Copenhagen Business School

Kandidatafhandling

Statistical Modeling of Extreme Values

Statistisk modellering af ekstreme værdier

Forfatter:

Anders C. T. J. Pedersen

Student number: 93648

Cand.Merc.Mat

Vejleder:

Dorte Kronborg

Department of Finance

Antal sider: 77

Antal tegn: 113.144

Afleveringsdato: 31.07.2019

CBS COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Resumé

Denne kandidatafhandling er et studie i Ekstrem Værdi Teori (EVT), og diverse statistiske metoder til at beskrive adfærden af ekstreme hændelser. Udvalgte dele af teorien beskrives og med et afsæt i relevant finansielt data efterprøves dele af denne. Fælles for afhandlingen er antagelsen af uafhængigt identisk fordelt data (iid). Del 1 berører asymptotiske modeller, og dækker både blok - og grænseværdi modeller, dette efterprøves og analyseres på to serier af aktiekurser, Mærsk B og Novo Nordisk B. Dernæst introduceres et teoretisk indblik i Poisson processer til modellering af ekstreme værdier. Til slut i del 1 introduceres multidimensionelle modeller af ovenstående, og afsluttes med at berøre modellering af afhængige serier af data. Del 2 omhandler brugen af EVT til at modellere finansiel risiko, herunder Value-at-Risk (VaR) og Expected Shortfall (ES) og disses sammenhæng med EVT. Del 3 fokuserer på diskussion af de forskellige metoder og modeller, og afsluttes med en konklusion af det analyserede, såvel som afhandlingen generelt.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Extreme value theory (hence EVT) is an area of great interest across a broad spectrum of companies, especially in the financial sector. Both insurance companies, investment firms and banks are more than interested in estimating monetary amounts at risk, for their individual investments. Focusing on the investment firms and banks, especially different positions in assets are attractive to look at.

When talking extreme values, it is equivalent to look at both tails of distribution curves, and with regards to the different positions to take in a stock or equity, this corresponds to short and long positions (Gilli and Këllezi - 2006). Usually when discussing risk, one would look at how much is at stake would some bad outcome happen, e.g. for the insurance companies, only one side of the tail is interesting, as this would tell something about under what circumstances would there be many instances of e.g. house fires. This is not the same case for an investor (be it an investment firm or a bank), as one would have to take into account what sort of position is taken in the asset.

This thesis will use two equities, and the daily closing prices as foundation for the empirical analysis. The two equities are chosen based on different parameters.

- Local firms - Danish firms from the C20 index are more relatable.
- Large firms - Larger firms make it easier to find information, and are traded more often.
- Old firms - With regards to the literature, many datapoints are necessary for most of the analysis.

Based on these parameters, the C20 index was reduced, and since some of the firms on it, were not always publicly traded, it was reduced further. By choice of the third parameter, old firms, Maersk and Novo Nordisk where singled out, being two of the oldest traded. Both have had their firms, and equities split at some point, and since there is no reason to research further into the differences, the Maersk B equity and Novo Nordisk B equities were chosen as the equities that would make up the dataset. More on the dataset in the following subsection, Data exploration.

When handling financial risk, it is often hard to reasonably determine the true distribution of the data, and financial data has a great risk of being heavy-tailed. Distribution of data with heavy-tails causes great challenges for institutions, when it comes to correctly assessing how much is at risk. Here EVT presents theories of how to more correctly estimate various distributions, based on the extreme values observed, and thus better approximate worst case outcomes.

The overall goal of this thesis is to research how to locate extreme values, with practical application on data of the two equities chosen. The research will revolve around different methods of doing so, both when looking at the equities one by one, but also when they are combined. The latter is covered in the multivariate theories, and aim to research if an investor can say something about risk on one asset, based on the current state of another.
Finally, measures of financial risk will constitute a basis for putting EVT theory into perspective. The parameters of EVT will be used to build the widely used risk measures Value-at-Risk (VaR) and Expected Shortfall (ES).

1.2 Problem formulation

The overall purpose of the thesis is to research extreme value theory. Different methods and approaches will be evaluated based on their possibility/ability to estimate extreme events. Methods will be tested on the Maersk B and Novo Nordisk B equities.

In order for this research to be fulfilled, the following research questions are raised:

- How to locate extreme values, and how are these classified in the different approaches of EVT?
- What methods of estimation are favorable to use in EVT, and which of these are best?
- Analyse the extreme events on Maersk B and Novo Nordisk B equities.
- Evaluate the multivariate theories of EVT

1.3 Methodology

This thesis uses a deductive approach, as theory will be researched and then tested on relevant data. It has a theoretical methodology, and will aim to answer the research questions, mainly by searching the literature and secondly with empirical test and data analysis.

The thesis is split into three main parts, each covering different aspects of the topic of extreme value theory. Part 1 covers the theory of block maxima, threshold, point process and multivariate extremes. It also covers the data analysis of block maxima models, threshold models and the multivariate models. Part 2 covers the risk management aspects of EVT, and will discuss the usage. Part 3 rounds of the thesis by putting some of the topics into different context, provides topics for further research, and ends with a conclusion that picks up the relevant findings of the thesis.

The primary source of literature for this thesis, is An Introduction to Statistical Modeling of Extreme Values (Coles, S., 2001 - Appendix A.1). Besides the primary literature, Quantitative Risk Management (J. McNeil, A. et al, 2005 - Appendix A.1) has been used as replacement for some of the equations and theorems. Various articles from academic journals has been used for parts 2 and 3 of the thesis. All literature used will be acknowledged in the appendix.

For the statistical work, and the data analysis the statistical software **R** has been used. Different packages in **R** has been utilized in order to perform the various analytical work needed throughout the thesis. These are the following, with short descriptions:

- **extRemes**: Functions for modeling extreme value distributions, e.g. for "Mean residual life plot".
- **evir**: Functions for modeling extreme value data, e.g. "Mean excess" and modeling the distributions.
- **evd**: Functions for modeling multivariate extremes

Besides the packages used, a few custom functions were made in order to manipulate the data to the extent needed. These functions along with the relevant **R**-code, the analytical work of the thesis, is supplied in the appendix.

1.4 Limitation

In order to comprehend the content of this thesis, it is assumed that the reader has a general knowledge of financial -, statistical - and probability theory.
To reasonably answer and evaluate the research questions raised on the problem formulation part, the following limitations have been imposed on this thesis.
With regards to keeping the extent of the thesis at a sensible level, only some of the theory is researched. The parts of the theory chosen is based on the idea that the basic elements of EVT should be researched, and from there build up the theoretical difficulty

throughout the thesis.
The data analysis parts of the thesis are limited to what is referred to as classical EVT (block maxima models and threshold models), and something about the multivariate theory. Neither Point process or dependent series chapters will have analysis parts to them, and thus are only there to broaden the theoretical aspects of the thesis, and to provide perspective about the data structure and methods.
Within the theoretical parts of the thesis, not all models and theorems are deduced fully, since it would be out of the scope of the thesis to complete every calculation and proof fully.
The objective of the data analysis is to enhance the theoretical knowledge with real life examples, and to provide insight in practical application of the theories described.

1.5 Data exploration

The data has been extracted from Bloomberg on 16/04/2019, and contains the full length of closing prices available. For Maersk this results in data from 06/10/1989 up until 16/04/2019. For Novo Nordisk this results in data from 24/07/1991 up until 16/04/2019. To have comparable datasets for the two equities, both sets have been reduced so they start 02/01/1992, being the first trading day of the year. Equally they have both been shorted off so they end 28/12/2018. By doing this both equities are of comparable length, and also have complete years, and quarters for further analysis.

After shorting off the datasets, further cleaning of the data is needed. Cleaning of the datasest was done, as there are days (mostly at the start of their lifetime), where the equities where not traded. Since it is relevant to have as many possible datapoints available, cleaning has been carried out such that is there are no trades for day $t = n$, then the closing price of day $t = n - 1$. Other possibilities would have been to simulate the data, but this is outside the scope of this thesis.

To get an initial overview of how the data is structured, and to begin the research,

data was plotted as this gives a very nice view of the two equities. Both the daily closing prices and their equivalent log returns are plotted, log returns being a well used transformation of closing prices and thus returns. The transformation allows to look at the data in a desirable way, especially when it comes to EVT, as the interest is to look at the changes. The definition of log return can be found various places, this is from[1]:

$$\mathrm{LN}(P_n/P_{n-1}) \tag{1.1}$$

With P_n denoting the closing price of the equity at day n.

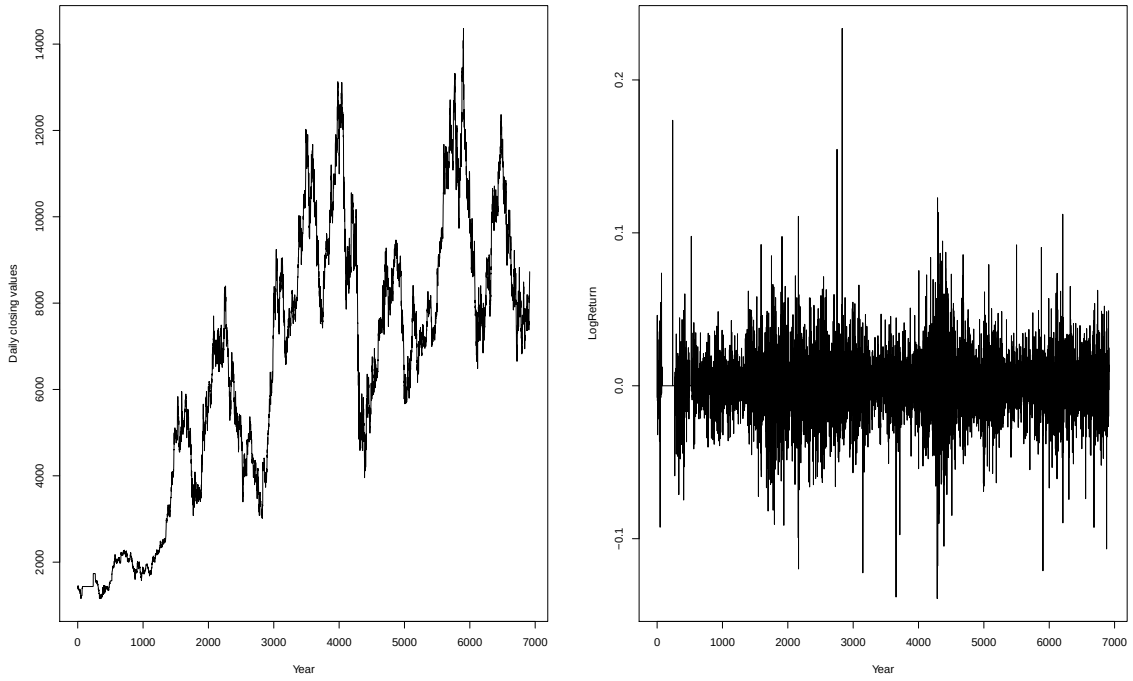

Figure 1: Left: Daily closing values of Maersk B, Right: Log Returns

These two plots give a nice initial overview of the two equities, and areas of interest within the reduced time period. Plots of the Novo Nordisk equity can be found in the appendix. The plots of the two equities log returns really gives a good indication of how large, and where the extremes are to be found.

[1] `https://quantivity.wordpress.com/2011/02/21/why-log-returns/`

Part I

Extreme Value Theory

2 Classical Extreme Value Theory and Methods

The purpose of this section is to explain the theory behind the fundamental theories of Extreme Value Theory. These includes the asymptotic model of block maxima, and the theory concerning estimation of the distribution of maxima, and the return levels. Later in the section threshold models will be introduced and explained. To round off the section, analysis of the data of the two equities introduced in section (1.5).
The work in this section is referenced to (Coles, S., 2001, Chapter 3 - Appendix A.1) unless other sources are stated.

A key parameter in Extreme Value Theory, is the definition of maxima of a distribution, since the extreme values will be some of the maximum values of a given distribution. Defining M_n as the maxima of a sequence of iid random stochastic variables, such:

$$M_n = \max\{X_1, ..., X_n\} \tag{2.1}$$

This maxima can be derived if the distribution of the underlying data is known, but as this is often not the case, as this is one of the aims of the upcoming theories a reformulation is appropriate. For some sequence of constants $\{a_n\}$ and $\{b_n\}$, the maxima can be linearly reformulated as:

$$M_n^* = \frac{M_n - b_n}{a_n} \tag{2.2}$$

for $\{a_n > 0\}$.

The aim is to search for limiting distributions for maxima M_n^* , and this is done in this chapter with the use of asymptotic models.
The classical approach to Extreme Value Theory (hence EVT), is to use asymptotic models to estimate the distribution functions. This allows for regular estimators to be used, such as Maximum Likelihood, LogLikelihood and so forth. The three families of the distribution functions, given the classical theory are the Gumbel, Fréchet and Weibull families. The families are:

$$G(z) = \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, -\infty < z < \infty \tag{2.3}$$

$$G(z) = \begin{cases} 0, & z \leq b, \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{\alpha}\right\}, & z > b \end{cases} \tag{2.4}$$

$$G(z) = \begin{cases} \exp\left\{-\left[-\left(\frac{z-b}{a}\right)^{\alpha}\right]\right\}, & z < b, \\ 1, & z \geq b \end{cases} \tag{2.5}$$

Having three different distributions to choose from, makes the implementation from theory to practice a tough. Picking the right family is not easy, and there is no confirmation that the chosen family would be the correct one. Instead of having to struggle choosing the correct family, they are combined in the Generalized Extreme Value Distribution.

2.1 The Generalized Extreme Value Distribution - GEV

Given the existence of parameters $\{a_n > 0\}$ and $\{b_n\}$, thus for a non-degenerate distribution function G:

$$\lim_{n\to\infty} \Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \to G(z) \tag{2.6}$$

So basically the limit states that a distribution of maxima can be approximated by the GEV for some large value on n.
If and only if that is true, then G is a member of the joined family of distribution functions:

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{2.7}$$

defined on $\{z : 1+\xi(z-\mu)/\sigma > 0$ for $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$.
The parameters, ξ, σ, μ are the three central parameters of the distributions.

- ξ is the shape parameter, which determines which of the three families a distribution is.
- σ is the scale parameter.
- μ is the location parameter.

The reformulation of the maxima M_n can pose a challenge in practice, but this is fixed by rewriting, and assuming large enough value for n:

$$\begin{aligned} &\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \approx G(z) \\ &\iff \Pr\{M_n \leq z\} \approx G\left\{\frac{z-b_n}{a_n}\right\} \\ &\qquad = G^*(z) \end{aligned} \tag{2.8}$$

where G^* is a different member than G of the GEV family. Thus according to equations (2.7) and (2.6), if M_n^* can be approximated by a one of the GEV family, then M_n can be approximated by a different member of the GEV family. This leads to the outlining of the first models, the block maxima models.

2.2 Block Maxima Models

Observing a set of iid observations: $x_1, x_2...$, the first of the models for estimating extreme values, is the block maxima approach. Here a block size is chosen, depending on the size of the dataset, and within each block the k biggest observations are chosen as extremes. This has the potential limitation that in a single period there can be more extremes than chosen, and in others there can be chosen extremes that in reality are not really extremes. For financial data it is usually reasonable to chose block sizes according to some periods of interest, this includes months, quarters or years, whereas for other data types other types of blocks might prove more reasonable.

Under the assumption of independent observations, this will be revisited in section (5), a series of independent stochastic variables $X_1, X_2...$ are observed. These observations are blocked appropriately to the chosen block size n, so the block maxima of the series is $M_{n,1}, ..., M_{n,m}$, that can then be fitted to the GEV distribution.
One of the great challenges when using the block maxima methods, concerns choice of block size. With block maxima methods, the amount of datapoints are dramatically reduced, as the block size increases.
The fact that the amount of datapoints are reduced, poses the challenge of reasonably estimating the distribution of the maxima, as estimation becomes less feasible, as for e.g. yearly maxima for a ten year old equity would result in having to estimate the distribution based on ten datapoints.

2.2.1 Return level

The return level z_p is a measure of how often an extreme value will be exceeded, with some probability p. This is relevant, in the case of financial data, it can give an indication of how probable some extreme outcome is over a given period of time, e.g. a statement like: "The equity experiences negative returns of more than 9.5% on average every 20 years".

