



Ca' Foscari
University
of Venice

Master's Degree programme
in Finance

Final Thesis

Extreme Value Theory for Portfolio Risk Management

Supervisor

Ch. Prof. Davide Raggi

Graduand

Nicola Ferretti

Matriculation Number 887605

Academic Year

2021 / 2022

A mio nonno.

Index

Introduction	7
Chapter I Portfolio Optimization	9
1.1. Markowitz Model	9
1.2. The effects of diversification	15
1.3. Role of correlation	18
1.3.1. Coefficient of correlation between R_1 and R_2 equals $+1$ ($\rho_{1,2} = +1$)	18
1.3.2. Coefficient of correlation between R_1 and R_2 equals -1 ($\rho_{1,2} = -1$)	19
1.4. Criticisms of the Modern Portfolio Theory	20
Chapter II Synchronization of the Returns	23
2.1. State-Space Model	25
Chapter III Extreme Value Theory	27
3.1. Modelling Block Maxima.....	27
3.1.1. The Fisher-Tippet Theorem and the GEV Distribution.....	27
3.1.2. Estimating the GEV distribution by the Maximum Likelihood.....	30
3.1.3. Return Level.....	31
3.2. Peaks Over Thresholds (POT).....	32
3.2.1. The Limiting Distribution of Extremes Over High Threshold and the GPD distribution	33
3.2.2. Estimating the GPD by the Maximum Likelihood.....	36
3.2.3. Estimating the Tails of the Loss Distribution	36
3.3. Risk Measures	37
Chapter IV Real Data Application	39
Conclusion	51
Bibliography	53
Annex A Code	55

Introduction

The term “extreme” refers to the complete class of phenomena whose occurrence is rare, and which can lead, precisely because of their unusual character, to disastrous environmental, economic, and human impacts.

The probabilistic theory dealing with these events is called Extreme Value Theory and aims to develop mathematical methods and models able to describe and to predict the occurrence of such rare phenomena.

Initially, the analysis of extremes was introduced in hydrology and meteorology in order to study flood levels and natural disasters. However, in recent years the domain of application of this statistical methodology has managed to include other disciplines and applied sciences such as finance, traffic prediction, insurance, and structural engineering.

The first studies on extreme values date back to the first part of the twentieth century, when Tippet and Fisher (1928) stated an asymptotic argument which represents the cornerstone of extreme value theory: the Extremal Types Theorem. In the subsequent years, the asymptotic theory was extended and codified by Gnedenko (1948) and Gumbel (1958), while the characterization of extremes as observation exceeding and high threshold is due to Pickands (1970).

In this thesis, the role of Extreme Value Theory (EVT) in risk management is addressed as a method for modelling and measuring extreme risks applied to a portfolio. The synchronization of the markets is of fundamental importance in this application to know how it affects the composition of the best portfolio, in particular the correlation between returns of the stocks in the portfolio, and the calculation of risk measures.

The present thesis is articulated in four chapters. Firstly, Chapter 1 contains an overview of the Markowitz model used to compose different portfolio and select the best one according to the Sharpe ratio. Furthermore, in the other paragraph of the chapter are presented the importance of diversification in a portfolio and how it affects the risk, and the role of the correlation in this model. Lastly, the criticisms of the Markowitz model are described.

The problem of diversification of the market and the returns' stocks is defined in Chapter 2. Here, State-Space models are described, which is the model that is used to synchronize data, and which are the main characteristics of these models.

The two main types of models are then described in Chapter 3 about the EVT. The first class of model is the Block Maxima. In particular, it is studied how to derive it from the Fisher-Tippet Theorem and how to estimate the distribution of the model through the maximum likelihood function. Finally, it is possible to calculate specific risk measures using the estimates of the Block Maxima coefficients. Subsequently, a similar process is used to describe the second class of model: the Peak Over Threshold (*POT*) model. Also, for this model, the main characteristics are studied and, through the likelihood function, the coefficient for the computation of the risk measures are estimated. In the last paragraph are defined the two main risk measure that can be calculated using the estimates of the POT model: the Value-at-Risk (*VaR*) and the Expected Shortfall (*ES*).

Finally, in Chapter 4, all previously mentioned techniques and models are applied to a data consisting of stock prices, over the last seven years (2015-2021) from all over the World, with the aim of analyzing the behaviour of the VaR and ES as the initial assumptions of -the models change. In particular, three Python codes are created in order to have different scenarios to be analyzed. It is possible to examine the codes in *Annex A*.

