter of the overall JPM returns. This is in line with what we might expect. Either the gains or losses should govern the shape parameter of the overall returns. In this case, I suspect it will be the gains that govern the shape parameter estimate of 0.45 of JPM returns.

3.1.6 Losses

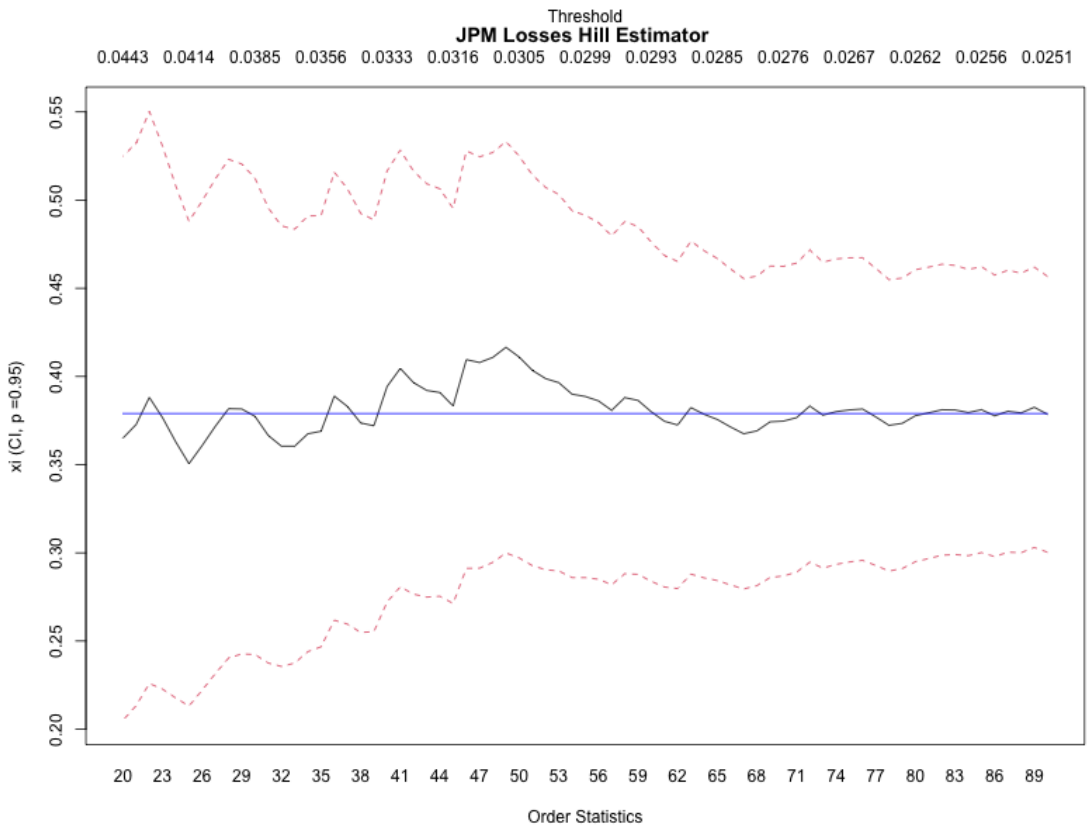

Figure 14: JPM Losses Hill Estimator

The JPM losses hill estimate is about 0.379.
The shape parameter of 0.379 is quite a bit smaller than that of the gains and the overall returns. Similar to before, this means that the shape parameter of the overall distribution of returns for JPM is governed by the extremes of JPM gains, not the losses.

3.1.7 Code

```r
## Setup:
library(quantmod)
library(lubridate)
getSymbols(Symbols = c("000425.SZ", "JPM"), from = as.Date(today()) - years(5), to = today(), auto.assign = TRUE)
XCMG_returns <- dailyReturn(`000425.SZ`, type='log')
JPM_returns <- dailyReturn(JPM, type='log')
# Split into cleaned gains and losses
clean_returns <- function(returns) {
    clean_returns=as.numeric(returns[!is.na(returns) & !is.infinite(returns)])
}
XCMG_gains <- clean_returns(na.omit(ifelse(XCMG_returns > 0, XCMG_returns, NA)))
XCMG_losses <- clean_returns(na.omit(ifelse(XCMG_returns < 0, abs(XCMG_returns), NA)))
JPM_gains <- clean_returns(na.omit(ifelse(JPM_returns > 0, JPM_returns, NA)))
JPM_losses <- clean_returns(na.omit(ifelse(JPM_returns < 0, abs(JPM_returns), NA)))
XCMG_returns <- clean_returns(na.omit(XCMG_returns))
JPM_returns <- clean_returns(na.omit(JPM_returns))
## Hill Estimator:
library(evir)
hill_plot <- function(returns, start=20, end=320, main="ttl", u=0) {
  plot <- hill(returns, option='xi', start=start, end=end, main=paste(main,"\n"))
  segments(x0=start, y0=u, x1=end, y1=u, col="blue")
}
hill_plot(XCMG_returns, start=20, end=108, main="XCMG Hill Estimator", u=0.464)
hill_plot(JPM_returns, start=20, end=120, main="JPM Hill Estimator", u=0.45)
hill_plot(XCMG_gains, start=20, end=110, main="XCMG Gains Hill Estimator", u=0.47)
hill_plot(XCMG_losses, start=20, end=106, main="XCMG Losses Hill Estimator", u=0.363)
hill_plot(JPM_gains, start=20, end=137, main="JPM Gains Hill Estimator", u=0.46)
hill_plot(JPM_losses, start=20, end=90, main="JPM Losses Hill Estimator", u=0.379)
# Omitted: Saving PNG Files
```

Listing 4: Hill Estimator

3.1.8 Hill Estimators: Analysis

After deriving the above estimates for ξ, we observe that they fall in line with what theory would suggest. Particularly, the estimates for XCMG and JPM returns (all returns) matches one of their respective gains/losses shape estimates. The other remaining estimate is less than both other estimates. This falls in line with our expectation that the shape parameter of all returns is typically governed by one of the shape parameters between the gains and losses, since all returns use the extremes from both gains and losses.

It's important to continue providing estimates for ξ, since the hill estimator is just a best guess. After all, extremes are volatile and rare, hence they are extremes. It would be naive to select only one estimator and call it quits.

3.2 Pickand's Estimator

The Pickands Estimator is given as follows:

$$\hat{\xi}_{k,n}^{(P)} = \log_2 \left(\frac{X_{(k,n)} - X_{(2k,n)}}{X_{(2k,n)} - X_{(4k,n)}} \right)$$

$$\text{as } n \to \infty,\ k = k_n \to \infty, \text{ and } \frac{k_n}{n} \to 0.$$

Pickand's Estimator Equation [2, 4]

Pickand's estimator uses 3 order statistics to estimate ξ. If the above holds, then $\hat{\xi}_{k,n}^{(P)} \to \xi$ in probability.

We use the Pickand's Estimator at various thresholds to get a better idea of the shape parameter ξ for XCMG and JPM extreme returns. Hopefully it will validate our results from the Hill Estimator in the previous section.

In the following, I show the plots and give a brief description of the estimates for each of the returns. Interpretations for this section are included at the end of this section. **The thresholds selected are motivated from those found in the Empirical Mean Excess plots, however we are primarily varying the thresholds to find where the plot stabilizes**.