The return level, z_p can be estimated by inverting equation (2.7):

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right], & \text{for } \xi \neq 0, \\ \mu - \sigma\log\{-\log(1-p)\}, & \text{for } \xi = 0 \end{cases} \tag{2.9}$$

where $G(z_p) = 1 - p$. Specifically z_p is the return level related to the return period $1/p$. When it comes to visualizing the return levels, the parameters of the GEV has some nice secondary properties.
Defining $y_p = -log(1-p)$, so equation (2.9) can be expressed as:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - y_p^{-\xi}\right], & \text{for } \xi \neq 0, \\ \mu - \sigma\,\log y_p, & \text{for } \xi = 0 \end{cases} \tag{2.10}$$

results in, that if z_p is plotted against y_p on a logarithmic scale, the curvature of the plot will reveal the nature of parameter ξ:

- If $\xi = 0$ the plot will be linear
- If $\xi < 0$ the plot will be convex with an asymptotic limit at: $\mu - \sigma/\xi$, as $p \to 0$
- If $\xi > 0$ the plot will be concave

These plots are useful for validation of the model, and will be revisited in the empirical part of this section.

2.2.2 Maximum likelihood estimation

For the maximum likelihood estimators, ξ also impacts the possibilities of estimating these:

- for $\xi > -0.5$, maximum likelihood estimates are regular, and have asymptotic properties

- for $-1 < \xi < -0.5$, maximum likelihood estimators are obtainable, but does not have asymptotic properties
- for $\xi < -1$, maximum likelihood estimators are unobtainable

Assuming that $Z_1, ..., Z_m$ are independent variables distributed by the GEV, log-likelihood of the GEV parameters, in the case of $\xi \neq 0$ are:

$$\ell(\mu, \sigma, \xi) = -m \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{m} \log \left[1 + \xi \left(\frac{z_i - \mu}{\sigma}\right)\right] - \sum_{i=1}^{m} \left[1 + \xi \left(\frac{z_i - \mu}{\sigma}\right)\right]^{-1/\xi} \tag{2.11}$$

so long as

$$1 + \xi \left(\frac{z_i - \mu}{\sigma}\right) > 0, \text{for } i = 1, ..., m \tag{2.12}$$

If equation (2.12) is not fulfilled the likelihood = 0, resulting in the log-likelihood = $-\infty$. The alternate case of $\xi = 0$, i.e. the Gumbel distribution, and the log-likelihood is:

$$\ell(\mu, \sigma) = -m \log \sigma - \sum_{i=1}^{m} \left(\frac{z_i - \mu}{\sigma}\right) - \sum_{i=1}^{m} \exp\left\{-\left(\frac{z_i - \mu}{\sigma}\right)\right\} \tag{2.13}$$

Maximization of equations (2.11) and (2.13) results in maximum likelihood estimates for the full GEV family. Noting that numerical optimization of the MLE is necessary to achieve correct results. The estimates $\{\hat{\xi}, \hat{\sigma}, \hat{\mu}\}$ are asymptotically multidimensional normally distributed with regards to ξ.

Deduction of the return level is then done by substituting in the maximum likelihood estimates in equation (2.10):

$$\hat{z}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left[1 - y_p^{-\hat{\xi}}\right], & \text{for } \hat{\xi} \neq 0, \\ \hat{\mu} - \hat{\sigma} \log y_p, & \text{for } \hat{\xi} = 0 \end{cases} \tag{2.14}$$

Variance of the return level estimation can be deduced by use of the delta method:

$$\text{Var}(\hat{z}_p) \approx \nabla z_p^T \text{V} \nabla z_p \tag{2.15}$$

with V being the variance-covariance matrix of the estimates $\{\hat{\xi}, \hat{\sigma}, \hat{\mu}\}$, and:

$$\begin{aligned} \nabla z_p^T &= \left[\frac{\partial z_p}{\partial \mu}, \frac{\partial z_p}{\partial \sigma}, \frac{\partial z_p}{\partial \xi} \right] \Big|_{(\hat{\mu},\hat{\sigma},\hat{\xi})} \\ &= \left[1, -\xi^{-1}(1 - y_p^{-\xi}), \sigma\xi^{-2}(1 - y_p^{-\xi}) - \sigma\xi^{-1} y_p^{-\xi}) \log(y_p) \right] \Big|_{(\hat{\mu},\hat{\sigma},\hat{\xi})} \end{aligned} \tag{2.16}$$

where $y_p = -\log(1 - p)$.

As mentioned, the assumptions regarding independence of the extreme observations might be unrealistic when working with financial data. Being unable to satisfy this assumption, makes the maximum likelihood estimations somewhat unreliable. For this reason, profile likelihood will be considered a reasonable alternative for estimating the return levels.

2.2.3 Profile likelihood

Profile log-likelihood is the method of maximizing a parameter vector ϕ, which in the case of GEV is $\phi = (\mu, \sigma, \xi)$. Log-likelihood for ϕ is $\ell(\phi_i, \phi_{-i})$, where ϕ_{-i} is the notation of all other parameters of ϕ, except ϕ_i. The profile log-likelihood for ϕ_i is:

$$\ell_p(\phi_i) = \max_{\phi_{-i}} \ell(\phi_i, \phi_{-i}) \tag{2.17}$$

Profile likelihood becomes very useful, as it can provide a confidence interval for the return level z_p. This is done by isolating one of the parameters in equation (2.9), here μ is isolated, and thus the density is reparameterized with $\{z_p, \sigma, \xi\}$:

$$\mu = z_p + \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right] \tag{2.18}$$

With the isolation of one of the parameters, it now allows for the calculation of the profile likelihood of the return level, by maximization of the other parameters. This method will be used to calculate the confidence intervals for the return levels, in the data analysis section at the end of the chapter.

2.2.4 Max-stability of distributions

The final theoretical definitions to be made concerning the GEV distribution, is the aspect of max-stability of a distribution. Max-stability means that taking the maxima values of a distribution will lead to a stable distribution, i.e. the same distribution as the underlying, though with other values of μ and σ. A distribution G, is called max-stable, if for $n = 2, 3, ...$, constants $\alpha_n > 0$ and β_n exists, then:

$$G^n(\alpha_n z + \beta_n) = G(z) \tag{2.19}$$

A distribution is max-stable if, and only if, it is a GEV distribution.
Proving this statement is rather straight forward, and is done by considering the following.
M_{nk} is the maxima of a vector of $n \times k$ variables. This is equivalent to maxima of a one-dimensional vector of lenght $n \times k$ and the maxium of k maxima, where each k is the maxima of n observations. Supposing also, the limiting distribution of $(M_n - b_n)/a_n \to G$,

thus by equation (2.6) for a large enough n:

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \approx G(z) \tag{2.20}$$

Then for any $k > 0$, as also nk is large enough:

$$\Pr\left\{\frac{M_{nk} - b_{nk}}{a_{nk}} \leq z\right\} \approx G(z) \tag{2.21}$$

Per definition, M_{nk} could be considered as the maximum of k variables distributed equivalently to M_n, so:

$$\Pr\left\{\frac{M_{nk} - b_n}{a_n} \leq z\right\} = \left[\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\}\right]^k \tag{2.22}$$

Thus by equations (2.21) and (2.22):

$$(2.21) \Rightarrow \Pr\{M_{nk} \leq z\} \approx G\left(\frac{z - b_{nk}}{a_{nk}}\right) \tag{2.23}$$

$$(2.22) \Rightarrow \Pr\{M_{nk} \leq z\} \approx G^k\left(\frac{z - b_n}{a_n}\right) \tag{2.24}$$

So it shows that G^k is the same as G, in all aspects but μ and σ, and therefore all members of the GEV family are max-stable.

For now this concludes the theoretical outline of the block maxima models, and the next section will cover the threshold models.

2.3 Threshold Models

Threshold models take another approach to finding maxima than the explained in the previous section.
The work in this section is referenced to (Coles, S., 2001, Chapter 4 - Appendix A.1)
Instead of grouping observations on blocks, and only categorizing the block maxima as extreme, threshold models use a chosen threshold, and categorize every observation above (or below for minima), as being extreme. To quantify this, the probability functions is presented as being the conditional probability:

$$\Pr\{X > u + y \mid X > u\} = \frac{1 - F(u+y)}{1 - F(u)}, y > 0 \tag{2.25}$$

M_n defines to the same as in the previous section, hence the $\Pr\{M_n \leq z\} \approx G(z)$ where $G(z) = \text{Function}(2.7)$, for some given σ and $\mu > 0$, and ξ.
From this the distribution function of our conditional excess term $(X - u|X > u)$, is approximated as:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \tag{2.26}$$

defined on $\{y : y > 0 \text{ and } (1 + \xi y/\tilde{\sigma}) > 0\}$ were $\tilde{\sigma} = \sigma + \xi(u - \mu)$.

Equation 2.26 defines the family of distributions known as the generalized Pareto family. The connection between block maxima, and the generalized Pareto family is so, that if block maxima is approximated by a distribution G, the threshold excess will have an analogous approximating distribution of the generalized Pareto family.
The associated values of the GEV distribution of block maxima will uniquely determine the parameters of the generalized Pareto distribution (hence GPD) of the threshold excess, especially but not limited to the value of the parameter ξ, as this will be identical in the two distributions.

As with the GEV distribution, the GPD has a set of special cases, that will determine how the distribution function behaves. When $\xi < 0$ the distribution will take form of a short-tailed version of Pareto type II. If $\xi = 0$ the distribution will be exponential, and can be interpreted as:

$$\lim_{\xi \to 0}(2.26) \Rightarrow$$
$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right) \tag{2.27}$$

for $y > 0$.
In the case of $\xi > 0$, the distribution G will take form of an ordinary Pareto distribution, defined by parameters $\alpha = \frac{1}{\xi}$ and $k = \frac{\beta}{\xi}$.

2.3.1 Choice of threshold

Taking a sequence of iid datapoints $x_1, ..., x_n$, the datapoints exceeding the threshold u will be identified as the extremes. These extremes are labeled $x_{(1)}, ..., x_{(k)}$, hence the threshold exceedances can be defined as $y_j = x(j) - u$ for $j = 1, ..., k$. The distribution of these exceedances can be approximated by a form of the generalized Pareto family.
As with the block maxima approach, there is a payoff between bias and variance, as the choice of threshold can be chosen too small, leading to large bias as the asymptotic rules will be violated as there will be too many observations being identified as extremes. On the other hand, a threshold chosen too large, will likely lead to large variance, since the amount of possible extremes will be too small to reasonably estimate the model.

Available for choosing the optimal threshold, there are two methods which offers different approaches. One approach searches to explore how stable, the parameter estimates in the model are, whereas the other approach is more pragmatic, trying out different threshold levels.

The pragmatic approach utilizes the mean of the GPD, given that Y has a GPD, with parameters σ and ξ, then

$$\mathrm{E}(Y) = \frac{\sigma}{1-\xi}, \tag{2.28}$$

if $\xi < 1$, and if $\xi \geq 1$ the mean will be infinite.
A series of datapoints, $X_1, ..., X_n$ exceeding the chosen threshold u_0, denoted by an arbitrary $\mathbf{X}$.
By inserting in equation 2.28 will result in:

$$\mathrm{E}(\mathbf{X} - u_0 \mid \mathbf{X} > u_0) = \frac{\sigma_{u_0}}{1-\xi}, \tag{2.29}$$

if, given that $\xi < 1$.
Here σ_{u_0} denotes the scale parameter, to the chosen threshold u_0. Since if the GPD is valid above the threshold u_0, equation (2.29) can be rewritten, since the GPD should also be valid for a threshold $u > u_0$, given a change in the scale parameter σ, so that:

$$\begin{aligned} \mathrm{E}(\mathbf{X} - u \mid \mathbf{X} > u) &= \frac{\sigma_u}{1-\xi} \\ &= \frac{\sigma_{u_0} + \xi u}{1-\xi} \end{aligned} \tag{2.30}$$

It shows that the expected value of the excess, the $\mathbf{X} - u$ term, is linear dependent of the threshold u. By construction, equation (2.30) gives a numerical estimate of the mean of maxima above the threshold. Since the mean is changing linearly with changes of u, it leads to the method of looking at the following:

$$\left\{ \left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) \right) : u < x_{\max} \right\} \tag{2.31}$$

as the $x_{(i)}, ..., x_{(n_u)}$ are the n_u excess observations of the chosen threshold, and the $x_{\max}$ being the largest observed of the X_i.

This result in what is called the mean residual life plot. The plot is useful for checking the threshold chosen. The plot should be linear in u for at suitable threshold, given that the GPD presents an accurate approximation of the excess distribution. Using the mean residual life plot as a tool for choosing the threshold is to be done with caution, since the estimation on the higher values are based on very few observations, and the smaller values are not considered to be extreme.
Not all plots will result in a straightforward interpretation, as the plot might not be perfectly outlined, an thus hard to interpret.

The mean residual life plot will be utilized in the data analysis subsection for choice of threshold.

2.3.2 More than one threshold

Since the mean residual life plot has its challenges, complementary techniques for choice of threshold are sometimes useful. An idea is to fit the GPD for several different thresholds, and assessing the stability of the parameters estimated.
If a GPD, for some threshold u_0, is acceptable for modelling the exceedances, then for higher values of u, exceedances will also follow a GPD. For the two distributions $\xi_0 = \xi$, but σ is expected to change except when $\xi = 0$, as:

$$\sigma_u = \sigma_{u_0} + \xi(u - u_0) \tag{2.32}$$

A reparameterization of the equation (2.32) can fix σ, so it will be constant with respect to u, thus both σ and ξ will be (near-)constant for a reasonable minimum threshold choice u_0:

$$\sigma^* = \sigma_u - \xi u \tag{2.33}$$

This essentially leads to estimates for both $\hat{\sigma^*}$ and $\hat{\xi}$, which should be assessed with regards to different levels of u, and thus the minimum viable threshold of the GPD can be estimated for some confidence interval. For $\hat{\xi}$ the confidence interval can be read directly from the variance-covariance matrix V, and similarly the confidence intervals for $\hat{\sigma^*}$ can be extracted by the delta method:

$$\text{Var}(\hat{\sigma^*}) \approx \nabla {\sigma^*}^T \text{V} \nabla \sigma^* \tag{2.34}$$

where

$$\nabla {\sigma^*}^T = \left[\frac{\partial \sigma^*}{\partial \sigma_u}, \frac{\partial \sigma^*}{\partial \xi} \right] = [1 - u] \tag{2.35}$$

Concluding that different methods of finding a reasonable threshold for the GPD are available, and seems to be useful when used in union. This then leads to the estimation of the different parameters.

2.4 Estimation methods for threshold models

Having chosen a threshold, estimation of the parameters of the GPD can be done be various approaches. These include maximum likelihood, probability-weighted moments and more, but maximum likelihood is the most straightforward, and is a procedure frequently used in other aspects alike, so it will be the method of choice.
The work in this section is referenced to (J. McNeil, A. et al, 2005, Chapter 7.2.2 - Appendix A.1)
β in (J. McNeil, A. et al, 2005) is equivalent to σ in (Coles, S., 2001) and has been replaced to match notation of this thesis.
Preparing the relevant data for the maximum likelihood, focus is at the N_u excess terms of the threshold u, labeling these excesses $y_1, ..., y_j = \mathbf{Y}_j$. The density of the GPD is denoted $g_{\sigma,\xi}$, the log-likelihood can be written and calculated:

$$\begin{aligned} \ell(\sigma, \xi; Y_1, ..., Y_{N_u}) &= \sum_{j=1}^{N_u} \log g_{\sigma,\xi}(\mathbf{Y}_j) \\ &= -N_u \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^{N_u} \log\left(1 + \xi \frac{\mathbf{Y}_j}{\sigma}\right) \end{aligned} \tag{2.36}$$

with the constraints that $\sigma > 0$, and $1 + \xi \mathbf{Y}_j / \sigma > 0$ for all j, as if this is not the case $\ell(\sigma, \xi; Y_1, ..., Y_{N_u}) = -\infty$.
The solution to the maximization problem results in a GPD model, $G_{\hat{\sigma},\hat{\xi}}$ for the distribution of the excess.

As the log-likelihood assumes the data to be iid, it makes sense to take a look at the case where this assumption is not met, especially being relevant when looking at financial data, as data of this type is usually dependent.
This is because the structure is resembling a Markow-chain, and thus being a datapoint x_i is dependent on the previous x_{i-1}, where i denotes a point in time.