Chapter I Portfolio Optimization

Portfolio optimization is the process of selecting the best portfolio of given assets, or assets distribution, out of the set of all portfolios being considered. But which is the best portfolio available? This question is probably as old as the stock-market itself.

Harry Markowitz pioneered this stand of literature, and in his contribution “*Portfolio Selection*” published in 1952 in the *Journal of Finance*¹ provided the foundation for Modern Portfolio Theory (MPT) as a mathematical problem.

1.1. Markowitz Model

Until Markowitz’s study, a portfolio was considered diversified if it consisted of a series of individual assets, each valued separately from the others. The overall portfolio risk was derived from the sum of the individual risks of each security.

In the model proposed by Markowitz, it is stated that it is not enough to consider the characteristics of the asset individually, but it is necessary to account for the movements that occur between them, and which are represented by the covariance between the different assets. Markowitz indeed stated that if the investor also studies the correlation between the assets in the construction of a portfolio, then he will be able to build better portfolios with higher expected returns at the same level of risk or obtain a lower risk’s portfolio for the same expected return, compared to the portfolios constructed by ignoring the effects of the correlation between assets. In the Markowitz model, the risk is measured by the variance. Therefore, the variance of the portfolio is derived from the variance of the individual assets included in the portfolio and by the correlation between the assets. For the implementation of the model, the required inputs are the expected returns and the variance of each asset, and the covariance between the different securities.

Markowitz proved that an investor, under precise assumptions, can build optimal portfolios which, through an efficient combination of mean, variance, and covariance, maximize the expected returns given a certain level of risk, or minimize the risk given a certain expected return. This result is achieved through the diversification of the assets in the portfolio. Markowitz also defined the concept of diversification, in which it is stressed the need to combine highly correlated assets, but it is convenient to consider assets belonging to not necessarily *correlated* sectors. It is for this reason that the Markowitz

¹ Markowitz, H. (1952), *Portfolio Selection*, The Journal of Finance, Vol. 7, No. 1, pp. 77-91.

model, based on the Mean-Variance criteria for portfolio selection, is still considered the starting point of the Modern Portfolio Theory.

The portfolio selection process can be divided into two macro-phases: the first one is based on the analysis of the historical returns of the assets whose data are used to predict their future performance; the second macro-phase considers the analysis outlined above to select the best portfolio combination.

Before proceeding with the description of the Markowitz model, it is necessary to describe the assumptions needed to define this model:

- Absence of taxes and transaction costs.
- Analysis is based on a single-period model of investment.
- Perfect market competitiveness.
- An investor is risk-averse.

Consider a portfolio with n different assets. Let μ_i and σ_i^2 be the corresponding mean and variance of a single asset i and let $\sigma_{i,j}$ be the covariance between R_i and R_j . Suppose that the relative amount of the value of the portfolio invested in asset i is x_i . If R is the return of the whole portfolio, then:

$$\mu = E[R] = \sum_{i=1}^n \mu_i x_i \quad (1.1)$$

$$\sigma^2 = Var[R] = \sum_{i=1}^n \sum_{j=1}^n \sigma_{i,j} x_i x_j \quad (1.2)$$

$$\sum_{i=1}^n x_i = 1 \quad (1.3)$$

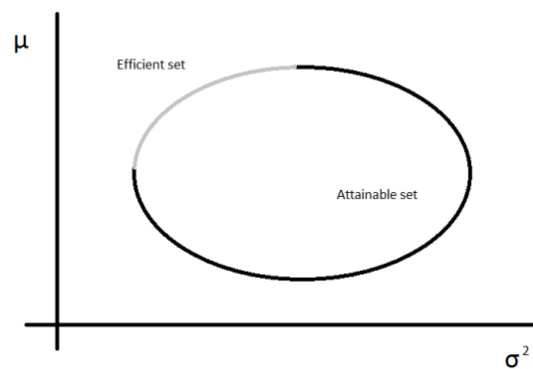
$$x_i \geq 0 \text{ for } i = 1, 2, \dots, n \quad (1.4)$$

Equation (1.1) is the portfolio return defined as the proportion-weighted sum of the assets' returns that compose the portfolio; Equation (1.2) is the portfolio volatility that is a function of the covariances between the different assets. Condition (1.3) defines that the sum of weights must be equal to 1; Condition (1.4) allows to take only long positions, alternatively, it defines the impossibility to short-sell any asset². For different choices of