3.2.1 XCMG

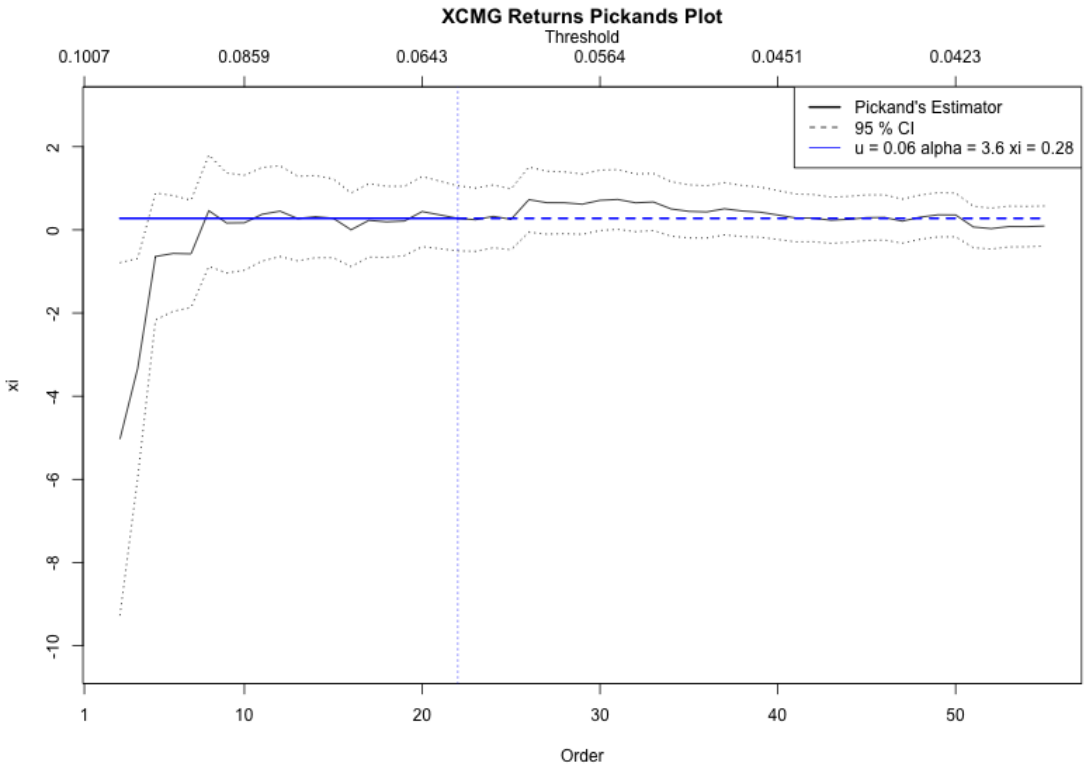

Figure 15: XCMG Pickands Estimator

Pickand's estimator estimates $\xi^{(P)} = 0.28$ for all XCMG returns.

3.2.2 Gains

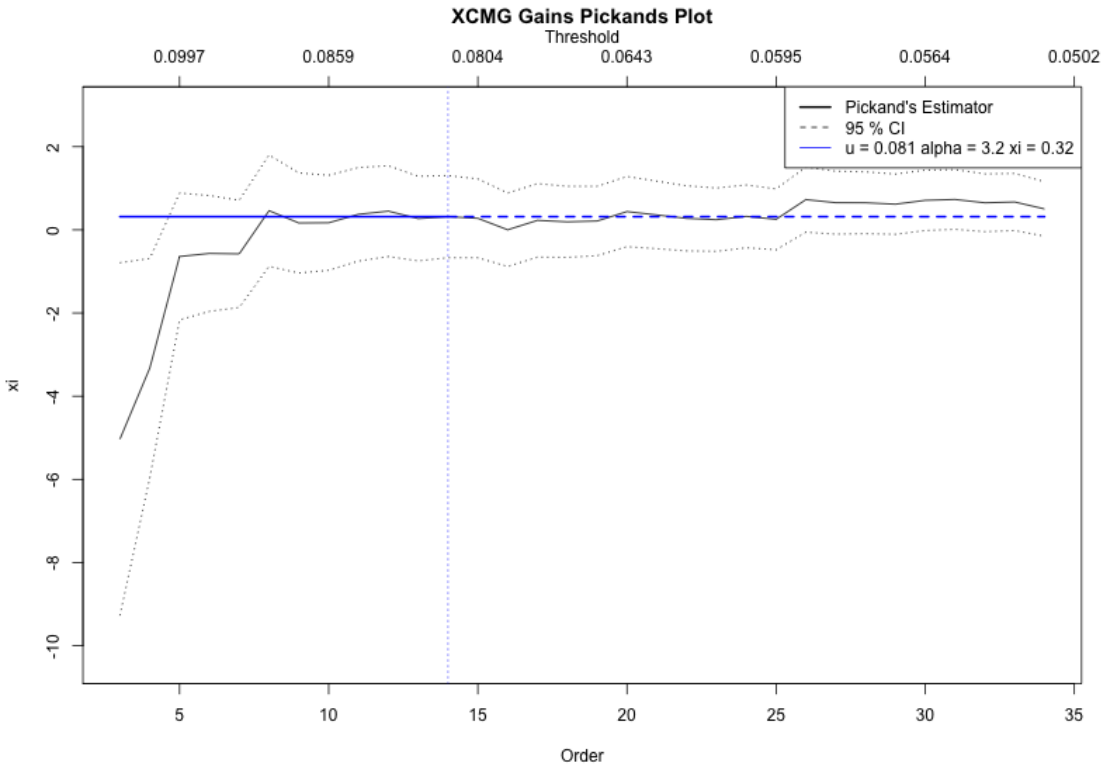

Figure 16: XCMG Gains Pickands Estimator

Pickand's estimator estimates $\xi^{(P)} = 0.32$ for XCMG gains.