If the underlying data is serially dependant, but there does not seem to be tenancies towards clustering in the extremes, it would suggest an extremal index of $\theta = 1$, for the underlying process.
Being connected to the modeling of dependent series, a more thorough definition will follow in section (5). For now the loose definition of θ will suffice, reference to (Coles, S., 2001, page 97):

$$\theta = (\text{limiting mean cluster size})^{-1} \tag{2.37}$$

When $\theta = 1$ is the case, the asymptotic approaches of a limit model with a Poisson process determining the occurrence of extremes, and the excess terms are iid by GPD. In the case where there is evidence of clustering in the extremes, the underlying process is assumed to have an extremal index of $\theta = 0$, this assumption of iid excess terms is not satisfied.
Considering the calculation of the log-likelihood by a quasi-maximum likelihood method, this problem can be reworked, and the estimations of the parameters be carried out in a reasonable fashion. (J. McNeil, A. et al, 2005, page 279)

Since there is now determined a model for the excess of a given threshold u, it is also possible to fit a model to a larger threshold v. Following Equation 2.25 and Equation 2.26:

$$\begin{aligned}
\tilde{F}_v(x) &= \frac{\tilde{F}(v+x)}{\tilde{F}(v)} = \frac{\tilde{F}(u+(x+v-u))}{\tilde{F}(u)} \frac{\tilde{F}(u)}{\tilde{F}(u+(v-u))} \\
&= \frac{\tilde{F}_u(x+v-u)}{\tilde{F}_u(v-u)} = \frac{\tilde{G}_{\sigma,\xi}(x+v-u)}{\tilde{G}_{\sigma,\xi}(v-u)} \\
&= \tilde{G}_{\sigma+\xi(v-u),\xi}(x)
\end{aligned} \tag{2.38}$$

It shows that for higher thresholds, the excess distribution continues to stay a GPD where the shape parameter ξ stays the same, and the scale parameter σ, which is positive linearly dependant with the higher threshold v compared to u.

Keeping to the possibility space of $\xi < 1$, the function of the excess mean is determined as:

$$e(v) = \frac{\sigma + \xi(v-u)}{1-\xi} = \frac{\xi v}{1-\xi v} + \frac{\sigma - \xi u}{1-\xi} \tag{2.39}$$

given that $u \leq v < \infty$ if $0 \leq \xi < 1$ or if $\xi < 0$ that $u \leq v \leq u - \sigma/\xi$. This method is useful for graphically checking and chosen a reasonable threshold, which will be followed up in the next subsection.

The rest of this section references (Coles, S., 2001, Chapter 4.3.3 - Appendix A.1)
Having analysed the given parameters in the model, for checking the threshold choice will lead to the graphical checking. But another convenient method is to observe the model in terms of return levels and quantiles, instead of the parameter values.
Given that a GPD with parameters σ and ξ is a good model for the exceedances of a chosen threshold u, by a variable $\mathbf{X}$, for $x > u$

$$\begin{aligned} \Pr\{\mathbf{X} > x \mid \mathbf{X} > u\} &= \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi} \implies \\ \Pr\{\mathbf{X} > x\} &= \zeta_u \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi} \end{aligned} \tag{2.40}$$

Denoting $\zeta_u = \Pr\{\mathbf{X} > u\}$.

x_m being the return level, that is the level that will be exceeded one time every m observations, on average that is, will be the solution of:

$$\begin{aligned} \zeta_u &= \left[1 + \xi\left(\frac{x_m - u}{\sigma}\right)\right]^{-1/\xi} = \frac{1}{m} \\ \text{Isolating } x_m \implies x_m &= u + \frac{\sigma}{\xi}\left[(m\zeta_u)^{\xi} - 1\right] \end{aligned} \tag{2.41}$$

This requires m to be ample large enough to guarantee that $x_m > u$, with the assumption satisfied that $\xi \neq 0$.

In the situation where $\xi = 0$, the GPD will be an exponential pareto distribution, resulting in:

$$x_m = u + \sigma \log(m\zeta_u) \tag{2.42}$$

Since x_m by construct is the m-observation return level, plotting the former against the latter on a logarithmic scale results a plot that grants the same qualitative characteristics as the return level plots with regards to the GEV model.
This returns to the same graphical and numerical analysis of the plots:

- If $\xi = 0$, a function with linear properties.
- If $\xi < 0$, it holds convex properties.
- If $\xi > 0$, it holds concave properties.

Regarding the similarities to the return level plots of the GEV model, where it was reasonable to look at the N-year returns, that is the case where once every N years it would be exceeded.
Rewriting this to fit the GPD case of the m-observation returns, n_y denotes the amount of observations in a year resulting in: $m = N \times n_y$ - will result in a N-year return level of, given that the case of $\xi \neq 0$:

$$z_n = u + \frac{\sigma}{\xi}\left[(Nn_y\zeta_u)^{\xi} - 1\right] \tag{2.43}$$

else if $\xi = 0$ the N-year return level is, since again the GPD will be an exponential pareto distribution:

$$z_n = u + \sigma \log(Nn_y\zeta_u) \tag{2.44}$$

2.4.1 Maximum likelihood estimation

This subsection references (Coles, S., 2001, Chapter 4.3.3 - Appendix A.1)
When it comes to estimating the return levels, the shape and scale parameters, σ and ξ, the relevant maximum likelihood estimates are substituting the the parameter values. The estimate for ζ_u however, is the probability that a given observation exceeds the chosen threshold.
This comes down to the estimator of a sample proportion of observations exceeding threshold u:

$$\hat{\zeta}_u = \frac{k}{n} \tag{2.45}$$

As the amount of threshold excess observations are following a binomial distribution, finding the maximum likelihood estimate of $\text{Bin}(n, \zeta_u)$ is equivalent with Equation 2.45. Computing the variance-covariance matrix V is now possible since the estimates has replaced the parameter values:

$$\text{V}(\hat{\zeta}_u, \hat{\sigma}, \hat{\xi}) = \begin{bmatrix} \hat{\zeta}_u(1-\hat{\zeta}_u)/n & 0 & 0 \\ 0 & v_{1,1} & v_{1,2} \\ 0 & v_{2,1} & v_{2,2} \end{bmatrix} \tag{2.46}$$

as the binomial properties of ζ_u, the $\text{Var}(\hat{\zeta}_u) \approx \hat{\zeta}_u(1-\hat{\zeta}_u)/n$, and the $v_{i,j}$ just refers to the i, j'th element of the variance-covariance matrix of $(\hat{\sigma}, \hat{\xi})$ - leading to the calculation of the variance of the $\hat{x}_m$ term, by the delta method:

$$\text{Var}(\hat{x}_m) \approx \nabla x_m^T \text{V} \nabla x_m \tag{2.47}$$

with

$$\begin{aligned}\nabla x_m^T &= \left[\frac{\partial x_m}{\partial \zeta_u}, \frac{\partial x_m}{\partial \sigma}, \frac{\partial x_m}{\partial \xi}\right]\Big|_{(\hat{\zeta}_u,\hat{\sigma},\hat{\xi})} \\ &= \left[\sigma m^{\xi}\zeta_u^{\xi-1}, \xi^{-1}\{(m\zeta_u)^{\xi}-1)\}, -\sigma\xi^{-2}\{(m\zeta_u)^{\xi}-1)\} + \sigma\xi^{-1}(m\zeta_u)^{\xi}\log(m\zeta_u)\right]\Big|_{(\hat{\zeta}_u,\hat{\sigma},\hat{\xi})}\end{aligned} \tag{2.48}$$

2.4.2 Profile likelihood

This subsection references (Coles, S., 2001, Chapter 4.3.3 - Appendix A.1).
For parameter estimation of the GPD, as with the GEV distribution of the block maxima method, it is usually more reliable to utilize the profile log-likelihood as described in equation (2.17). As with the block maxima model, a reparameterization is done in order to get the return level as a parameter, and σ is isolated:

$$\sigma = \begin{cases} \frac{(x_m-u)\xi}{(m\zeta_u)^{\xi}-1}, & \text{if } \xi \neq 0 \\ \frac{x_m-u}{\log(m\zeta_u}, & \text{if } \xi = 0 \end{cases} \tag{2.49}$$

As with the block maxima model, this can be used to give a set of confidence intervals for the return level.

With the traditional theories of EVT now explained, along with methods of estimation of the parameters, the next section will cover empirical implementation of these methods and models.

2.5 Data analysis: Maersk & Novo

This part of the section will revolve on implementing the theories described above. The theories of both block maxima and threshold excess will be applied for both equities, and minor conclusions and considerations will be noted as they are appropriate.

All work is based on the theories presented, but have no other reference than the models, as this section is part of the empirical analysis of this thesis.

2.5.1 Block maxima models

For the block maxima models, appropriate sizes of blocks will be tested, followed by distribution estimation, and concluding with parameter estimation by the use of maximum likelihood and profile likelihood.
Both data tables of the equities has been reduced so all blocks has similar size, i.e. incomplete monts of the first and last years has been removed. Therefore the data for both the Novo and the Maersk equities consists of log-returns from 1992 to 2018, and the block maxima for monthly, quarterly and yearly maxima graphically displayed: The

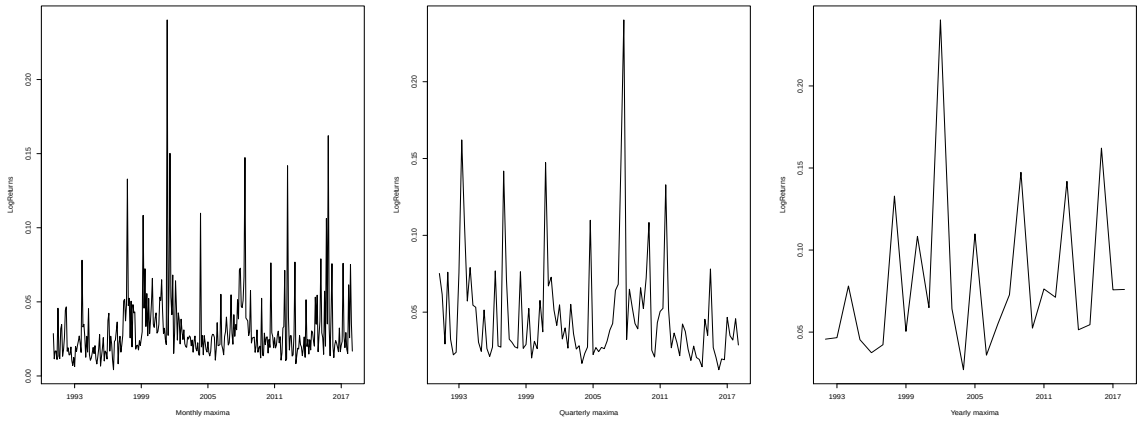

Figure 2: Left: LogReturns - Novo equity monthly maxima, Middle: LogReturns - Novo equity quarterly maxima, Right: LogReturns - Novo equity yearly maxima

considerations made in theory, that block maxima removes a lot of datapoint can be

clearly seen by the granularity of the plots, but it is clear that the quarterly and yearly captures the most distinct maxima. On the monthly basis, only few clusters of maxima can be seen.
For the Maersk equity, similar block maxima are produced:

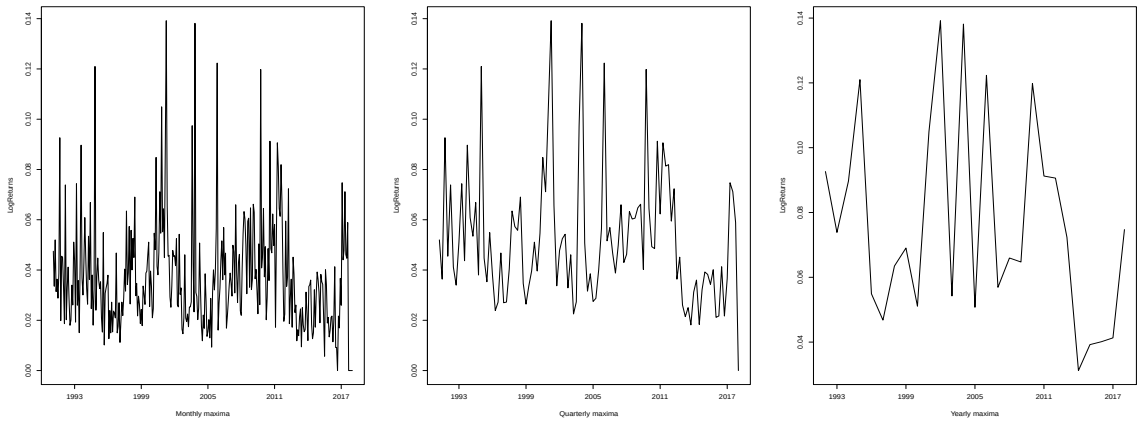

Figure 3: Left: LogReturns - Maersk equity monthly maxima, Middle: LogReturns - Maersk equity quarterly maxima, Right: LogReturns - Maersk equity yearly maxima

The same observations concerning the granularity of the maxima can be spotted in the plots of the Maersk equity maxima. The biggest difference to be seen from these, is that there would seem to be slightly more clustering of maxima for the Maersk equity than the Novo equity. Based on considerations of granularity and having enough observations within each block, the monthly maxima probably contains too few observations to make reasonable assumptions of the distribution of the maxima. The yearly blocks might be too rough in concern to the amount of datapoints, and therefore the quarterly maxima will be the main point of focus for modelling purposes.

2.5.2 GEV distributions

Having settled upon the quarterly maxima for modelling the GEV distribution, the observed points are plotted against the modelled density of the GEV distribution, to

see if the model results in a reasonable fit. For both equities the distributions of the

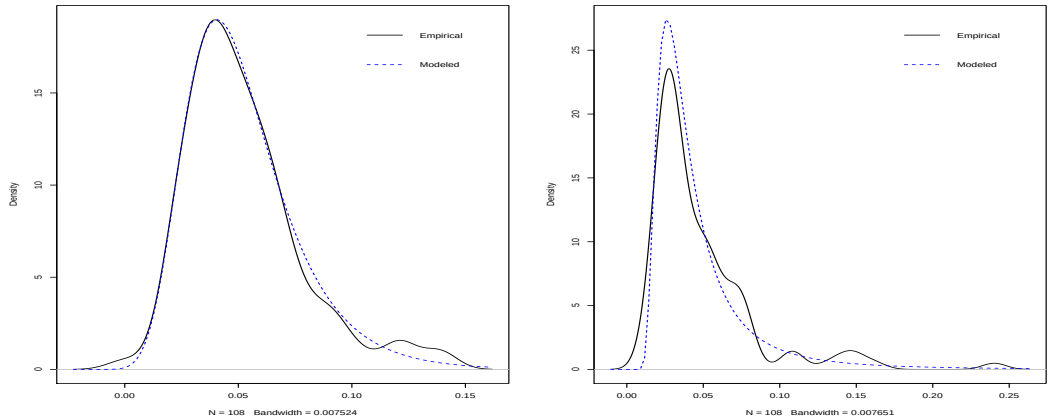

Figure 4: Left: Density empirical vs. modeled - Maersk Quarterly maxima, Right: Density empirical vs. modeled - Novo Quarterly maxima

maxima seem to be a good fit. Comparing the two, the modelled density of the Maersk equity seem to be more heavy-tailed than the corresponding for the Novo equity. Since the both seem to be acceptable models for the distributions of maxima, parameters can be estimated.

2.5.3 Maximum likelihood & Profile likelihood

The parameters will be estimated using the maximum likelihood method.