² When short sales are prohibited, single assets may lie on the frontier. When short sales are allowed, portfolios can be constructed that offer the same expected return and lower variance of

x_1, \dots, x_n the investor will get different combination of μ and σ^2 , meaning that the characteristics of the portfolio will change. The set of all possible (μ, σ^2) combinations is called the *attainable set*. Those (μ, σ^2) with maximum μ for a given σ^2 or less and minimum σ^2 for a given μ or more are called the efficient set, or efficient frontier; while the remaining segment of the attainable set is the inefficient part. Since an investor wants a high profit and a small risk, he/she wants to maximize μ and minimize σ^2 and therefore he/she should choose a portfolio that lies on the efficient frontier.

Figure 1: The attainable set in the (μ, σ^2) plane.

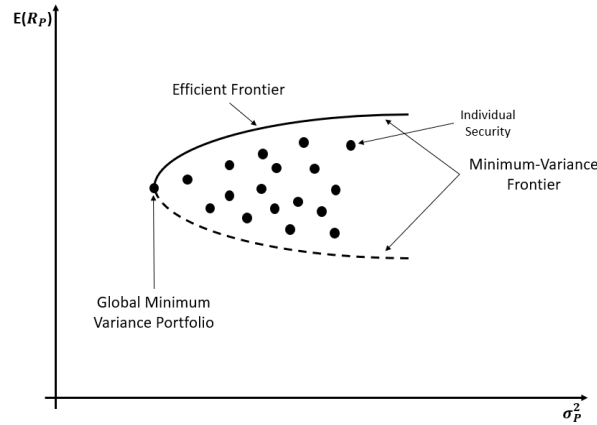


Source: Markowitz H. (1952), Portfolio Selection, The Journal of Finance, Vol. 7, No. 1., pp. 77-91.

In Figure 1 it can be noticed that each portfolio composition that lies on the efficient frontier is more attractive than any other portfolio in the inefficient part, this is because for a given σ^2 any efficient portfolios have a higher μ , or for a given μ they have a lower σ^2 . This is the reason why an investor should choose a portfolio that lies on the efficient frontier.

the single assets that lie on the frontier. These portfolios typically will have short positions in low-expected-return assets.

Figure 2: Efficient frontier in the (μ, σ^2) plane.



Source: Own production.

Figure 2 is focused on the efficient frontier, and the most important portfolio composition is the Global Minimum-Variance Portfolio that defines the separation between the efficient and the inefficient frontier. As it can be seen, for any portfolio that lies on the dashed part of the frontier (inefficient frontier) there exist a portfolio with the same volatility and a larger expected return positioned directly above it. The Global Minimum-Variance Portfolio is the one with the lowest variance among all the efficient portfolios.

Once the efficient frontier has been determined, the second part of the optimization is to select which is the best portfolio. In order to find the optimal portfolio, a possibility is to use the Capital Allocation Line³ (CAL) with the highest Sharpe ratio (the steepest slope) that allows to find the tangency condition between the efficient frontier and the CAL itself. The function of the Capital Allocation Line⁴ is given by:

$$CAL: E(r_C) = r_f + \sigma_C \frac{E(r_P) - r_f}{\sigma_P} \quad (1.5)$$

The slope of the CAL defines the function of the Sharpe ratio:

$$S = \frac{E(r_P) - r_f}{\sigma_P} \quad (1.6)$$

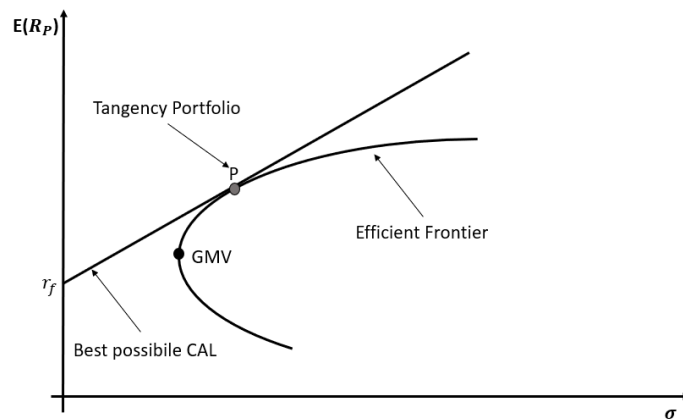
³ The capital allocation line measures the risk of risky and risk-free assets. Its slope is known as the “reward-to-volatility ratio”. In this thesis, for simplicity, the risk-free rate is not considered.