3.2.3 Losses

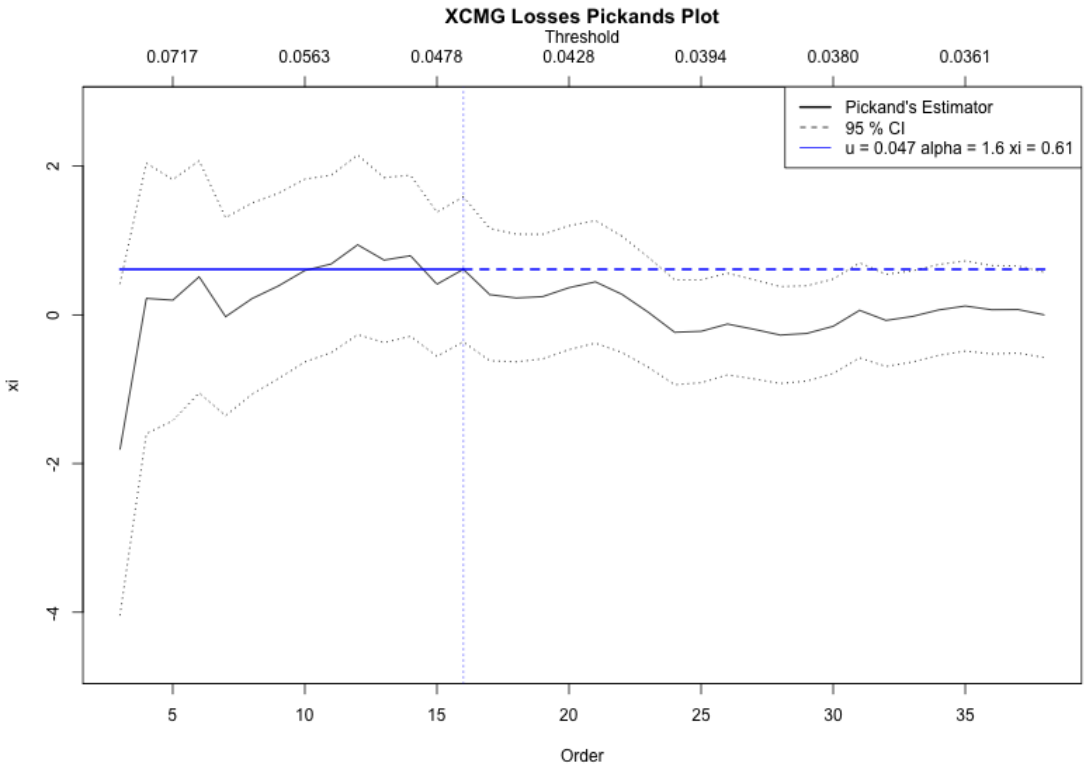

Figure 17: XCMG Losses Pickands Estimator

Pickand's estimator estimates $\xi^{(P)} = 0.61$ for XCMG losses.

The plot appears to have some volatility, but not too unstable.

3.2.4 JPM

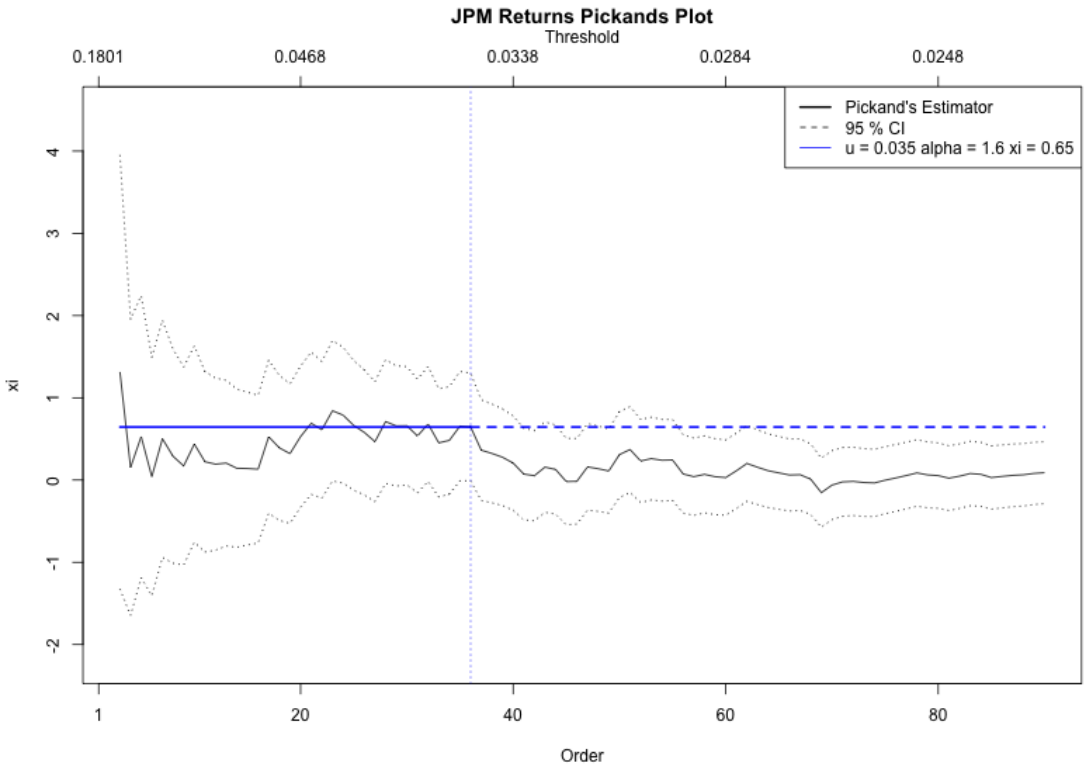

Figure 18: JPM Pickands Estimator

Pickand's estimator estimates $\xi^{(P)} = 0.65$ for all JPM returns.
This plot appears to approach the estimate, just before dropping to a lower estimate for slightly smaller thresholds.

3.2.5 Gains

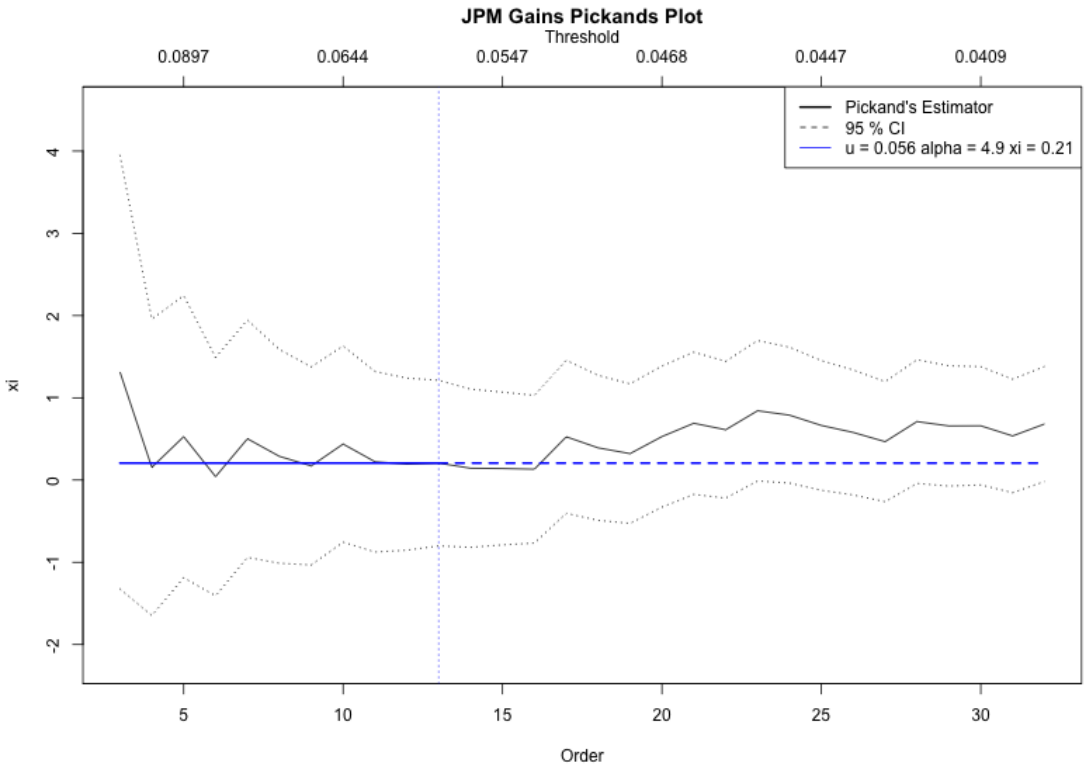

Figure 19: JPM Gains Pickands Estimator

Pickand's estimator estimates $\xi^{(P)} = 0.21$ for all JPM returns.