Parameter estimates	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$
Maersk equity	0.041 (0.002)	0.019 (0.001)	0.012 (0.059)
Maersk 95% conf	(0.0368, 0.0449)	(0.0165, 0.0222)	(-0.1034, 0.128)
Novo equity	0.031 (0.002)	0.014 (0.001)	0.424 (0.089)
Novo 95% conf	(0.0279, 0.0341)	(0.0117, 0.0173)	(0.2495, 0.5976)

Table 1: Parameter estimation by MLE - Quarterly maxima

Since the estimated $\hat{\xi} > -0.5$, as suggested in section (2.2.2), the maximum likelihood estimates are valid estimations for both equities. Since theory suggested that profile likelihood might offer a better estimation, and also provide a graphical assessment of the likelihood, $\hat{\xi}$ is estimated by profile likelihood to compare:

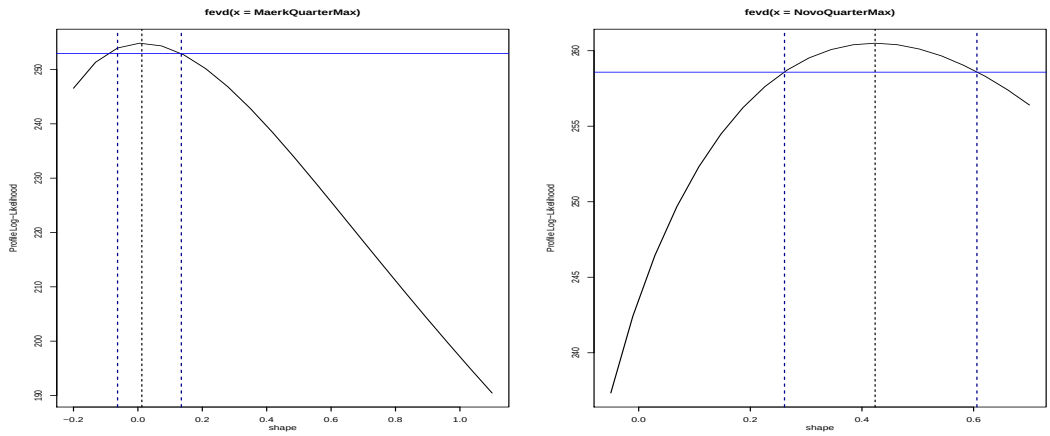

Figure 5: Left: Profile Likelihood for ξ - Maersk Quarterly maxima, Right: Profile Likelihood for ξ - Novo Quarterly maxima

The comparison for the estimates of $\hat{\xi}$ by profile likelihood:

Parameter estimates	$\hat{\xi}$	95% conf
Maersk equity	0.012	(-0.0632, 0.1349)
Novo equity	0.424	(0.3613, 0.5778)

Table 2: Parameter estimation by Profile likelihood - Quarterly maxima

The profile likelihood estimates are similar to the maximum likelihood for $\hat{\xi}$, and therefore maximum likelihood estimates are regarded valid for estimation of the return levels.

2.5.4 Return levels

With all parameters estimated, return levels can be calculated from substituting the estimates into equations (2.10) and (2.47). Since $\hat{\xi} \neq 0$ for both equities:

$$\begin{aligned} \hat{z}_p^{\text{Maersk}} &= \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left[1 - y_p^{-\hat{\xi}}\right] = 0.041 - \frac{0.019}{0.012}\left[1 - y_p^{-0.012}\right] \\ \hat{z}_p^{\text{Novo}} &= \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left[1 - y_p^{-\hat{\xi}}\right] = 0.031 - \frac{0.014}{0.424}\left[1 - y_p^{-0.424}\right] \end{aligned} \tag{2.50}$$

And thus results for different return periods:

$\hat{z}_p$	$p = 0.05$	$p = 0.025$	$p = 0.0125$
Maersk equity	0.116	0.131	0.145
Novo equity	0.168	0.228	0.308

Table 3: Return level estimation by MLE - Quarterly maxima

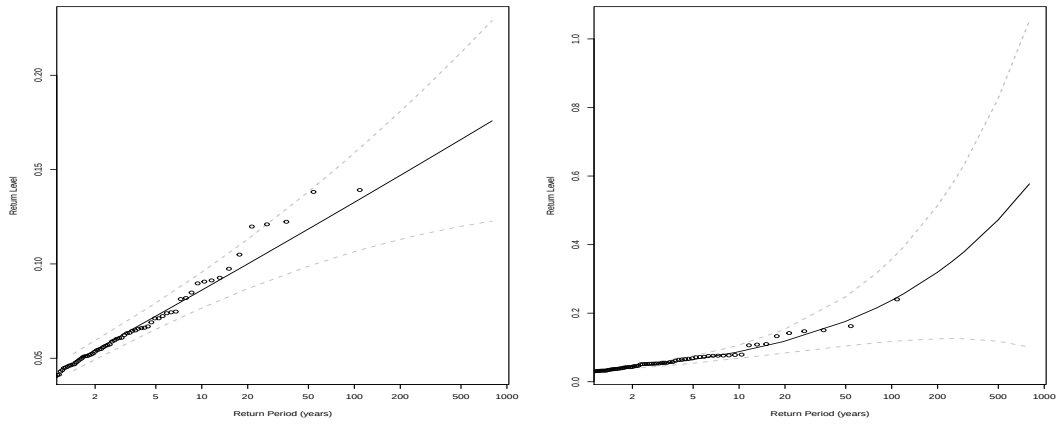

Figure 6: Left: Return level vs. return period - Maersk Quarterly maxima, Right: Return level vs. return period - Novo Quarterly maxima

The dotted lines on figure (6) indicates the 95% confidence interval of the estimation line.

2.5.5 Threshold models

For the threshold models a proper large enough threshold needs to be chosen. By examining the mean residual life plots of the equities, an area of reasonable thresholds can be found, and from there different thresholds can be tried. On the left the Maersk equity and on the right the Novo equity:

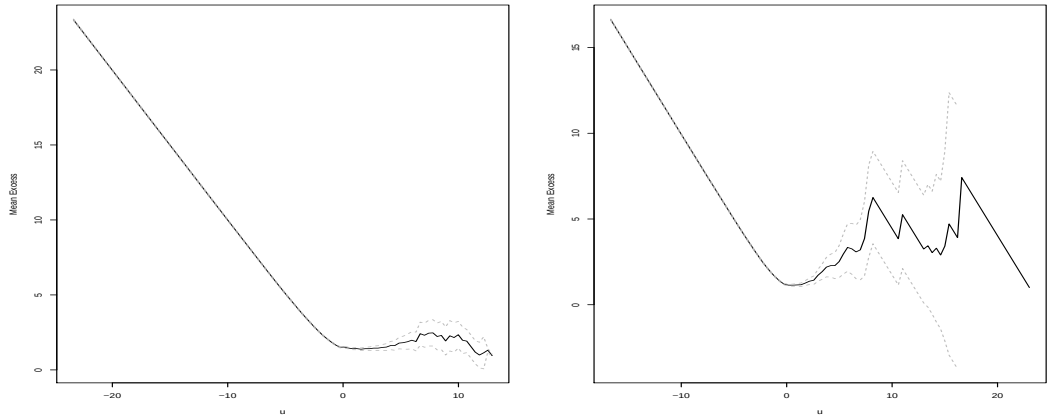

Figure 7: Left: Mean residual life plot - Maersk equity,
Right: Mean residual life plot - Novo equity

For both equities the LogReturns has been re-scaled as $\tilde{X} \rightarrow 100\tilde{X}$, to enable any visual changes to the mean residual life plots. For Maersk, the plot shows curvature between $0 < u < 7$, and for Novo the curvature is between $0 < u < 5$, which suggests thresholds of 0.07 and 0.05 respectively.

2.5.6 Maximum likelihood & Profile likelihood

With thresholds chosen, parameters can be estimated. Calculating $\hat{\zeta}$, by equation (2.45):

$$\hat{\zeta}_u^{\text{Maersk}} = \frac{k}{n} = \frac{26}{6670} = 0.0039$$
$$\hat{\zeta}_u^{\text{Novo}} = \frac{k}{n} = \frac{56}{6757} = 0.0083 \tag{2.51}$$

And the remaining parameters can be estimated by maximum likelihood

Parameter estimates	$\hat{\sigma}$	$\hat{\xi}$
Maersk equity	0.032 (0.011)	-0.336 (0.270)
Novo equity	0.014 (0.003)	0.486 (0.219)

Table 4: Parameter estimation by maximum likelihood - Threshold

And as with the GEV, estimates can be tested by profile likelihood also for the GPD:

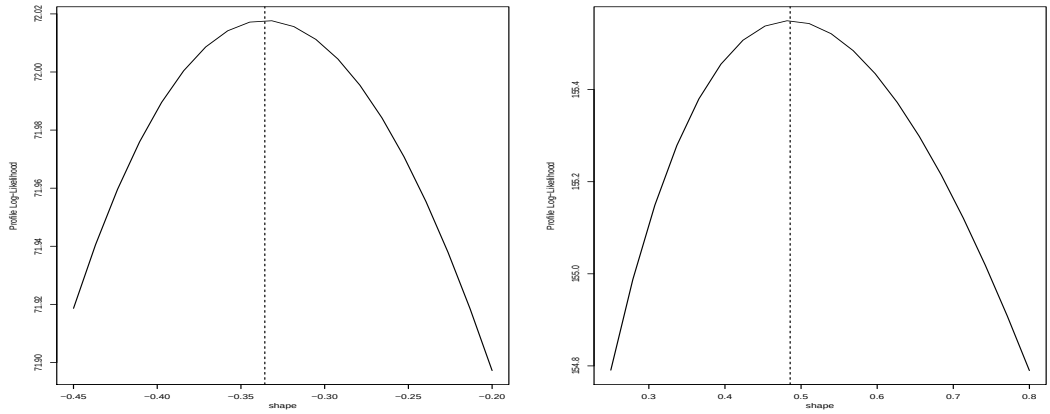

Figure 8: Left: Profile Likelihood for ξ - Maersk GPD,
Right: Profile Likelihood for ξ - Novo GPD

Parameters estimated by the profile likelihood are equivalent to the estimated parameters by maximum likelihood.

Having estimated the parameters of the model, the return levels can be estimated.

2.5.7 Return levels

For the GPDs return levels are calculated, and plots represent return levels on return periods, in m-observations. This can be transformed to a yearly return level, by taking the amount of trading days, and scaling m.

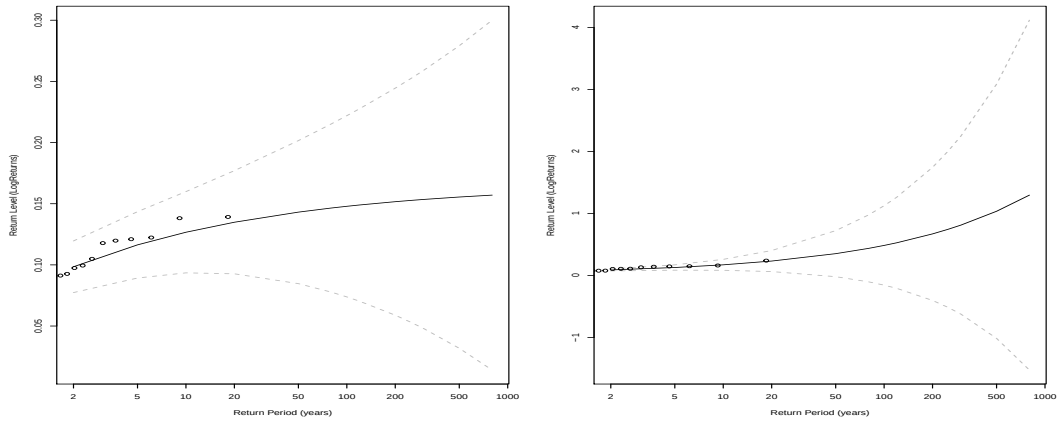

Figure 9: Left: Return level vs. return period - Maersk GPD,
Right: Return level vs. return period - Novo GPD

3 Point Process Characterization

This type of characterization does not, in a statistical sense, offer anything that was unobtainable with the methods introduced in the previous chapter.
This section will only be a theoretical research of point process characterization, hence no empirical implementation will be carried out.
The work in this section is referenced to (Coles, S., 2001, Chapter 7 - Appendix A.1) unless other sources are stated.

The definition of a point process, is a given stochastic principle for the instances and placement of point events. When the principle applies to a set $\mathscr{A}$, describing e.g. a time period, the point process model could be used to outline the amount of instances of a given event.
This leads to being able to calculate the probability of the event, in a period of time - or the compliment, expected time between two instances of an event. Looking at at a sequence of random values $Y_1, ...Y_n$ within $\mathscr{A}$, the positive integer random variable $N(A) \forall A \subset \mathscr{A}$ is:

$$N(A) = \sum_{i=1}^{n} I_{\{Y_i \in A\}} \tag{3.1}$$

Thereby being a counter of points or vectors Y_i in the subset A. The expected amount can be defined for any subset as:

$$\Lambda(A) = \mathrm{E}\{N(A)\} \tag{3.2}$$

Defined as being the intensity measure related to the given process.
If the parameter λ is constant, this process is defined as being homogeneous. However if the parameter lambda is variable, that is a dependent of time, the process is defined as being non-homogeneous.

From this the intensity function of the process can be derived, under the assumption that $A = [a_1, x_1] * ... * [a_k, x_k] \subset \mathbb{R}^k$:

$$\lambda(\mathbf{x}) = \frac{\partial \Lambda(A)}{\partial x_1 ... \partial x_k} \Longleftrightarrow \Lambda(A) = \int_A \lambda(\mathbf{x}) d\mathbf{x} \tag{3.3}$$

3.1 Poisson Process

The process is a Poisson point process, for $\mathscr{A} \subset \mathbb{R}^k$, if and only if two conditions are met:
- (a): For $A \subset \mathscr{A}$, and $k \geq 0$,

$$\Pr\{N(A) = k\} = \begin{cases} e^{-\Lambda(A)} \frac{\Lambda(A)^k}{k!}, & \Lambda(A) < \infty, \\ 0, & \Lambda(A) = \infty \end{cases} \tag{3.4}$$

This can be multidimensional, and can be both homogeneous - and non-homogeneous Poisson process.
- (b): For $m \geq 1$, if $A_1, ..., A_m$ are non-overlapping subsets of $\mathscr{A}$, then the stochastic variables $N(A_1), ..., N(A_m)$ are independent.

So this shows that, the amount of points within a chosen interval are distributed by a Poisson distribution, thus being mutually independent to the amount of points within disjoint intervals, and the mean value is dependent on the length of the interval.
The independent occurrence of points within the Poisson process is probably the most interesting property, as it perfectly aligns with the modelling of white noise, e.g. in financial loss. Though the variable amount of points in different subsets of $\mathscr{A}$ might not be favorable, this is consequently due to a larger average amount of points in some subsets than others, by construct not as a result of points influencing the placement of other points, but rather due to external/natural noise. This makes it, that the Poisson process faces challenges when looking at a subsets of $\mathscr{A}$ where a clustering of points is natural, which could be e.g. in a recession period, where financial loss is expected to be occurring more frequently.

3.1.1 Statistical application

Initially observing a set of points $x_1, ..., x_n$ within set $\mathscr{A}$. Under the assumption that the intensity function $\lambda(\cdot)$ belongs to the parametric family of models, $\lambda(\cdot;\theta)$, which leaves the greatest challenge, with regards to the validity of the model, to estimate the θ-vector. Observing the case of the set $\mathscr{A}$ being one-dimensional with $\mathscr{A} \subset \mathbb{R}$, and points $x_1, ..., x_n$ being the outcome of a Poisson process on $\mathscr{A}$, with intensity $\lambda(\cdot;\theta)$ for a given value of θ.

Denoting $I_i = [x_i, x_i + \delta_i]$ as a set of limited sized intervals surrounding the observations, and $\mathfrak{I} = \mathscr{A} \setminus \bigcup_{i=1}^{n} I_i$. Hence, Poisson properties gives:

$$\Pr\{N(I_i) = 1\} = \exp\{-\Lambda(I_i;\theta)\}\Lambda(I_i;\theta) \tag{3.5}$$

wherein

$$\Lambda(I_i;\theta) = \int_{x_i}^{x_i+\delta_i} \lambda(u)\,du \approx \lambda(x_i)\delta_i \tag{3.6}$$

by substituting equation (3.6) into equation (3.5):

$$\Pr\{N(I_i) = 1\} \approx \exp\{-\lambda(x_i)\delta_i\}\lambda(x_i)\delta_i \approx \lambda(x_i)\delta_i \tag{3.7}$$

since for small values of $\delta_i \Rightarrow \exp\{-\lambda(x_i)\delta_i\} \approx 1$, leading to:

$$\Pr\{N(\mathfrak{I}) = 0\} = \exp\{-\Lambda(\mathfrak{I})\} \approx \exp\{-\Lambda(\mathscr{A})\} \tag{3.8}$$

as all δ_i are small values. This results in the likelihood:

$$\begin{aligned} L(\theta; x_1, ..., x_n) &= \Pr\{N(\mathfrak{I}) = 0, N(I_i) = 1, i = 1, ..., n\} \\ &= \Pr\{N(\mathfrak{I}) = 0\} \prod_{i=1}^{n} \Pr\{N(I_i = 1\} \\ &\approx \exp\{-\Lambda(\mathscr{A};\theta)\} \prod_{i=1}^{n} \lambda(x_i;\theta)\delta_i \end{aligned}$$

Resulting in the intensity, by division of δ_i:

$$L(\theta; x_1, ..., x_n) = \exp\{-\Lambda(\mathscr{A}; \theta)\} \prod_{i=1}^{n} \lambda(x_i; \theta) \tag{3.9}$$

where, as defined in (3.3):

$$\Lambda(\mathscr{A}; \theta) = \int_{\mathscr{A}} \lambda(x; \theta) dx \tag{3.10}$$

Taking this into a specific representation, where $x_1, ..., x_n$ being points of a one-dimensional Poisson process, are observed at an interval $\mathscr{A} = [0, T]$, and the intensity parameter λ is unknown:

$$\Lambda(\mathscr{A}; \lambda) = \lambda T$$

Resulting in the likelihood function to be:

$$L(\lambda; x_1, ..., x_n) = \exp\{-\lambda T\} \lambda^n$$

Using the likelihood estimator, and maximizing this:

$$\ell(\lambda) = \log L(\lambda; x_1, ..., x_n) = -\lambda T + n \log \lambda$$

Thus giving the maximum likelihood estimate, describing the rate of which points in the one-dimensional Poisson process are occurring:

$$\hat{\lambda} = \frac{n}{T}$$

If the Poisson process is non-homogeneous, this is not straightforward and potentially very difficult to find.