This plot shows a relatively consistent estimate, possibly slightly underestimating as larger and smaller thresholds are slightly above the horizontal blue line.

3.2.6 Losses

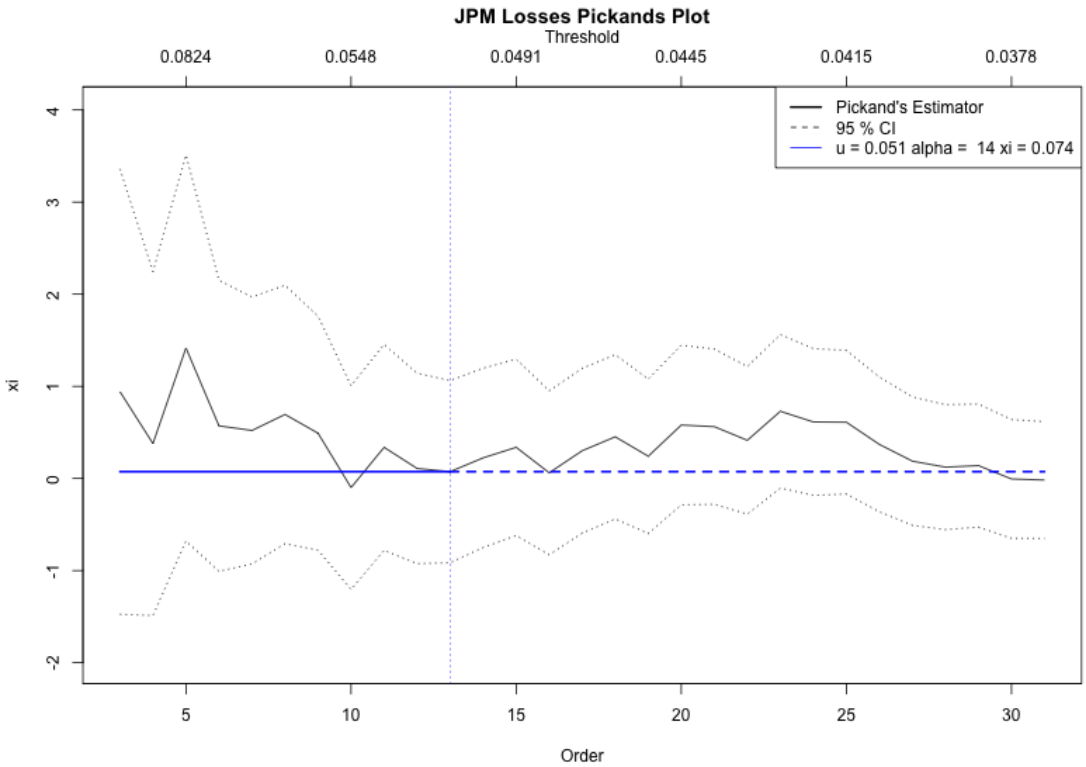

Figure 20: JPM Losses Pickands Estimator

Pickand's estimator estimates $\xi^{(P)} = 0.074$ for all JPM returns.

This estimate is possible underestimating ξ. We notice that most of the possible thresholds around our selected threshold are above the chosen threshold. We will soon compare this estimate to that of the Hill's estimator to make a better decision about which to use, and to inform the possible range of ξ.

3.2.7 Code

```
## SETUP:
library(quantmod)
library(lubridate)
getSymbols(Symbols = c("000425.SZ", "JPM"), from = as.Date(today())
    - years(5), to = today(), auto.assign = TRUE)
XCMG_returns <- dailyReturn(`000425.SZ`)
JPM_returns <- dailyReturn(JPM)
clean_returns <- function(returns) {
  clean_returns=as.numeric(returns[!is.na(returns) & !is.infinite(
    returns)])
}
XCMG_gains <- clean_returns(ifelse(XCMG_returns > 0, XCMG_returns,
    NA))
XCMG_losses <- clean_returns(ifelse(XCMG_returns < 0, abs(XCMG_
    returns), NA))
JPM_gains <- clean_returns(ifelse(JPM_returns > 0, JPM_returns, NA))
JPM_losses <- clean_returns(ifelse(JPM_returns < 0, abs(JPM_returns)
    , NA))
XCMG_returns <- clean_returns(XCMG_returns)
JPM_returns <- clean_returns(JPM_returns)
## Pickand's Plots:
library(evmix)
plot_pickands <- function(returns, threshold, plot_title="") {
  excesses <- returns[returns > threshold]
  pickandsplot(excesses, main=plot_title, xlab="Order", ylab="xi", y
    .alpha=FALSE, legend.loc="topright")
}
plot_pickands(XCMG_returns, threshold=0.0150, plot_title="XCMG
    Returns Pickands Plot") #xi=0.28
plot_pickands(XCMG_gains, threshold=0.0225, plot_title="XCMG Gains
    Pickands Plot") #xi=0.32
plot_pickands(XCMG_losses, threshold=0.0200, plot_title="XCMG Losses
     Pickands Plot") #xi=0.61
plot_pickands(JPM_returns, threshold=0.0080, plot_title="JPM Returns
     Pickands Plot") #xi=0.65
plot_pickands(JPM_gains, threshold=0.0194, plot_title="JPM Gains
    Pickands Plot") #xi=0.21
plot_pickands(JPM_losses, threshold=0.0250, plot_title="JPM Losses
    Pickands Plot") #xi=0.49
#Omitted: Saving PNG Files
```

Listing 5: Pickands Estimator

3.3 Pickand's Estimator: Interpretation

The estimates found for XCMG and JPM returns are all positive. This is consistent with the findings from the Hill Estimator.

We observe that the estimates for ξ vary more than expected from the Hill estimates. Particularly, we expect that one of the estimates be close to that of all returns, since all returns are made up of gains and losses. However, the estimate for XCMG losses is much larger than the estimate of XCMG all returns. Similarly, the estimate for all JPM returns is significantly greater than that of JPM gains and losses.

These somewhat counterintuitive findings are possibly due to the way Pickand's estimator works. It only computes an estimate from three order statistics, making the estimate more volatile than Hill's Estimator. They are also possibly due to minor differences in opinion of how to exactly select the threshold from the mean excess plots. However, whichever choice one goes with, we can never find the best, exact solution for extreme value parameters or the thresholds because there are few extremes to work with. That is why more estimators are helpful in narrowing down our best guess of ξ.

We can at least conclude that the shape parameter is likely positive.

We will compute one more estimate for ξ to verify this claim of positivity and to provide even more robustness over what the best estimates for the shape parameter are for each of these returns.

3.4 Generalized Pareto Distribution: Method of Probability Weighted Moments

My third method for estimating ξ is to use the method of Probability Weighted Moments (PWM) to fit the shape parameter and scale parameters of the Generalized Pareto Distributions (GPD) of XCMG and JPM returns using the POT package.

Before conducting this analyses, I need to ensure the first moment of the returns is finite.[2,4] I plot the MaxSum Ratio plots for p=$\{1, 2, 3\}$ of the gains and losses for XCMG and JPM.

3.4.1 MaxSum Ratios

p=1

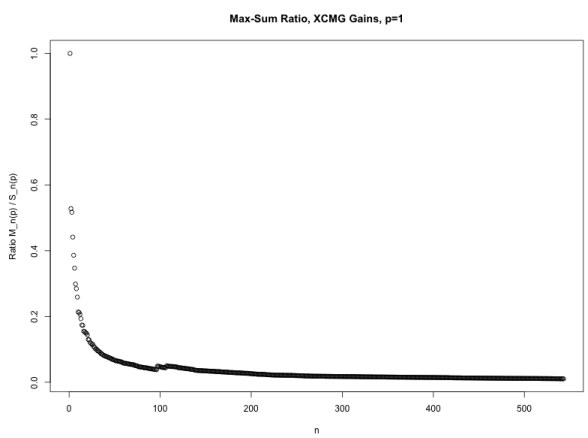

Figure 21: XCMG Gains Ratio: $p = 1$

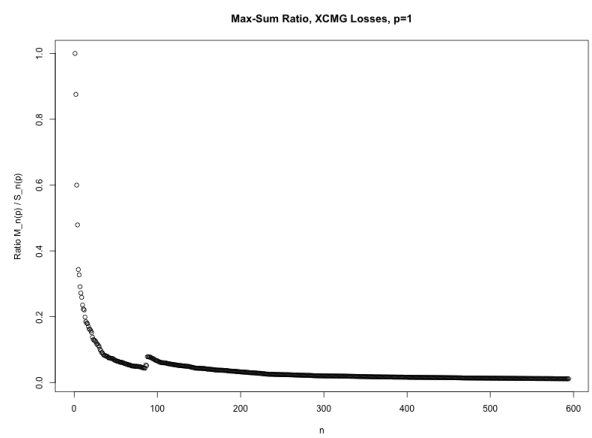

Figure 22: XCMG Losses Ratio: $p = 1$

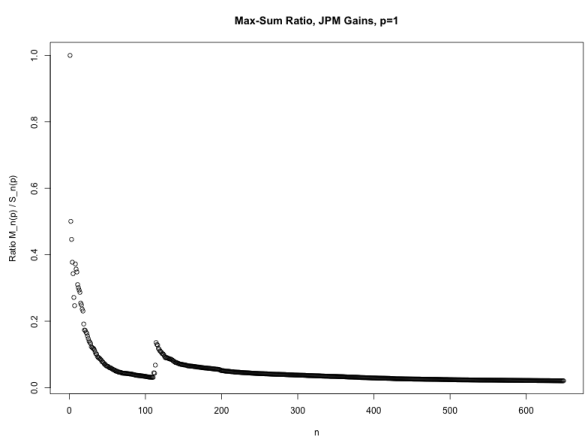

Figure 23: JPM Gains Ratio: $p = 1$

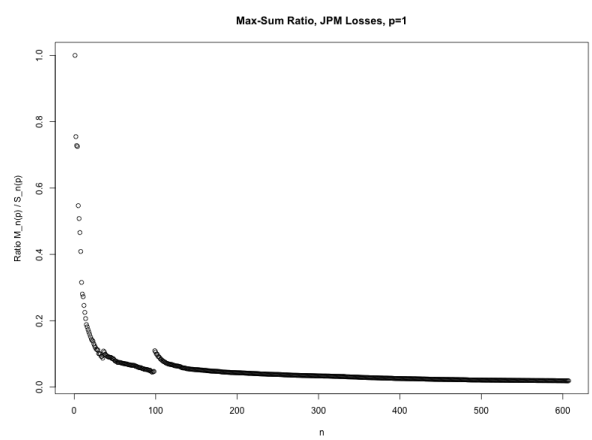

Figure 24: JPM Losses Ratio: $p = 1$

p=2

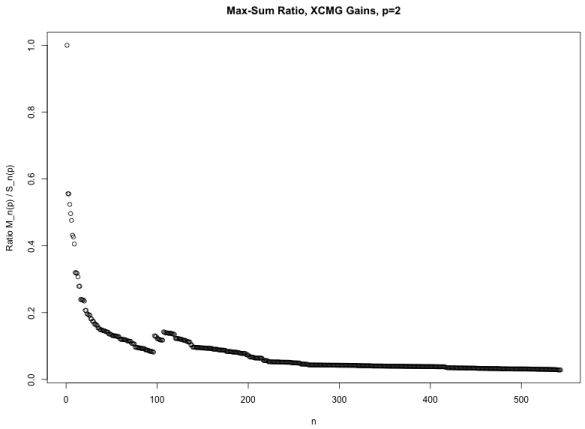

Figure 25: XCMG Gains Ratio: $p = 2$

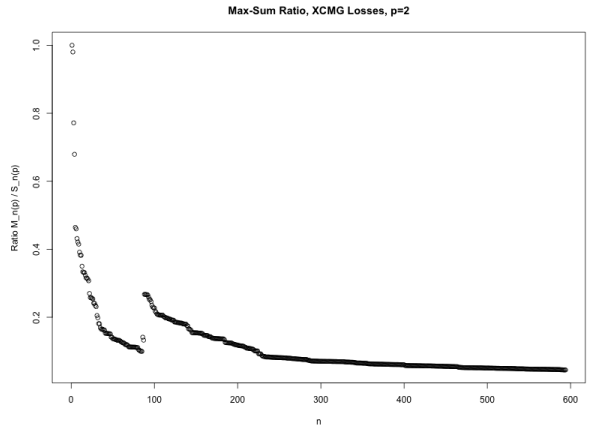

Figure 26: XCMG Losses Ratio: $p = 2$

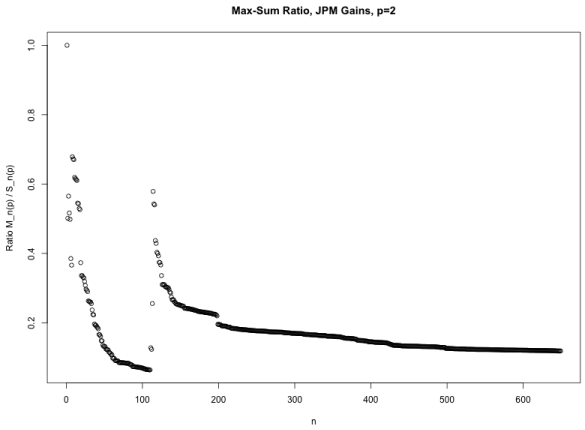

Figure 27: JPM Gains Ratio: $p = 2$

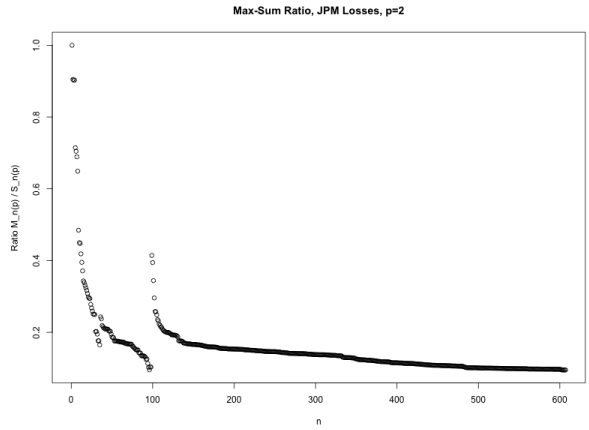

Figure 28: JPM Losses Ratio: $p = 2$

p=3

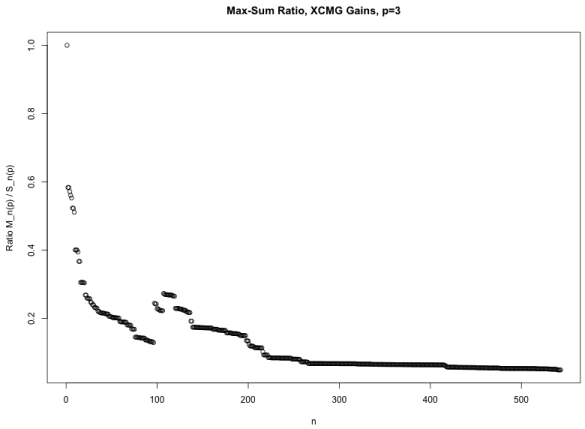

Figure 29: XCMG Gains Ratio: $p = 3$

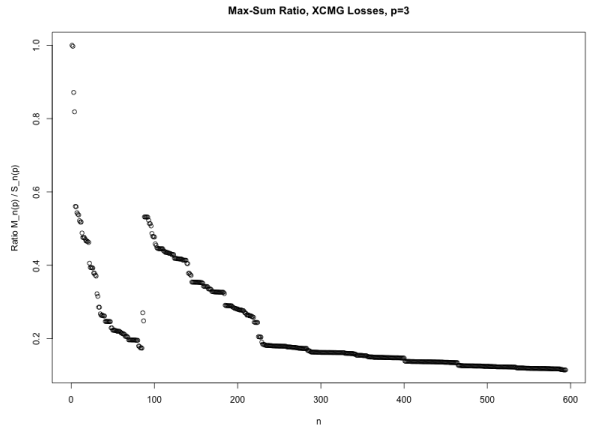

Figure 30: XCMG Losses Ratio: $p = 3$

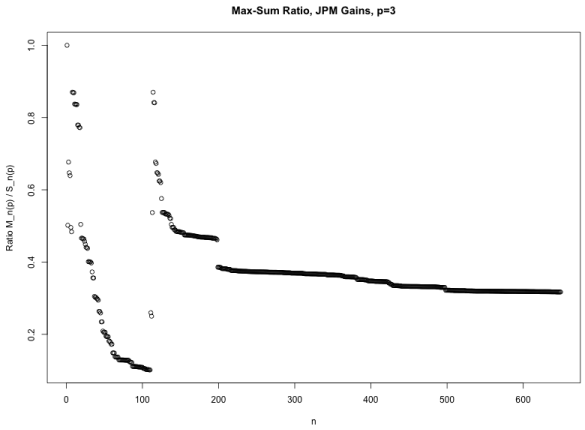

Figure 31: JPM Gains Ratio: $p = 3$

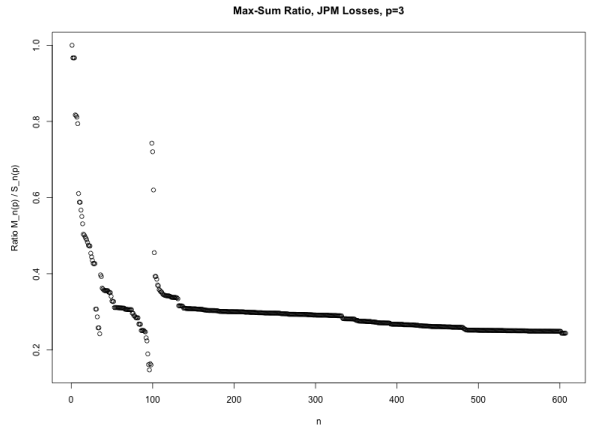

Figure 32: JPM Losses Ratio: $p = 3$

MaxSum Ratio Analysis

Daily Gain/Loss returns for XCMG and JPM are in RV(-p) for some $p > 0$, with p determined by the above plots. Most importantly, the first moments of XCMG and JPM gains and losses looks to converge to 0.