3.1.2 Convergence

Defining $N_1, N_2, ...$ as point processes on the set $\mathscr{A}$. The sequence is said to converge in distribution to N, denoted:

$$N_n \xrightarrow{d} N,$$

if

$$\lim_{n\to\infty} F_{N_n}(x) = F_N(x) \forall\, x$$

Using the same notation as in section (2), with $M_n = \max\{X_1, ..., X_n\}$, and assuming that $X_1, ..., X_n$ are iid, with a mutual distribution F.
By equation (2.2), there will be a sequence of constants $\{a_n > 0\}$ and $\{b_n\}$ resulting in:

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \to G(z) \tag{3.11}$$

and G being the GEV distribution function such that:

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{3.12}$$

For some parameters $\sigma, \mu > 0$ and ξ. An array of point processes N_n on $\mathbb{R}^2$ such that

$$N_n = \left\{\frac{i}{n+1}, \frac{X_i - b_n}{a_n} : i = 1, ..., n\right\} \tag{3.13}$$

is the definition of the different iid point processes.
By construction this constrains the time aspect, $\frac{i}{n+1}$, so it will always stay within the region of $[0, 1]$, for any $n \to \infty$.
Each point $n \in N_n$ has a certain probability of being in the set of $A = [0, 1] \times (u, \infty)$, for a

large chosen threshold u of:

$$p_A = \Pr\left\{\frac{X_i - b_n}{a_n} > u\right\} \approx \frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}$$

By assumption of the array of point processes being iid, it follows that

$$N_n(A) \sim \text{Bin}(n, p_A) \tag{3.14}$$

Being distributed binomial, the asymptotic distribution of $N_n(A)$ for $n \to \infty$ is a Poisson limit, with intensity measure given in equation (3.6), thus being $\text{Poi}(\Lambda(A))$, where:

$$\Lambda(A) = \left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}$$

Since it has already been established that there is homogeneity (TJEK AT DER ER KOMMET MED LÆNGERE OPPE) with regards to the time factor, this Poisson limit can be generalized for any set of A, with the construction of $A = [t_1, t_2] \times (u, \infty)\ \forall [t_1, t_2] \subset [0, 1]$:

$$\Lambda(A) = (t_2 - t_1)\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi} \tag{3.15}$$

This is shown to have a connection to the GEV theories described earlier, an example of this is case where $X_1, ..., X_n$ is an array of independent exponential variables, and the distribution G is the Gumbel distribution from section (2), where the limits of the distribution is $(z_-, z_+) = (-\infty, \infty)$, and with respect to the Gumbel distribution $\mu = 0$, $\sigma = 1$ for $\xi \to 0$, such that equation (3.13) $\Rightarrow$

$$N_n = \left\{\frac{i}{n+1}\ ,\ (X_i - n) : i = 1, ..., n\right\}$$

Thus converging with intensity measure $\Lambda(A)$, towards a Poisson process, with the set $A = [t_1, t_2] \times (z, \infty)$ for $z > \infty$:

$$\Lambda(A) = (t_2 - t_1) \exp(-z)$$

Another example is the case where $X_1, ..., X_n$ is an array of independent variables, which are distributed uniformly $\mathrm{U}(0,1)$, with the limiting distribution $G(z) = \exp(z)$, for $z < 0$, resulting in the limits of the distribution $(z_-, z_+) = (-\infty, 0)$, thus the point processes are

$$N_n = N_n = \left\{ \frac{i}{n+1} \, , \, n(X_i - 1) : i = 1, ..., n \right\}$$

Thus converging with intensity measure $\Lambda(A)$, towards a Poisson process, with the set $A = [t_1, t_2] \times (z, 0)$ for $z < 0$:

$$\Lambda(A) = (t_2 - t_1) \, (-z)$$

The Poisson point process offers an alternative to threshold models, with what is called Peaks-Over-Threshold (POT). POT models look at extremes as one would in queue theory, and the interval between extreme events are to be seen as waiting time.
This chapter served only to put the theory into a broader perspective, and the link to the multivariate theory.

4 Multivariate Extremes

Following the threshold and point process approaches of the previous chapters, this chapter will focus on multivariate models and methods. Having more than one series of extreme points gathered in one distribution.
The properties of the multivariate distributions can be estimated in a similar fashion to the other topics.
The focus of this thesis is economic theory, and therefore the multivariate sequence of vectors $(X_1, Y_1), ..., (X_n, Y_n)$, is to be seen as pairing of two stocks, and their given log returns, as explained in the introduction. Were one to consider $\{X_i\}$ and $\{Y_i\}$ separately, one would be observing two independent random sequences.

The multivariate theories revolve around the use of **copulas**, a topic that is interesting, but might not be accurate with regards to financial theory. For now it will be accepted as plausible, but will be touched upon further in the discussion part, with regards to the article (Mikosch, T - 2005, Copulas: Tales and Facts). The work in this section is referenced to (Coles, S., 2001, Chapter 8 - Appendix A.1) unless other sources are stated.

4.1 Componentwise Maxima

Using the multivariate sequence of vectors, a redefinition of M_n is required, and this chapter will now define the componentwise maxima vector as follows.
If:

$$M_{x,n} = \max_{i=1,...,n} \{X_i\}\ ,\ M_{y,n} = \max_{i=1,...,n} \{Y_i\} \tag{4.1}$$

the componentwise maxima is defined as:

$$\boldsymbol{M}_n = \left(M_{x,n}, M_{y,n}\right) \tag{4.2}$$

It is paramount to state that the index i does not need to match, that is maxima in one half of the sequence e.g. X_i with $i = 1$, does not mean that Y_i with $i = 1$ need to be maxima.

The distributions of the two different sequences need to have a common underlying distribution. For the purpose of this theoretical example, the Fréchet distribution is assumed, as this is the easiest to formulate.
Let:

$$F(z) = \exp\left(\frac{-1}{z}\right),\ z > 0 \tag{4.3}$$

From section (2), this being the instance, in which the GEV distribution takes parameter values of: $\xi = 1$, $\sigma = 1$, $\mu = 0$, resulting in:

$$\Pr\left\{\frac{M_n}{n} \leq z\right\} = \exp\left(\frac{-1}{z}\right),\ z > 0 \tag{4.4}$$

Rescaling the componentwise vector:

$$\boldsymbol{M}_n^* = \left(\frac{M_{x,n}}{n}, \frac{M_{y,n}}{n}\right) = \left(M_{x,n}^*, M_{y,n}^*\right) \tag{4.5}$$

Gives the marginal result.

Now the limiting distribution, of the componentwise maxima $\boldsymbol{M}_n^*$ can be inferred from equation (2.8). In the instance of:

$$\Pr\left\{M_{x,n}^* \leq x, M_{y,n}^* \leq y,\right\} \xrightarrow{d} G(x, y) \tag{4.6}$$

with G being a non-degenerate distribution function. Then the distribution function G, is of the form:

$$G(x,y) = \exp\{-V(x,y)\}, x > 0, y > 0 \tag{4.7}$$

with $V(x,y)$ as

$$V(x,y) = 2\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) dH(w), \tag{4.8}$$

and H is a distribution function on the set $[0,1]$, with a constraint of the mean such that

$$\int_0^1 w dH(w) = \frac{1}{2} \tag{4.9}$$

In the case where H is a differentiable function, equation (4.8) can be written as:

$$V(x,y) = 2\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) h(w) dw, \tag{4.10}$$

given the density of $H = h$.

If H is not differentiable, but meets the constraint of the mean, equation (4.9), for value of w:

$$V(x,y) = x^{-1} + y^{-1}, \tag{4.11}$$

this holds where:

$$h(w) = \begin{cases} \frac{1}{2} & w = 0, \\ \frac{1}{2} & w = 1, \\ 0 & \text{for other values} \end{cases} \tag{4.12}$$

Thus G in accordance with equations (4.7) and (4.11), the distribution function, in this case is:

$$G(x,y) = \exp\{-(x^{-1} + y^{-1})\}, \tag{4.13}$$

with $x > 0$ and $y > 0$. This corresponds to the case with independent marginal distributions.

This family of distributions are named the bivariate extreme value distributions. Generalizing these results, the entire class of bivariate extreme value distributions can be described in a manner similar to the GEV distributions:

$$\tilde{x} = \left[1 + \xi_x \left(\frac{x - \mu_x}{\sigma_x}\right)\right]^{1/\xi_x} \tilde{y} = \left[1 + \xi_y \left(\frac{y - \mu_y}{\sigma_y}\right)\right]^{1/\xi_y}$$
$$\Rightarrow G(x,y) = \exp\{-V(\tilde{x},\tilde{y})\} \tag{4.14}$$

following usual assumptions of the distribution functions, that $\left[1 + \frac{\xi_x(x-\mu_x)}{\sigma_x}\right] > 0$ and likewise $\left[1 + \frac{\xi_y(y-\mu_y)}{\sigma_y}\right] > 0$ making both parts of the distributions GEV, with (ξ_x, σ_x, μ_x) and (ξ_y, σ_y, μ_y) individually, and V is satisfying equation (4.6).
One of the nice properties of the GEV, was the max-stability implied on the distributions, ensuring stability of the excess.
In a similar fashion, it can be implied on the multivariate maxima, from equation (4.6). Taking for any given fixed $a > 0$:

$$V(a^{-1}x, a^{-1}y) = aV(x,y) \tag{4.15}$$

Thus V is homogeneous of order -1. Using this along with equation (4.7), it is inferred that:

$$G^n(x,y) = G(n^{-1}x, n^{-1}y), n = 2,3,...,\infty \tag{4.16}$$

Therefore it is resolves that, if (X, Y) is distributed by a given function G, thus M_n must also be distributed by function G (except the rescaling by n^{-1}).
This all gives a nice rationale to how the bivariate distributions limit off the multivariate distributions, alas the feasible region of limits are not quite limited to these.

Looking over this feasible region, any distribution function H adhering to the mean constraint in equation (4.9), and defined on the region $[0, 1]$, will be a satisfying limiting distribution to equation (4.6).
Since this results in a non-finite parameterization within the feasible limiting distributions, estimating parameters within these distribution function might be difficult.
A prominent solution to this, is to utilize certain sub-families of the limiting distributions H, thus resulting in finding sub-families of G - using sub-families that allow for parameterization. By choice of these sub-families, it is possible to approximate the entire family based on one member.

An example is a logistic family:

$$G(x, y) = \exp\left\{-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}\right\} \qquad x > 0, y > 0 \tag{4.17}$$

Defined where $\alpha \in (0, 1)$. To derive equation (4.17), the density of H, in equation (4.8), assuming $0 < w < 1$, is defined as:

$$h(w) = \frac{1}{2}(\alpha^{-1} - 1)\{w(1-w)\}^{-1-1/\alpha}\{w^{-1/\alpha} + (1-w)^{-1/\alpha}\}^{\alpha-2} \tag{4.18}$$

Due to symmetry of w, equation (4.12), the mean constraint of equation (4.9) is fulfilled.

One of the primary useful properties of the logistic distribution family, is the adaptable nature, since as $\alpha \to 1$, representing variables that are perfectly independent:

$$(4.17) \Rightarrow G(x,y) \to \exp\{-(x^{-1}+y^{-1})\}$$

and the reversed case of dependant variables, with $\alpha \to 0$

$$(4.17) \Rightarrow G(x,y) \to \exp\{-\max(x^{-1},y^{-1})\}$$

thus the logistic family is able to approximate every layer of the sub-family of H, with varying levels of dependency from one extreme to the other.

Having described the nature of the componentwise maxima, next part is to look further into the structure of these.

4.1.1 Structure variables

In order to have an easier way of describing the return levels of the multivariate distributions, a term Z is implied. Z is derived from the relationship between M_x and M_y, being able to take forms of either $\max\{M_x, M_y\}$, $\min\{M_x, M_y\}$ alternatively $M_x + M_y$. This results in the notation of Z:

$$Z = \phi(M_x, M_y)$$

with the distribution function of Z being:

$$\Pr\{Z \le z\} = \int_{A_z} g(x,y)\,dx\,dy \tag{4.19}$$

with g being the probability density function of (M_x, M_y).

Circling back to the N-year return level, this can be calculated for the structure variable Z, by solving:

$$G_Z(z) = 1 - \frac{1}{N}. \tag{4.20}$$

This has some major potential complications, as the distribution function is only available in its integral formulation, equation (4.19). The most straight forward way of solving this, and getting estimated return levels of Z, is to incorporate univariate methods.

Observing the vector of componentwise maxima $(m_{x,1}, m_{y,1}), ..., (m_{x,k}, m_{y,k})$, a vector $z_1, ..., z_k$ is the structure variable according to the given observations, thus $z_i = \phi(m_{x,i}, m_{y,i})$. Assuming that z_i are distributed by a GEV distribution, univariate methods are now able to be imposed, as the observations are now of a univariate vector, rather than the multivariate one.
Having simplified this, one must also take into consideration that new challenges may arise. In particular the assumptions concerning the GEV distribution of this vector are relatively weak, and also this methods requires that data for all underlying variable are available at the same time, thus invalidating this method if there are empty datapoints. This could arise if one equity is traded on a given day, but the other is not. This further strengthens the introductory parameter, that large firms were to be chosen, as these are assumed to be traded more often.

For now one of the essential models have been expanded from a univariate plane to a multivariate one. The next subsection will expand another theoretical model to the multivariate plane.

4.1.2 Multivariate point process

The first to be considered, is the point process model.
Expanding the univariate representation, observing $(x_1, y_1), (x_2, y_2)...$ as a vector of independent bivariate points, assumed to be distributed by a function adhering to Fréchet margins in section (3), and satisfying the same convergence as the multivariate block maxima models:

$$\Pr\left\{M^*_{x,n} \leq x, M^*_{y,n} \leq y\right\} \rightarrow G(x, y)$$

Noting $\{N_n\}$ as a sequence of point processes, thus:

$$N_n = \{(n^{-1}x_1, n^{-1}y_1), ..., (n^{-1}x_n, n^{-1}y_n)\} \tag{4.21}$$

It applies, as in the univariate case, for

$$N_n \xrightarrow{d} N$$

are lower bound in $(0, 0)$, as N is a non-homogeneous Poisson process defined in $(0, \infty) \times (0, \infty)$. This being the same definition as the univariate case, but on a plane rather than one dimensional. The intensity function with respect to N can be shown to be:

$$\lambda(r, w) = 2\frac{dH(w)}{r^2} \tag{4.22}$$

where

$$r = x + y, w = \frac{x}{r} \tag{4.23}$$

To understand this, it was found helpful to look at this the same way as one can derive complex numbers from angles of the unit circle.
Like calculating complex numbers by use of the unit circle, this method is changing

coordinates to distance to origin, by r, and the angle, by $w \in [0,1]$.
Assuming H can be differentiated, and having density h, $h()$ is a measure of the frequency of the events, as w measures the proportionate size of (x, y).

Depending on the state of independence between extreme observations, a mostly independent relationship would assume small y/n values if x/n are is large, and the other way around. This would result in $h(w)$ taking a large value when w is close to the outer cases, as $w = 1$ or $w = 0$, and thus taking small values in between. In the opposite independence state, being close where extremes are near dependent, resulting in $x/n \approx y/n$, $h(w)$ would be largest around $w = 0.5$.