The second moments of everything except XCMG gains appears to not converge to 0. This means it's likely that $1 \leq p < 2$ for XCMG losses, JPM gains, and JPM losses.[2] However, to be sure, we plot the MaxSum Ratios for all gains/losses with $p = 3$.

For $p = 3$, all gains and losses plots appear to not approach 0. This implies that the parameter for XCMG gains is some $p \leq 2 < 3$.[2]

This means all first moments are finite, allowing us to continue with the Method of Probability Weighted Moments.[2],[4]

PWM Estimates

I now use the method of PWM to compute the shape and scale estimates for each GPD.

XCMG

All returns $\hat{\xi}_X^{(pwm)} = 0.15397963$, scale= 0.01513291.

Gains $\hat{\xi}_{X+}^{(pwm)} = 0.1649580$, scale= 0.0137954.

Losses $\hat{\xi}_{X-}^{(pwm)} = 0.14826463$, scale= 0.01084621.

JPM

All returns $\hat{\xi}_J^{(pwm)} = 0.33190280$, scale= 0.01023387.

Gains $\hat{\xi}_{J+}^{(pwm)} = 0.31312970$, scale= 0.01050621.

Losses $\hat{\xi}_{J-}^{(pwm)} = 0.40358411$, scale= 0.00881030.

All shape parameter estimates are positive, similar to the other estimations.

The estimates for ξ of XCMG returns are smaller in magnitude than those found from the Hill's estimator. This may due to the threshold. In Hill's method, the threshold is selected by identifying a point of convergence in the hill plot, whereas here the threshold is motivated directly by the Empirical Mean Excess Plots.

The estimates of gains and losses for each shape parameter are similar to the shape estimates of "all returns". This is a good sign, as it matches expectations mentioned before.

This test yielded robust estimates for ξ.