The limiting distribution in equation (4.6) can now be derived for the point process model:

$$\Pr\left\{M^*_{x,n} \leq x, M^*_{y,n} \leq y\right\} = \Pr\{N_n(A) = 0\} \tag{4.24}$$

With A being the multivariate interpretation of the univariate case:

$$A = \left\{\frac{(0,\infty) \times (0,\infty)\}}{\{(0,x) \times (0,y)\}}\right\} \tag{4.25}$$

This leads to

$$\Pr\{N_n(A) = 0\} = \exp\{-\Lambda(A)\} \tag{4.26}$$

where $\Lambda(A)$ is:

$$\begin{aligned}\Lambda(A) &= \int_A 2\frac{dr}{r^2}dH(w) \\ &= \int_{w=0}^{1}\int_{r=\min\{x/w,\,y/(1-w)\}}^{\infty} 2\frac{dr}{r^2}dH(w) \\ &= 2\int_{w=0}^{1}\max\left(\frac{w}{x},\frac{1-w}{y}\right)dH(w)\end{aligned} \tag{4.27}$$

bounded from $(0,0)$.

When implementing this procedure, it is under the assumption that N_n is fairly well approximated by a Poisson limit, of a given region, and convergence of this is secure with respect to regions bound away from $(0,0)$.

Choosing an appropriate region that allows for smooth calculations have its benefits, and the easiest regions are when $A = \{(x,y) : x/n + y/n \geq r_0\}$, for some large value of r_0, as:

$$\Lambda(A) = 2\int_A \frac{dr}{r^2} dH(w) = 2\int_{r=r_0}^{\infty} \frac{dr}{r^2} \int_{w=0}^{1} dH(w) = \frac{2}{r_0} \tag{4.28}$$

resulting in $\Lambda(A)$ not being dependent of the parameters of H, instead being constant. Furthermore this will ease the likelihood calculation, still assuming h being the density of H:

$$\begin{aligned} L(\theta; (x_1,y_1),...,(x_n,y_n)) &= \exp\{-\Lambda(A)\} \prod_{i=1}^{N_A} \lambda\left(\frac{x_{(i)}}{n}, \frac{y_{(i)}}{n}\right) \\ &\propto \prod_{i=1}^{N_A} h(w_i) \end{aligned} \tag{4.29}$$

denoting $w_i = x_{(i)}/(x_{(i)} + y_{(i)})$, for points $(x_{(i)}, y_{(i)})$ of $N_A \in A$.

As with the multivariate block maxima, the multivariate point process will see implementation at the end of this chapter.

The final installment of the multivariate expansions of the models is the bivariate threshold excess, with examples being the Fréchet distribution.

4.1.3 Bivariate threshold excess

In the univariate case of the threshold excess model, the family of distributions of which the tail was approximated was defined as:

$$G(x) = 1 - \zeta\left\{1 + \frac{\xi(x-u)}{\sigma}\right\}^{-1/\xi} \qquad , x > u \tag{4.30}$$

for a chosen threshold u. The goal being to approximate a distribution such that $F(x) \approx G(x)$, instead in the bivariate case the goal is to approximate the combined distribution $F(x,y)$, for $x > u_x, y > u_y$, for a dual set of model parameters $(\zeta_x, \sigma_x, \xi_x), (\zeta_y, \sigma_y, \xi_y)$.
Say (X,Y) has the joint distribution F, are transformed such that $(\tilde{X}, \tilde{Y})$ is distributed by $\tilde{F}$ having boundaries that are standard Fréchet, for some suitable choice of u_x, u_y, thus:

$$\tilde{X} = -\left(\log\left\{1 - \zeta_x\left[1 + \frac{\xi_x(X-u_x)}{\sigma_x}\right]^{-1/\xi_x}\right\}\right)^{-1} \tag{4.31}$$

$$\tilde{Y} = -\left(\log\left\{1 - \zeta_y\left[1 + \frac{\xi_y(Y-u_y)}{\sigma_y}\right]^{-1/\xi_y}\right\}\right)^{-1} \tag{4.32}$$

Following the results of equation (4.7), it is for a suitable large value of n known that:

$$\begin{aligned}\tilde{F}(\tilde{x},\tilde{y}) &= \left\{\tilde{F}^n(\tilde{x},\tilde{y})\right\}^{1/n} \\ &\approx \left[\exp\left\{-V\left(\frac{\tilde{x}}{n},\frac{\tilde{y}}{n}\right)\right\}\right]^{1/n} \\ &= \exp\{-V(\tilde{x},\tilde{y})\}\end{aligned} \tag{4.33}$$

Since $\tilde{F}$ is but a transformation of F, these must be equal, thus following equation (4.8):

$$F(x,y) \approx G(x,y) = \exp\{-V(\tilde{x},\tilde{y})\} \qquad \forall\, x > u_x, y > u_y \tag{4.34}$$

whereas $\tilde{x}$ and $\tilde{y}$ are the representation of x and y defined as their given transformations of equations (4.31) and (4.32).
A complication in the light of the bivariate expansion of the univariate case, arises as the interpretation of the results can be hindered as there is definitely the possibility of a single of the components, either x_i or y_i that exceeds the chosen threshold, but not the

other. An example of this would be:

$$\begin{aligned} R_{0,0} &=]-\infty, u_x[\times]-\infty, u_y[& R_{1,0} &= [u_x, \infty[\times]-\infty, u_y[\\ R_{0,1} &=]-\infty, u_x[\times[u_y, \infty[& R_{1,1} &= [u_x, \infty[\times[u_y, \infty[\end{aligned} \tag{4.35}$$

Where the matrix R, equation (4.35), describes the relationship between maxima, so $R_{0,0}$ being where neither are above the threshold, and $R_{1,1}$ are both above threshold. The two cases where either one or the other is above are in $R_{0,1}$ or $R_{1,0}$, here the density of $\tilde{F}$ is invalid, as the restrictions are not valid, so the partial likelihood needs to be calculated by another form:

$$\Pr\{X = x, Y \leq u_y\} = \frac{\partial F}{\partial x}|_{(x,u_y)} \tag{4.36}$$

Generalizing this to all possible outcome regions across the data, the total likelihood is formulated as:

$$L(\theta; (x_1, y_1), ..., (x_n, y_n)) = \prod_{i=1}^{n} \psi(\theta; (x_i, y_i)) \tag{4.37}$$

$$\psi(\theta; (x_i, y_i)) = \begin{cases} F(u_x, u_y) & \text{if}(x, y) \in R_{0,0} \\ \frac{\partial F}{\partial x}|_{(x,u_y)} & \text{if}(x, y) \in R_{1,0} \\ \frac{\partial F}{\partial y}|_{(u_x,y)} & \text{if}(x, y) \in R_{0,1} \\ \frac{\partial^2 F}{\partial x \partial y}|_{(x,y)} & \text{if}(x, y) \in R_{1,1} \end{cases} \tag{4.38}$$

Thereby allowing for maximization of the likelihood function as per usual.

4.1.4 Asymptotic independence

One of the greater challenges with using the theories in practice, is that there is assumed across the broad spectrum to be independence between maxima. Looking at a bivariate case of a vector of random variables, where $\rho \in [0, 1]$ denotes the correlation between

observed variables x and y, values of $\rho < 1$ will cause the limiting distribution of equation (4.7) to be valid, as observations are independent.
Where ρ is adjacent or equal to 1 though, variables are most likely showing severe signs of dependency, thus modelling these would lead to overestimation of the dependency with extrapolation. In order to explore this, a new variable is defined:

$$\chi = \lim_{z \to z_+} \Pr\{Y > z | X > z\} \tag{4.39}$$

denoting z_+ as being the end-point of the distribution function F, thereby χ is a measure of, given one variable is large, that also the other variable be large.
The extremal case where $\chi = 0$ is problematic, as here X and Y will be asymptotically independent.
Denoting F_X as the marginal distribution of X and F_Y as the marginal distribution of Y, then define χ as:

$$\chi = \lim_{u \to 1} \Pr\{F_Y(Y) > u | F_X(X) > u\} \tag{4.40}$$

whilst in this case, for $0 < u < 1$:

$$\begin{aligned} \chi(u) &= 2 - \frac{\log \Pr\{F_X(X) < u, F_Y(Y) < u\}}{\log \Pr\{F_X(X) < u\}} \\ &= 2 - \frac{\log \Pr\{F_X(X) < u, F_Y(Y) < u\}}{\log u} \end{aligned} \tag{4.41}$$

then it can be shown that,

$$\chi = \lim_{u \to 1} \chi(u) \tag{4.42}$$

In extension of that, it is also not too difficult to deduce, that uniformly for u, the bivariate distribution, being distributed by $G(x, y) = \exp\{-V(x, y)\}$:

$$\chi(u) = 2 - V(1, 1) \tag{4.43}$$

Hence, χ is a measure that provides some clarity on the dependency between variables, when distributions are displaying asymptotically dependency.

When the distributions however, are asymptotically independent, χ fails to deliver, and a substitute representation of χ is needed. Still keeping $0 < u < 1$, denoting:

$$\begin{aligned}\bar{\chi} &= \frac{2\log\Pr\{F_X(X) > u\}}{\log\Pr\{F_X(X) > u, F_Y(Y) > u\}} - 1 \\ &= \frac{2\log(1-u)}{\log\Pr\{F_X(X) > u, F_Y(Y) > u\}} - 1\end{aligned} \tag{4.44}$$

thus,

$$\bar{\chi} = \lim_{u \to 1} \bar{\chi}(u) \tag{4.45}$$

Having these two measured established, there now exists a reasonable way to estimate how the dependency structure is between variables.
Using the two measures χ and $\bar{\chi}$ in conjunction with one another, allows for a measurement of the underlying dependency. When variables are asymptotically independent $\bar{\chi} < 1$, and thus $\chi = 0$. For the inverted case, variables that are asymptotically dependent $\bar{\chi} = 1$.
Measuring the strength of the dependency calculations of χ is the preferred in the dependant case, and $\bar{\chi}$ preferred in the independent case.

Implementing this is not straightforward for any of the multivariate models, as the block maxima approach leads to a significantly smaller sample set, that in its nature would cause the application to be very insecure. But on the other hand, both threshold modelling and point process approaches are more populated, and have enough data-points to make valid estimations. Usage of χ and $\bar{\chi}$ are reasonable ways of validating choice of model in regards to the data, and equally, choice of threshold in the threshold models.
Empirical implementation and analysis will be carried out as part of the next section.

4.2 Data analysis: Maersk & Novo

Implementation of the multivariate theories proved to be very challenging. The **R** libraries that offer functions for these types of distributions were unable to provide satisfactory results.
The analysis part is therefore reduced to include some calculations of the bivariate GPD, and an overview of the distribution of data.

Looking at the bivariate GPD, the thresholds from the univariate case are re-used, and to get an overview of what the probability of extreme cases happening at the same time. The graphical overview of the data: The x-axis represents the Novo equity, and the y-axis represents the Maersk equity.

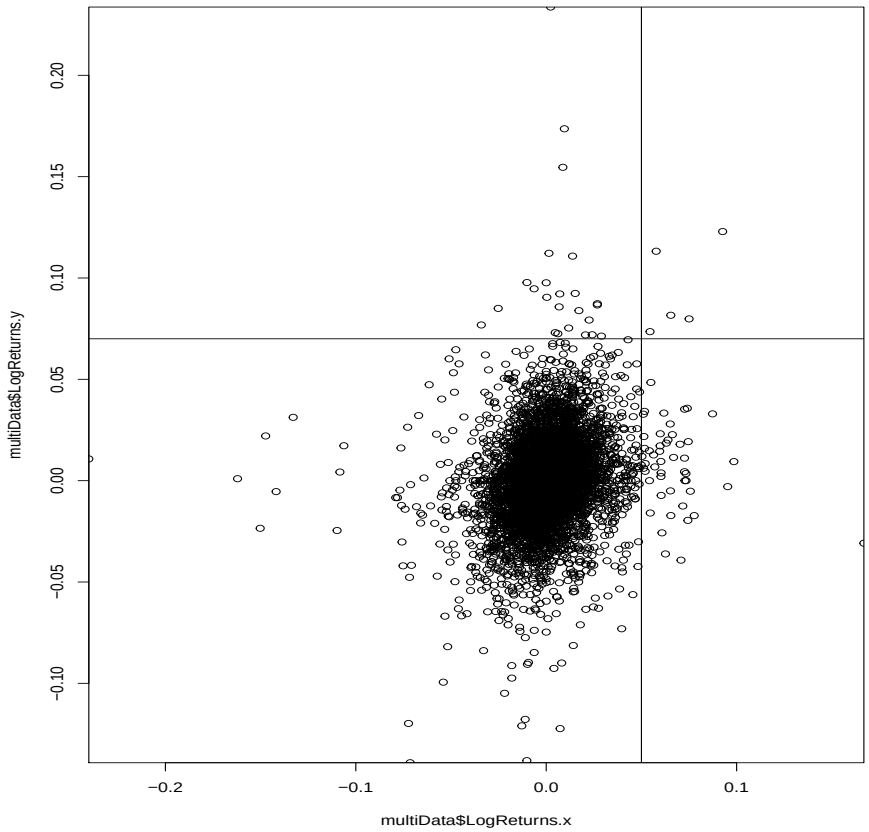

Figure 10: LogReturns with exceedances

It can be seen from figure (10), that only a few of the observations are extreme with regards to both equities. To look further into this, probabilities for the bivariate GPD is calculated, these probabilies are based on observations exceeding the thresholds of log returns plotted:

- Novo: 0.05 log return threshold = $\hat{x}$
- Maersk: 0.07 log return threshold = $\hat{y}$

Marginal exceedances	$\hat{x}$ and $\hat{y}$	$\hat{x} = 0.06$ $\hat{y} = 0.08$	$\hat{x} = 0.07$ $\hat{y} = 0.09$
Novo exceedance	0.008	0.004	0.002
Maersk exceedance	0.004	0.002	0.002
Both exceeding	0.0007	0.0004	0.0003

Table 5: Probabilities of the bivariate GPD

Finally the conditional probabilites of the equities surpassing the thresholds, given the other surpasses:

$$\widehat{\text{MN}} = \Pr\{\text{Maersk} > \hat{y} | \text{Novo} > \hat{x}\} \tag{4.46}$$

$$\widehat{\text{NM}} = \Pr\{\text{Novo} > \hat{x} | \text{Maersk} > \hat{y}\} \tag{4.47}$$

Conditional exceedances	$\hat{x}$ and $\hat{y}$	$\hat{x} = 0.06$ $\hat{y} = 0.08$	$\hat{x} = 0.07$ $\hat{y} = 0.09$
$\widehat{\text{MN}}$	0.090	0.099	0.110
$\widehat{\text{NM}}$	0.167	0.156	0.144

Table 6: Conditional Probabilities of the bivariate GPD

As to be expected, the marginal probabilities drop as the thresholds increase, as the extreme observations become fewer and fewer.
On the other hand the conditional probabilities stay at a high level, or even rise a little. This could suggest that there is some external market factor that drives both equities up at the same point in time.

As mentioned, it has proven difficult to implement the multivariate in the form of analysis of the equities. Full **R** code is supplied in the appendix.
Multivariate theory will be touched upon further in the discussion section.

5 Aspects of the underlying dependency

For now all scope of this assignment has regarded theories of modelling data with an underlying independent structure. As discussed there is reason to suspect that there is not complete independence in the data chosen for the empirical research.
As with data such as the equity data, a given value observed in period $t = n$, is dependent on the previous value observed in period $t = n - 1$. This corresponds to a case where the underlying data structure is that of a Markow-chain, and thus incidentally not being independent.
One argument with regards to the empirical work, especially those of the block maxima, both the basic case and the bivariate case, is that blocking the observations smooth out the underlying dependency. An example of this, is that observations will be between one day and almost two months apart for the monthly blocks, between one day and almost half a year apart for the quarterly blocks and so forth.

With these aspects considered, theory on how EVT can be used on series of dependent data will now be explained.
The work in this section is referenced to (Coles, S., 2001, Chapter 5 - Appendix A.1) unless other sources are stated. This section will solely be theoretical with no empirical implementation. This is meant as reflection on the use of iid theories, with data of dependent underlying structure.

5.1 EVT for modelling dependent data

One way of looking at it, is to assume the data is a stationary series. Here observations are potentially The trick is to have a formulation that expresses the independence, or near-independence of observations being suitably far apart, in terms of time.