3.4.2 Code

```r
## SETUP
library(quantmod)
library(lubridate)
getSymbols(Symbols = c("000425.SZ", "JPM"), from = as.Date(today()) - years(5), to = today(), auto.assign = TRUE)
XCMG_returns <- dailyReturn(`000425.SZ`)
JPM_returns <- dailyReturn(JPM)
# Split into Gains and Losses
XCMG_gains <- na.omit(ifelse(XCMG_returns > 0, XCMG_returns, NA))
XCMG_losses <- na.omit(ifelse(XCMG_returns < 0, abs(XCMG_returns), NA))
JPM_gains <- na.omit(ifelse(JPM_returns > 0, JPM_returns, NA))
JPM_losses <- na.omit(ifelse(JPM_returns < 0, abs(JPM_returns), NA))
rm(list=setdiff(ls(), c("XCMG_gains", "XCMG_losses", "JPM_gains", "JPM_losses")))
## MaxSum Ratio
plot_ratio <- function(returns, p, name="NA") {
  n <- length(returns)
  ratios <- numeric(n)
  for (i in 1:n) {
    M_n <- max(returns[1:i]^p)
    S_n <- sum(returns[1:i]^p)
    ratios[i] <- M_n/S_n
  }
  print(ratios[n-1])
  plot(1:n, ratios, type="p", main=paste("Max-Sum Ratio, ", name, ", p=", p, sep=""), xlab="n", ylab="Ratio M_n(p) / S_n(p)", col="black")
}
# Plot and Print for different p's
for (p in 1:3) {
  plot_ratio(XCMG_gains, p, name="XCMG Gains")
  plot_ratio(XCMG_losses, p, name="XCMG Losses")
  plot_ratio(JPM_gains, p, name="JPM Gains",)
  plot_ratio(JPM_losses, p, name="JPM Losses")
}
# Omitted: Saving PNG Files
```

Listing 6: MaxSum Ratio Plots

```r
## Setup:
library(quantmod)
library(lubridate)
getSymbols(Symbols = c("000425.SZ", "JPM"), from = as.Date(today()) - years(5), to = today(), auto.assign = TRUE)
XCMG_returns <- dailyReturn(`000425.SZ`, type='log')
JPM_returns <- dailyReturn(JPM, type='log')
clean_returns <- function(returns) {
  clean_returns=as.numeric(returns[!is.na(returns) & !is.infinite(returns)])
}
XCMG_gains <- clean_returns(na.omit(ifelse(XCMG_returns > 0, XCMG_returns, NA)))
XCMG_losses <- clean_returns(na.omit(ifelse(XCMG_returns < 0, abs(XCMG_returns), NA)))
JPM_gains <- clean_returns(na.omit(ifelse(JPM_returns > 0, JPM_returns, NA)))
JPM_losses <- clean_returns(na.omit(ifelse(JPM_returns < 0, abs(JPM_returns), NA)))
XCMG_returns <- clean_returns(na.omit(dailyReturn(`000425.SZ`, type='log')))
JPM_returns <- clean_returns(na.omit(dailyReturn(JPM, type='log')))

## Probability Weighted Moments (PWM):
library(POT)
plot_pwm <- function(x, u) {
  x <- as.numeric(coredata(x))
  excess <- x[x>u] - u
  gpd_estimates <- fitgpd(x, u, est='pwmb')
  print(gpd_estimates$fitted.values)
}
plot_pwm(XCMG_returns, u=0.0150) #scale=0.01513291, shape=0.15397963
plot_pwm(XCMG_gains, u=0.008) #scale=0.01684079, shape=0.0137954
plot_pwm(XCMG_losses, u=0.0225) #scale=0.009770835, shape=0.180511721
plot_pwm(JPM_returns, u=0.0200) #scale=0.01066543, shape=0.15560997
plot_pwm(JPM_gains, u=0.0194) #scale=0.01050621, shape=0.31312970
plot_pwm(JPM_losses, u=0.0250) #scale=0.008810297, shape=0.403584109
```

Listing 7: Method of Probability Weighted Moments

4 Extremal Index Estimation

The extremal index θ lives between 0 and 1.[2] Higher values indicate that extremes are less clustered and therefore more likely iid.

When extremes are iid, the magnitude of extremes is larger. We want to identify the extremal indices of XCMG and JPM returns to get an idea of how to best prepare for extreme events.

For example, if the extremal index is very close to 1, then extremes are going to be less correlated and therefore more protection against extreme events is needed. However, when extremal index is closer to 0, it's possible less precautions are necessary to take against extreme events.

Below, I've pasted output text for the blockTheta approach of computing the extremal indices. The extremal index for each set of returns is in the rightmost column. This index is estimated for multiple quantiles for robustness; we wish to avoid volatility in the extremes while also looking for an accurate measure of the index. My interpretation will follow after I've shown all of the plots.

4.1 XCMG

```
> blockTheta(XCMG_returns)

Title:
 Extremal Index from Block Method

Call:
 blockTheta(x = XCMG_returns)

Extremal Index:
   quantiles thresholds  N  K     theta
1      0.950 0.03517589 64 36 0.8890492
2      0.955 0.03729600 58 36 0.9835625
3      0.960 0.03943665 52 32 0.9021778
4      0.965 0.04239765 46 30 0.9246855
5      0.970 0.04470590 40 26 0.8654236
6      0.975 0.05022159 34 24 0.9143790
7      0.980 0.05671644 28 20 0.8775154
8      0.985 0.06032914 22 14 0.7277060
9      0.990 0.06976749 16 12 0.8404761
10     0.995 0.08372831 10  7 0.7456288
```

Figure 33: XCMG Extremal Index

4.1.1 Gains

```
> blockTheta(XCMG_gains, block=22)

Title:
 Extremal Index from Block Method

Call:
 blockTheta(x = XCMG_gains, block = 22)

Extremal Index:
   quantiles thresholds  N  K     theta
1      0.950 0.04433497 39 17 0.7298809
2      0.955 0.04511274 36 17 0.7930963
3      0.960 0.04721036 35 16 0.7280803
4      0.965 0.05022159 32 16 0.7987306
5      0.970 0.05366721 29 16 0.8839928
6      0.975 0.05637984 27 16 0.9513574
7      0.980 0.05882352 24 13 0.7622902
8      0.985 0.05975398 21 11 0.6866624
9      0.990 0.06144077 19 10 0.6685129
10     0.995 0.06484635 17  9 0.6527944
```

Figure 34: XCMG Gains Extremal Index

4.1.2 Losses

```
> blockTheta(XCMG_losses, block=22)

Title:
 Extremal Index from Block Method

Call:
 blockTheta(x = XCMG_losses, block = 22)

Extremal Index:
   quantiles thresholds  N  K     theta
1      0.950 0.03732500 30 18 0.9635669
2      0.955 0.03820598 27 18 1.0734502
3      0.960 0.03938737 24 16 0.9896363
4      0.965 0.04166666 21 15 1.0240838
5      0.970 0.04440497 18 14 1.0796350
6      0.975 0.04772728 15 11 0.9299040
7      0.980 0.05132449 12  8 0.7826318
8      0.985 0.05625880  9  7 0.8934759
9      0.990 0.06140349  6  5 0.9169126
10     0.995 0.07580175  3  3 1.0573682
```

Figure 35: XCMG Losses Extremal Index

4.2 JPM

```
> blockTheta(JPM_returns)

Title:
 Extremal Index from Block Method

Call:
 blockTheta(x = JPM_returns)

Extremal Index:
   quantiles thresholds  N  K     theta
1      0.950 0.02608773 68 31 0.6399711
2      0.955 0.02794445 62 30 0.6698295
3      0.960 0.02877931 56 27 0.6386169
4      0.965 0.03078233 49 22 0.5561704
5      0.970 0.03283916 43 19 0.5282083
6      0.975 0.03492313 37 16 0.5000505
7      0.980 0.04006235 31 13 0.4700636
8      0.985 0.04468598 24 11 0.5043348
9      0.990 0.04695584 18  9 0.5402774
10     0.995 0.05792113 12  6 0.5257899
```

Figure 36: JPM Extremal Index

4.2.1 Gains

```
> blockTheta(JPM_gains, block=22)

Title:
 Extremal Index from Block Method

Call:
 blockTheta(x = JPM_gains, block = 22)

Extremal Index:
   quantiles thresholds  N  K     theta
1      0.950 0.03283916 43 12 0.3479158
2      0.955 0.03314729 40 12 0.3749406
3      0.960 0.03492313 37 10 0.3217223
4      0.965 0.03733999 34 10 0.3509739
5      0.970 0.04006235 31  9 0.3390770
6      0.975 0.04153276 27  8 0.3392939
7      0.980 0.04468598 24  7 0.3274876
8      0.985 0.04623735 21  7 0.3751793
9      0.990 0.04695584 18  7 0.4387665
10     0.995 0.05397752 15  7 0.5277865
```

Figure 37: JPM Gains Extremal Index

4.2.2 Losses

```
> blockTheta(JPM_losses, block=22)

Title:
 Extremal Index from Block Method

Call:
 blockTheta(x = JPM_losses, block = 22)

Extremal Index:
   quantiles thresholds  N  K     theta
1      0.950 0.03223347 43 14 0.4421103
2      0.955 0.03267759 40 13 0.4282260
3      0.960 0.03447087 37 13 0.4641848
4      0.965 0.03493882 34 10 0.3567600
5      0.970 0.03711231 31  9 0.3438496
6      0.975 0.03778205 28  8 0.3307974
7      0.980 0.03953833 25  7 0.3172435
8      0.985 0.04161400 22  6 0.3026837
9      0.990 0.04333857 19  6 0.3513884
10     0.995 0.04596596 16  5 0.3409150
```

Figure 38: JPM Losses Extremal Index

4.3 Code

```r
## Setup:
library(quantmod)
library(lubridate)
getSymbols(Symbols = c("000425.SZ", "JPM"), from = as.Date(today())
    - years(5), to = today(), auto.assign = TRUE)
XCMG_returns <- dailyReturn(`000425.SZ`)
JPM_returns <- dailyReturn(JPM)
XCMG_gains <- clean_returns(na.omit(ifelse(XCMG_returns > 0, XCMG_
    returns, NA)))
XCMG_losses <- clean_returns(na.omit(ifelse(XCMG_returns < 0, abs(
    XCMG_returns), NA)))
JPM_gains <- clean_returns(na.omit(ifelse(JPM_returns > 0, JPM_
    returns, NA)))
JPM_losses <- clean_returns(na.omit(ifelse(JPM_returns < 0, abs(JPM_
    returns), NA)))
XCMG_returns <- clean_returns(na.omit(XCMG_returns))
JPM_returns <- clean_returns(na.omit(JPM_returns))

## Compute Block Thetas:
library(fExtremes)
blockTheta(XCMG_returns, block=22)
blockTheta(XCMG_gains, block=22)
blockTheta(XCMG_losses, block=22)
blockTheta(JPM_returns, block=22)
blockTheta(JPM_gains, block=22)
blockTheta(JPM_losses, block=22)
```

Listing 8: Extremal Index

4.4 Extremal Index: Interpretation

The XCMG returns extremal index appears to be about 0.87.
The XCMG gains extremal index appears to be about 0.79.
The XCMG losses extremal index appears to be about 1.

The JPM returns extremal index appears to be about 0.52.
The JPM gains extremal index appears to be about 0.37.
The JPM losses extremal index appears to be about 0.33.

We notice that the extremal indices for XCMG are mostly large. This indicates that XCMG has returns that do not cluster.[4] Therefore, portfolio investment strategies involving XCMG should be more risk averse.

(In the beginning, after eyeballing the XCMG log returns, I made a point that the extremes looked iid. This assumption is in line with what we found from the blockTheta estimates.)

For JPM returns, the extremal index is smaller, around 0.5. This is not very small, but certainly indicates some clustering of extremes.[4] Therefore, this implies that the extremes are less as large in magnitude compared to the case where the extremes are independent.

Therefore, when building a portfolio model with JPM returns, it may be smart to allow for extra risk in light of extra reward with JPM returns.

5 Conclusion

The shape parameters for XCMG and JPM returns (all returns, gains, and losses) are calculated above using three methods.

In practice, its common to see these analyses split into gains and losses. I incorporated all returns alongside gains and losses to add robustness to the estimates for ξ.

Pickands estimates were much more varied than Hills and PWMs. This is likely due to Pickands estimator using few order statistics.

In both Hill and Pickand's plots, the empirical mean excess threshold was not directly used. Instead, parts of the plot that stabilized around the excess mean thresholds were used. Moreover, for both cases, the estimates were plotted within a confidence interval that mostly contained the lower estimates found from fitting GPDs with the Probability Weighted Moments method.

The method of PWMs appears to make the most sense. It's sensible that the shape parameter of XCMG returns is smaller than that of JPM returns, as we observed in the log returns plot in the beginning, the JPM returns looked extremely volatile at COVID-19 times, and the XCMG returns were more randomly distributed over 5 years.

I end up choosing the PWM GPD-fit Estimates for ξ because I believe they are the most informative and robust estimates of ξ.

I concluded that XCMG has relatively iid extremes (indicated by $\hat{\theta} \approx 1$) with shape parameters:

All returns $\hat{\xi}_X = 0.154$

Gains $\hat{\xi}_{X+} = 0.165$

Losses $\hat{\xi}_{X-} = 0.148$,

and that JPM has more clustered extremes (indicated by $\hat{\theta} \approx 0.5$) with shape parameters:

All returns $\hat{\xi}_J = 0.332$

Gains $\hat{\xi}_{J+} = 0.313$

Losses $\hat{\xi}_{J-} = 0.404$.

5.1 Implications of Results

These results allow one to get a better idea of how the extreme values for XCMG and JPM gains and losses behave. Thus one can construct more robust models to forecast the returns of XCMG and JPM. Particularly, more risk averse strategies are advised for XCMG modeling, where extreme returns take on large values. And more risk-tolerable strategies are advised for modeling with JPM returns, where the extremes are expected to take on less reduced values because of clustering.

It makes sense that the JPM, Fincial services industry stock, would have a bigger shape parameter than the XCMG, industrial equipment manufacturer stock. This is because JPM deals with more volatile methods of revenue generation than a company like XCMG. It also makes sense that the extreme returns of JPM are more correlated than those of XCMG, for XCMG operates with construction companies all around the world, and JPM makes money from worldwide financial markets and companies. This may help to explain the differences in extremal indices for both stocks, as well as the differences in shape parameters of the extreme value distributions for each stock's returns.

6 References

[1] Gennady Samorodnitsky, Spring 2024, ORIE 4565 Extreme Values in Finance, Project Instructions.pdf, url=https://canvas.cornell.edu/courses/63072/files/10412847?module_item_id=2553247

[2] Gennady Samorodnitsky, Spring 2024, ORIE 4565 Extreme Values in Finance Course Notes, url=https://canvas.cornell.edu/courses/63072/modules

[3] R-Studio, 2024, url=https://posit.co/download/rstudio-desktop/

[4] Gennady Samorodnitsky, Lecturer, Spring 2024, ORIE 4565 Extreme Values in Finance, Cornell University