A stationary series $X_1, X_2, ...$ is said to satisfy the condition, $D(u_n)$, if for all $i_1 < ... < i_p < j_1 < ... < j_p$ with $j_1 - i_p > l$ (Coles, S., 2001, page 93 - Appendix A.1), such that

$$|\Pr\{X_{i1} \leq u_n, ..., X_{ip} \leq u_n, X_{j1} \leq u_n, ..., X_{jp} \leq u_n\} - \Pr\{X_{i1} \leq u_n, ..., X_{ip} \leq u_n\}\Pr\{X_{j1} \leq u_n, ..., X_{jp} \leq u_n\}| \leq \alpha(n, l) \tag{5.1}$$

where $\alpha(n, l) \to 0$ for some l_n, where $\frac{l_n}{n} \to 0$ for $n \to \infty$.
Looking at the above probability, this would only ever be true zero if the observations are independent. Although, under the alternate assumption that the data in question are not independent, the difference in probability above will not be true zero, by making n large enough, thus being spread enough apart in terms of time, the difference in probability will be close to zero.

Following the definition, that $M_n = \max\{X_1, ..., X_n\}$, and if the sequences of constants $\{a_n > 0\}$ and $\{b_n\}$, that

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \to G(z) \tag{5.2}$$

with G being a non-degenerate distribution function, and for real values of z.

Now with an example of this, by following the example (Coles, S., 2001, page 94-95 - Appendix A.1), where it is seen that a distribution of a series, X_i has a marginal Fréchet distribution, under the assumptions, and is also stationary: Letting $Y_0, Y_1, Y_2, ...$ be a sequence of independent random variables, distributed by:

$$F_Y(y) = \exp\left\{-\frac{1}{(a+1)y}\right\}, y > 0 \tag{5.3}$$

with $0 \leq a \leq 1$. Then the series X_i is defined as:

$$X_0 = Y_0, \qquad X_i = \max\{aY_{i-1}, Y_i\}, i = 1, ..., n \tag{5.4}$$

Thus, for $i = 1, ..., n$ and given $x > 0$:

$$\Pr\{X_i \leq x\} = \Pr\{aY_{i-1} \leq x, Y_i \leq x\} = \exp\left(-\frac{1}{x}\right) \tag{5.5}$$

Looking then at the series $X_1^*, X_2^*, ...$, with a similar marginal Fréchet distribution as the X_i series. Similarly, defining $M_n^* = \max\{X_1^*, ..., X_n^*\}$, and thus the probabilities related to M_n^*:

$$\Pr\{M_n^* \leq nz\} = [\exp\{-1/(nz)\}]^n = \exp(-1/z) \tag{5.6}$$

And similarly the probabilities related to M_n:

$$\begin{aligned}
\Pr\{M_n \leq nz\} &= \Pr\{X_1 \leq nz, ..., X_n \leq nz\} \\
&= \Pr\{Y_1 \leq nz, aY_1 \leq nz, ..., aY_{n-1} \leq nz, Y_n \leq nz\} \\
&= \Pr\{Y_1 \leq nz, ..., Y_n \leq nz\} \\
&= \left[\exp\left\{-\frac{1}{(a+1)nz}\right\}\right]^n \\
&= \{\exp\left(-\frac{1}{z}\right)\}^{\frac{1}{a+1}}
\end{aligned} \tag{5.7}$$

Where it has been utilized that the parameter $a \leq 1$. This gives the result that:

$$\Pr\{M_n^* \leq nz\} = [\Pr\{M_n \leq nz\}]^{\frac{1}{a+1}} \tag{5.8}$$

This relationship between M_n and M_n^* can be expanded to cover a wider group of stationary series: Keeping the definitions of X_i and X_i^*, and equally, M_n and M_n^*, then:

$$\Pr\left\{\frac{M_n^* - b_n}{a_n} \leq z\right\} \to G_1(z) \tag{5.9}$$

with $n \to \infty$ for normalizing sequences of $\{a_n > 0\}$ and $\{b_n\}$, where G_1 is a non-degenerate

distribution function, if and only if:

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \to G_2(z) \tag{5.10}$$

with

$$G_2(z) = G_1^{\theta}(z) \tag{5.11}$$

with $0 < \theta \leq 1$, θ is the extremal index, as touched upon in section (2.4).

To summarize what has been theorized for now, is basically that, maxima of a stationary series will converge, given sufficient $D(u_n)$ condition. Since the series maxima converge, equation (5.11) gives that its limiting distribution, is related to a limiting distribution of a independent series.
Actually, if G_1 is a GEV distribution, where parameters are (μ, σ, ξ), and where $\xi \neq 0$, then for G_1^{θ}:

$$\begin{aligned} G_1^{\theta}(z) &= \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}^{\theta} \\ &= \exp\left\{-\theta\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \\ &= \exp\left\{-\left[1+\xi\left(\frac{z-\mu^*}{\sigma^*}\right)\right]^{-1/\xi}\right\} \end{aligned} \tag{5.12}$$

with

$$\begin{aligned} \mu^* &= \mu - \frac{\sigma}{\xi}(1-\theta^{-\xi}) \\ \sigma^* &= \sigma\theta^{\xi} \end{aligned} \tag{5.13}$$

Notably the shape parameters of the two distributions are equal one another, thus if G_1 is distributed by the Gumbel distribution, then G_2 will also be distributed a Gumbel distribution, with location and scale parameters determined by:

$$\begin{aligned} \mu^* &= \mu + \sigma \log\theta \\ \sigma^* &= \sigma \end{aligned} \tag{5.14}$$

So all in all, the question that is lurking behind all this boils down to: How does all this affect the work done in previous sections.
For the block maxima models, the maxima observations are spread out increasingly, less for the monthly, but much (potentially) for the yearly. This would suggest that the block maxima models are quite robust in terms of the underlying dependency. With increasing block size, the models become more and more robust.

With regards to the threshold models, it is another case. Since there is reason to believe some underlying dependency in the data, extremes are likely to cluster up. To counter this, a method of declustering the data would be needed to be implemented. This boils down to setting up some time-based rule of when an extreme observation is independent. For the equity data, this could be anything from between 1 day, up to potentially strictly less than a month. By declustering the data, the new extremes would be the extremes of the clusters created, and the GPD would then be fitted to the new cluster maxima.

That concludes the thoughts on the underlying dependency of the data, and the first part of the thesis.
Now on to some perspective with regards to managing risk, based on EVT.

Part II

Risk Management with EVT

6 Extreme Value Theory for Risk Management

This chapter will focus on some articles, and act as perspectivation regarding the use of EVT as a tool for managing risk. The articles take different approaches to doing this, and this section will act as a summary is these. The main articles are (Gilli and Këllezi - 2006) and (McNeil, A. - 1999). Besides the articles (Röman, J. - 2017) has been used for definitions.

6.1 Value-at-Risk and Expected Shortfall

Two of the most commonly used measures of risk in the literature, are the measures VaR (Value-at-Risk) and ES (Expected Shortfall), which both are estimations based on the distribution of the underlying variables.
The definition of these measures is referenced to (Röman, J. - 2017).
The aim of VaR as a measure, is to find a given point on the distribution, say the need is to find, with 95% certainty, how much possible value is at risk of being actually lost, over some horizon of time.
Quantifying this, the formula for VaR is:

$$\text{VaR}_\alpha(\mathscr{L}) = inf\{c : P(\mathscr{L} > c) \leq 1 - \alpha\} \tag{6.1}$$

$$\text{VaR}_\alpha(\mathscr{L}) = F^{-1}(\alpha) \tag{6.2}$$

The 95% case would translate to the measure $\text{VaR}_{0.05}$, and the value calculated would be the amount at risk at that given confidence interval.

This type of measure has been widely discussed for being inadequate as a measure of risk, since it only gives an amount that is at risk, but not what would actually be lost, if the extreme loss case actually happened. Another major flaw of VaR is, that the measure is not a coherent risk measure (Röman, J. - 2017, chapter 2.9.9).

Without diving further, a risk measure is coherent, if it complies with the properties of, a) monotonicity, b) proportionality, c) translation and d) subadditivity.
For VaR, subadditivity is not satisfied, as two assets with similar VaR were to be joined in a portfolio, the joint VaR of the portfolio would not correspond to the additive VaR of the two assets separately. Not concluding that VaR is useless, but it needs the right setting to be of real value, which in term leads to VaR as a steppingstone for something better.
ES goes also by another name, which is Conditional VaR, since the calculation of ES can be put, so it is a probability, conditional that VaR is met. Going from VaR to ES is a matter of calculating the expected amount, actually being lost in the case that VaR_p is exceeded, thus the calculation:

$$\text{ES}_(\alpha) = E\{\mathscr{L}|\mathscr{L} \geq \text{VaR}_\alpha(\mathscr{L})\} \tag{6.3}$$

$$\text{ES}_(\alpha) = \frac{\int_\alpha^1 \text{VaR}_\gamma(\mathscr{L})d\gamma}{1-\alpha} \tag{6.4}$$

Contrary to VaR, ES is in fact a coherent risk measure, thus having some properties that are useful when working with simulations of different combinations of portfolios to compare risk across.
The positive thing for VaR compared to ES is, that VaR is far more easy to comprehend, whereas ES gets a bit harder. In the context of EVT, both EVT and these two risk measures aim to find extreme cases of events, based on some underlying distribution.
As discussed earlier in the paper, some models are of limited usage when it comes to heavy-tailed behavior of some distributions.

This heavy-tailed problem, is also apparent when looking at the relationship between VaR and ES, as two distributions might have the same VaR, but two very different ES. A graphical representation of this: In this theoretical example, exactly this behavior of

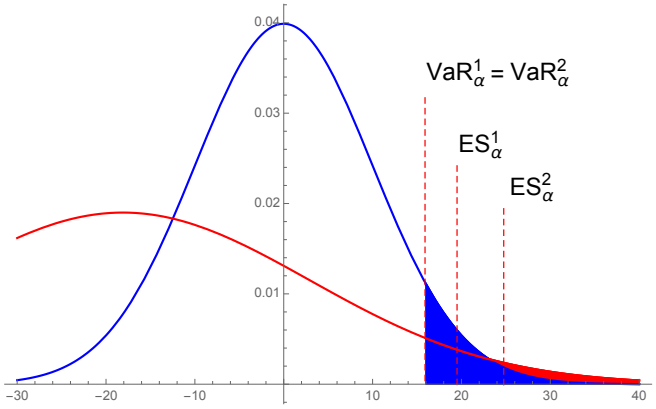

Figure 11: VaR and ES graphically

similar values for VaR, but different sizes for ES is apparent, in practise this results in ES being very heavy-tail vulnerable, and the estimation of the distribution therefore a crucial point in the estimation of risk.
So a standalone VaR, or even a standalone ES might not suffice for most risk managers, but the combination of these make for something more reliable and less vulnerable.

The following rewriting of VaR and ES is referenced to (Gilli and Këllezi - 2006).
Since the connection between EVT and the two risk measures are connected as mentioned above, and EVT works with actually estimating the tail of a distribution, VaR and ES can be rewritten to fit the terminology of this paper, and thus EVT.
Using GPD from equation (2.26), VaR is rewritten as $\widehat{\text{VaR}}$:

$$\widehat{\text{VaR}}_p = u + \frac{\hat{\sigma}}{\hat{\xi}}\left(\left(\frac{n}{N_u}^{\hat{\xi}} - 1\right)\right) \tag{6.5}$$

Thus by having explained VaR as a measure in the EVT terminology, and expressed with the parameters of the GPD, it is easy to put ES in the same, as ES is an expression of VaR:

$$\widehat{ES}_p = \widehat{\text{VaR}}_p + E\left(X - \widehat{\text{VaR}}_p | X > \widehat{\text{VaR}}_p\right) \tag{6.6}$$

The part of the equation where the expected value of is represented, is the same for the calculation of the mean excess function, given the threshold chosen is equal to $\widehat{\text{VaR}}_p$, and $\xi < 1$:

$$e(z) = E(X - z | X > z) = \frac{\sigma + \xi z}{1 - \xi} \sigma + \xi z > 0 \tag{6.7}$$

Since the term can be expressed like this, the formulation of $\widehat{ES}_p$ reformulates as:

$$\begin{aligned} \widehat{ES}_p &= \widehat{\text{VaR}}_p + \frac{\hat{\sigma} + \hat{\xi}(\widehat{\text{VaR}}_p - u)}{1 - \hat{\xi}} \\ &= \frac{\widehat{\text{VaR}}_p}{1 - \hat{\xi}} + \frac{\hat{\sigma} + \hat{\xi} u}{1 - \xi} \end{aligned} \tag{6.8}$$

6.2 Problems & Opportunities of Using EVT

Usual assumptions concerning EVT, revolves around losses when looking at financial data, especially when looking at the usage in concern to measuring risk involved.

This translates, as mentioned in section (1.1), to looking at only the one tail of the distribution, but only looking at half the maxima is obviously not reasonable when including more than just the usual sense of investment, where negative returns translates to losses.

Obviously a long position would result in the downside when the equity or stock or alike goes down in value, but an equivalent short position in the same, would have its downside when the equity goes up in value!
In other words, going from only looking at one tail, to including both tails in the data, as both tails have interesting tales to tell! Since in reality the distributions will not be symmetrical across the middle, an interesting point of research would be to figure out, whether or not there would be a significant difference between these, and thus a possibility to make make use of this.

Part III

Discussion - Conclusion

7 Discussion and further research

For the discussion, different views and aspects of the work done, and possible further work will be discussed. The first part will revolve around the data chosen, and the models included in the empirical work in the thesis. Alternate approaches for using EVT in accordance with other tools for looking at financial data will also be discussed. For the last part of the discussion, an area regarding the copulas of the multivariate theory will be discussed. Here, the article (Mikosch, T - 2005, Copulas: Tales and Facts - Appendix A.1) has laid the foundation for this discussion, and how and why copulas has gained so much popularity in the world of finance.
The general aim of this section is to bring other views than the ones presented in the previous sections. By doing so, rounding off the thesis and having other points of views to the theories and empirical work.

7.1 Analytical work and data selection

When looking over the data chosen for the analytical work, one of the key aspects for choosing the two equities were the amount of data points available. This is one the bigger challenges, as the interesting extremes are, as pointed out earlier, not appearing too often. When looking at some of the other equities that were considered, including Carlsberg, Pandora, ISS and so forth, they have in common that they are all part of the C25 Index, but are struggling with the amount of data available. This is due to their "young" lifetime as publicly traded equities, and thus available datapoints for the analytical work. When that being said, all of these would be interesting to review under the theories of EVT, as they probably have different underlying distributions as the two analysed in this thesis.

The considerations for choosing data for the analysis work, revolved around choosing some equities on the danish market, as this would make it more relatable, and also to have some prior knowledge of what and how the companies work and operate. This of cause puts limits on the available equities, and would of cause be a major consideration for further research to remove this limit, and broaden the search for interesting equities on the global market.
For a different approach than just looking at equities, portfolios might also be considered as possible suppliers of data for analytical work. The easy choice would be to pick out some of the already constructed portfolios, or indexes, i.e. C25 for danish equities, Euro Stoxx 50 for European equities.
Alternatively building a portfolio of equities oneself would also be quite interesting. Important considerations for building the portfolio should include considerations regarding the interdependence of the equities, lifetime for amount of datapoints, and how their different underlying distributions are. Selected carefully a portfolio with a nice variance spread would make for an interesting input into some of the models, as it would have some properties that are not available for single equities alone. The most reasonable for analysis on a portfolio, would be to assume it to be one datastring, and thus apply the classical theories of either block maxima or threshold models. With regards to the section about modelling dependent series, this would be a genuine concern when looking at such a portfolio. As mentioned in that section, using the block maxima models would be unhindered by the underlying dependency structure, when taking large enough blocks. On the other hand, when modelling threshold models on portfolios the underlying structure would have to be assessed properly, and possibly the need for using declustering of the data, and use the maxima of the clusters.

When looking at the models tested in the analytical sections of the thesis, other approaches would also have been valid. The classical models of the block maxima and threshold models are tested as they could, but approaches for testing the threshold models with regards to the fact that the distribution of data chosen might not be as

independent as needed. It would be interesting to research how implementation of the grouping of maxima would impact the threshold models. In particular great declines over larger periods would be smoothed out, as usually great declines are more frequent that prolonged periods of great inclines. For the data in particular this would probably impact that the heavy-tail behavior of the distributions, at least for the side that distributes decline in returns.
For the multivariate modelling, it would be interesting to research, how a transformation would impact the results. Especially if it would be possible to make use of the EVT as a predictive tool. The reasoning being, that by shifting the dates, by a week or a month or any other timeperiod, would it then be possible to set up a probability calculation such that it would be possible, to some degree, to predict decline or rise of another equity based on how the other equity is doing.

7.2 Thoughts on copulas

The article (Mikosch, T - 2005, Copulas: Tales and Facts - Appendix A.1), put the multivariate theories of EVT, the copulas, into some interesting context.
Looking first at the "popularity" of the term, with regards to academic papers published which utilizes the theories. The amount of papers published has spiked immensely, (Mikosch, T - 2005) reports an increase in search results on Google of the word "copula", increased from 10.000 in 2003 to 650.000 in 2005.
The same search done 30/07/2019 produced 770.000 results, and a search on Google Scholar of "copulas finance" produces 8.020 articles since 2015. To add further to this, (Embrechts and Hofert - N.D) also mentions this spike in publications.

The dilemma lies, with the reason for this rise in popularity in recent years, since the concept of copulas can be traced back to the middle of the last millennium (Embrechts and Hofert - N.D).
Maybe this is due to copulas being an "exotic" topic within statistics, but as regarded by (Mikosch, T - 2005), the concept does not add anything new, as it is merely a reformula-

tion of the original distribution.
As for the statistical properties of copulas, there is no clear proof that copulas offer a good fit of the underlying distributions of the data. The reason for using copulas is often because it provides some pretty mathematical formulations (Mikosch, T - 2005). Another thing to keep in mind, is that copulas does not solve the problem multidemsionality, that is the range of possible outcomes when working in higher dimensions. The mathematical formulation of the copulas might be pretty, but the underlying problem is not solved, as the multidemsionality problem does not magically disappear.

So all in all, the multivariate theories explained in this thesis, might not be viable in the field of finance after all, or put in another way: *The Copula Fashion - The Emperor's New Clothes* (Mikosch, T - 2005).
This concludes the discussion part, and the final point is to provide some conclusion to the models.

8 Conclusion

With the overall topic of the thesis, being the research of Extreme Value Theory. Throughout working on this thesis, various topics of EVT outlined, and some of these analysed on the equities of Maersk B and Novo Nordisk B. The different approaches of the block maxima, and the threshold models has been tested, and both have strengths and weaknesses.

With regards to this underlying dependency structure of the data, the block maxima approach might be best, as when it comes to dependency issues, these models handle it quite well. The best payoff between lack of data, and smoothing of dependency were found to be when using quarterly maxima.
The threshold models offer more efficient use of the data, and might therefore offer a more true approximation of the distribution of maxima, but they come short when it

comes to dealing with data which is not iid. Choosing a reasonable threshold is one of the bigger challenges, and for both equities this was analysed by use of the mean residual life plot.
The modeling of dependent series were outlined, and would probably be the better approach for analysing financial data, where there is reason to suspect an underlying dependency.

When it comes to the different methods of estimation, both maximum likelihood and profile likelihood were applied. For the maximum likelihood, this is a good method of estimation, when it is possible to do numerical solutions. The profile likelihood on the other hand, offers a graphical presentation, that is quite appealing, and a good estimation as well.

The analysis of the extremes of the two equities went according to plan, as for the the univariate models. With regards to the multivariate models, only some results were able to be found.
The multivariate theories has been outlined, but the analysis part based on this was hard to implement, and proved inconsequential. This was a setback, as the theories offers some interesting possibilities for risk managers, investors and the like. Having broadened the horizon with some interesting articles regarding the topic of copulas, the conclusion must be that the topic is fragile in regards to working with it as a tool for managing risk in finance.

A Appendix

A.1 References

Books:

- Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values*. Springer.
- J. McNeil, A., Frey, R. and Embrechts, P. (2005) *Quantitative Risk Management*. Princeton University Press.
- R. M. Röman, J. (2017) *Analytical Finance - Volume 1.* Palgrave Macmillan, ISBN: 978-3-319-34027-2

Academic articles:

- Mikosch, T. (2005) *Copulas: Tales and Facts*. `http://web.math.ku.dk/~mikosch/maphysto_extremes_2005/s.pdf`
- Embrechts, P. and Hofert, M. (N.D) *Statistics and Quantitative Risk Management for Banking and Insurance* `https://people.math.ethz.ch/~embrecht/ftp/qrm_stat_review.pdf`
- J. McNeil, A (1999)*Extreme Value Theory for Risk Managers*. `http://www.sfu.ca/~rjones/econ811/readings/McNeil%201999.pdf`
- J. McNeil, A (1998)*Calculating Quantile Risk Measures for Financial Return Series using Extreme Value Theory*. `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.487.8734&rep=rep1&type=pdf`
- X. Diebold, F., Schuermann, T. and D. Stroughair, J. (2000) Pittfalls and Opportunities in the Use of Extreme Value Theory in Risk Management *The Journal of Risk Finance, Vol 1 Issue 2, pp. 30-35*.
- Gilli, M. and Këllezi, E. (2006) An Application of Extreme Value Theory for Measuring Financial Risk *Computational Economics, 27: 207-228*.

A.2 Illustrations

Novo Nordisk B equity overview:

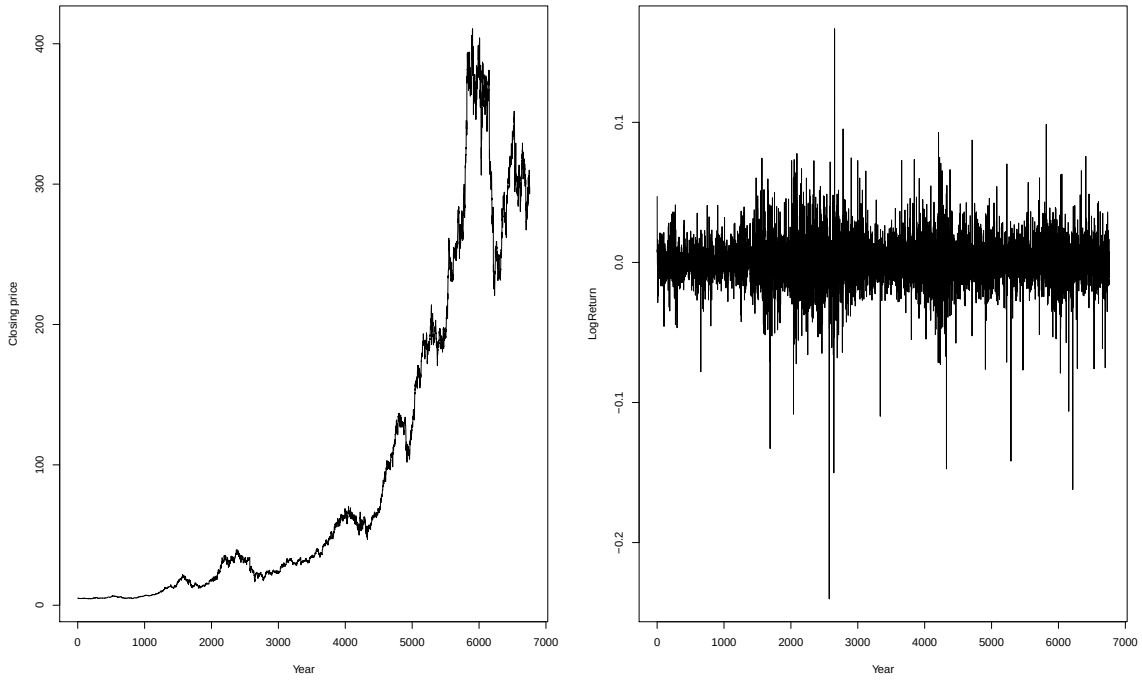

Figure A.1: Left: Daily closing values of Novo Nordisk B, Right: Log Returns

A.3 R-code

Maersk GEV code:

```
    MaerskRaw <- read.csv(
  file=file.choose(),
  header = T,
  sep = ";")


#Tager fra 1992 til 2018
MaerskReduceret <- get_data(MaerskRaw)
```

```
#Yearly GEV
MaerskYearMax <- -yearly_max(MaerskReduceret)
MaerskYM <- fevd(MaerskYearMax)
summary(MaerskYM)
plot(MaerskYM)

MaerskYM_rl <- return.level(MaerskYM, conf = 0.05,
return.period= c(2,5,10,20,50,100))
MaerskYM_rl

plot(MaerskYM, type="rl",
     main="Return Level Plot for Maersk w/ MLE",pch=16)

#Density of Block Maxima of the Yearly maxima
plot(MaerskYM,type="density",main="")
#Profile likelihood - parameter shape
profliker(MaerskYM, type="parameter", which.par=3, xrange = c(-0.25,1.5))

plot(MaerskYearMax, type="l",xlab = "Yearly maxima",
ylab ="LogReturns", xaxt="n")
axis(1, at=c(2,8,14,20,26),labels=c(1993,1999,2005,2011,2017))

#Quarterly GEV
MaerkQuarterMax <- -quarterly_max(MaerskReduceret)
MaerskQM <- fevd(MaerkQuarterMax)
summary(MaerskQM)
plot(MaerskQM)

MaerskQM_rl <- return.level(MaerskQM, conf = 0.05,
```

```
return.period= c(2,5,10,20,50,100))
MaerskQM_rl
plot(MaerskQM, type="rl",
     main="",pch=16)

#Density of Block Maxima of the Quarterly maxima
plot(MaerskQM,type="density",main="")
#Profile likelihood - paramter shape
profliker(MaerskQM, type="parameter", which.par=3,
xrange = c(-0.2009,1.11),main="")
profliker(MaerskQM, type="parameter", which.par=3, main="")


ci(MaerskQM, type="parameter", which.par=1 , method = "proflik",
verbose = FALSE)
ci(MaerskQM, type="parameter", which.par=2 , method = "proflik",
verbose = FALSE)
ci(MaerskQM, type="parameter", which.par=3 , method = "proflik",
verbose = TRUE, xrange = c(-0.2,1.1))
ci(MaerskQM, type="parameter", which.par=3 , method = "proflik",
verbose = FALSE)

ci(MaerskQM, type="parameter", which.par=1 , method = "normal",
verbose = FALSE)
ci(MaerskQM, type="parameter", which.par=2 , method = "normal",
verbose = FALSE)
ci(MaerskQM, type="parameter", which.par=3 , method = "normal",
verbose = FALSE)
```

```
plot(MaerkQuarterMax, type="l",xlab = "Quarterly maxima",
ylab ="LogReturns", xaxt="n")
axis(1,at=c(2,8,14,20,26)*4,labels=c(1993,1999,2005,2011,2017))
#Monthly max for graph
MaerskMonthlyMax <- -monthly_max(MaerskReduceret)
plot(MaerskMonthlyMax, type="l",xlab = "Monthly maxima",
ylab ="LogReturns", xaxt="n")
axis(1,at=c(2,8,14,20,26)*12,labels=c(1993,1999,2005,2011,2017))
```

Novo GEV code

```
NovoRaw <- read.csv(
  file=file.choose(),
  header = T,
  sep = ";")

#Tager fra 1992 til 2018
NovoReduceret <- get_data(NovoRaw)

NovoRev <- NovoReduceret[rev(order(NovoReduceret$LogReturns))]

plot(NovoReduceret$LogReturns, type="l", xlim=c(max, min))
plot(NovoReduceret$Last.Price, type="l")
plot(rev(NovoReduceret$LogReturns), type="l", ylab="LogReturn",
xlab="Year")
plot(rev(NovoReduceret$FixedClosing), type="l",
ylab="Closing price", xlab="Year")
#Yearly GEV
NovoYearMax <- -yearly_max(NovoReduceret)
NovoYM <- fevd(NovoYearMax)
summary(NovoYM)
```

```
plot(NovoYM)

NovoYM_rl <- return.level(NovoYM, conf = 0.05,
return.period= c(2,5,10,20,50,100))

plot(NovoYM, type="rl",
     main="Return Level Plot for Novo w/ MLE",pch=16)
NovoYM_rl

plot(NovoYM,type="density",main="")

plot(NovoYearMax, type="l",xlab = "Yearly maxima",
ylab ="LogReturns", xaxt="n")
axis(1, at=c(2,8,14,20,26),labels=c(1993,1999,2005,2011,2017))

#Quarterly GEV
NovoQuarterMax <- -quarterly_max(NovoReduceret)
NovoQM <- fevd(NovoQuarterMax)
summary(NovoQM)
plot(NovoQM)

NovoQM_rl <- return.level(NovoQM, conf = 0.05,
return.period= c(2,5,10,20,50,100))
NovoQM_rl
plot(NovoQM, type="rl",
     main="",pch=16)

plot(NovoQM, "hist")
```

```
plot(NovoQM,type="density",main="")

#Profile likelihood - paramter shape
summary(profliker(NovoQM, type="parameter", which.par=3,main=""))

ci(NovoQM, type="parameter", which.par=1 , method = "proflik",
verbose = FALSE)
ci(NovoQM, type="parameter", which.par=2 , method = "proflik",
verbose = FALSE)
ci(NovoQM, type="parameter", which.par=3 , method = "proflik",
verbose = FALSE)

ci(NovoQM, type="parameter", which.par=1 , method = "normal",
verbose = FALSE)
ci(NovoQM, type="parameter", which.par=2 , method = "normal",
verbose = FALSE)
ci(NovoQM, type="parameter", which.par=3 , method = "normal",
verbose = FALSE)

ci(NovoQM, type="parameter", which.par=3 , method = "proflik",
verbose = TRUE, xrange= c(-0.05,0.7))

plot(NovoQuarterMax, type="l",xlab = "Quarterly maxima",
ylab ="LogReturns", xaxt="n")
axis(1,at=c(2,8,14,20,26)*4,labels=c(1993,1999,2005,2011,2017))
```

Maersk GPD code

```
mrlplot(-MaerskReduceret$LogReturns*100, xlab="u")
threshrange.plot(-MaerskReduceret$LogReturns*100, type = "GP",
r = c(0,2), nint=100)
```

```
MaerskGPD <- fevd(x = -MaerskReduceret$LogReturns,
threshold = 0.07, units="LogReturns", type = "GP")
summary(MaerskGPD)

plot(-MaerskReduceret$LogReturns, type = "p",ylim = c(0,0.2),
ylab="LogReturns")
abline(h=0.07, col="purple", lty=2, lwd=2)

lengthMaersk <- length(-MaerskReduceret$LogReturns)
lengthMaersk
exceedMaersk <- sum(-MaerskReduceret$LogReturns>0.07)
exceedMaersk
ciMaersk <-exceedMaersk/lengthMaersk
ciMaersk

profliker(MaerskGPD, type="parameter", which.par=2,main="",
xrange = c(-0.45,-0.2))
ci(MaerskGPD, type="parameter", which.par=2 , method = "proflik",
verbose = TRUE,xrange = c(-0.455,-0.3))
plot(MaerskGPD, type="rl",
     main="",pch=16)
```

Novo GPD code

```
mrlplot(-NovoReduceret$LogReturns*100, xlab="u")

NovoGPD <- fevd(x = -NovoReduceret$LogReturns,
threshold = 0.05, units="LogReturns", type = "GP")
summary(NovoGPD)
```

```
plot(-NovoReduceret$LogReturns, type = "p",
ylim = c(0,0.2),ylab="LogReturns")
abline(h=0.05, col="purple", lty=2, lwd=2)

lengthNovo <- length(-NovoReduceret$LogReturns)
lengthNovo
exceedNovo <- sum(-NovoReduceret$LogReturns>0.05)
exceedNovo
ciNovo <-exceedNovo/lengthNovo
ciNovo

profliker(NovoGPD, type="parameter", which.par=2,
main="",xrange = c(0.25,0.8))
ci(NovoGPD, type="parameter", which.par=2 , method = "proflik",
verbose = FALSE)

plot(NovoGPD, type="rl",
     main="",pch=16)
```

Multiva GPD code

```
install.packages('evd')
library(evd)

plot(x = NovoQM$x, y = MaerskQM$x)

#bivariate model
BIVgpd <- gpdbiv(data1 = NovoReduceret$LogReturns,
data2 = MaerskReduceret$LogReturns, u1 = 0.05, u2= 0.07)
plot(BIVgpd)
```

```
interpret.gpdbiv(BIVgpd,0.05,0.07)
interpret.gpdbiv(BIVgpd,0.06,0.08)
interpret.gpdbiv(BIVgpd,0.07,0.09)


novoMulti <- matrix(NovoReduceret$LogReturns)
maerskMulti <- matrix(MaerskReduceret$LogReturns)

cbind(novoMulti,maerskMulti)

multiData <- merge(NovoReduceret,MaerskReduceret, by="X...Date")

pmvevd(multiData$LogReturns.x,multiData$LogReturns.y,
dep = 1, asy = c(1,2,3))

plot(x = multiData$LogReturns.x, y = multiData$LogReturns.y)
abline(h = 0.07)
abline(v = 0.05)
```