⁴ Bodie, Z., Kane, A., Marcus, A. J. (2018), *Investments, Eleventh Edition*, McGraw-Hill Education, New York.

Where $E(r_p)$ is the expected return of the portfolio, σ_p is its standard deviation (or the volatility) and r_f is the risk-free rate. The Sharpe ratio can be used to evaluate a portfolio's performance where actual returns are used in the formula (1.6), it represents the remuneration that the portfolio gives for each unit of risk, in other words it is the price of the risk.

Finally, from the tangency between the efficient frontier and the Capital Allocation Line with the highest Sharpe ratio, it is possible to find the optimal portfolio P , or Market Portfolio, as it can be seen in Figure 3.

Figure 3: Tangency between Capital Allocation Line (CAL) and Efficient Frontier



Source: Own production.

To complete the model, assuming that each investor can assign a utility to competing portfolios based on the expected return and risk of those portfolios, it can be stated that higher the utility values are assigned to portfolios with more attractive risk-return profiles. Therefore, the optimal portfolio is different for each investor due to the individual level of risk-aversion. An investor tries to maximize his/her utility function and to achieve this it can be used a quadratic function⁵:

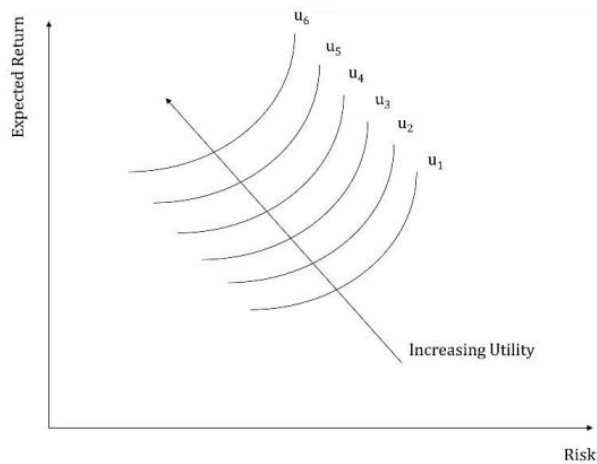
$$U = E(r) - \frac{1}{2}A\sigma^2 \quad (1.7)$$

⁵ Bodie, Z., Kane, A., Marcus, A. J. (2018), *Investments, Eleventh Edition*, McGraw-Hill Education, New York

Where A identifies the risk-aversion coefficient of the individual investor⁶ and the value $\frac{1}{2}$ represents a scaling factor; the utility (U) increases with high expected returns, but it decreases with a high volatility.

For each portfolio can be computed the utility value given the risk-aversion coefficient of the investor, and equally preferred portfolios lie in the Mean-Variance plane on an indifference curve which connects all the portfolios with the same level of utility.

Figure 4: Risk-Averse indifference curves



Source: Phillips P. J. (2009)⁷

As it can be seen in Figure 4, the higher is the indifference curve, the greater is the utility provided by the portfolios that lie on this curve:

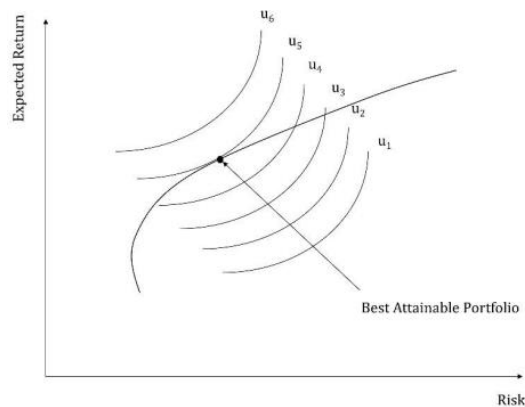
$$u_6 > u_5 > u_4 > u_3 > u_2 > u_1 \quad (1.8)$$

In the end, to find the best portfolio for a single investor, knowing the risk-aversion coefficient, it is enough to select the portfolio where the indifference curve is tangent to the efficient frontier, as shown in Figure 5.

⁶ In this model, it is assumed that the investor is risk-averse, therefore: $A > 0$.

⁷ Phillips, P.J. (2009), *Applying Modern Portfolio Theory to the Analysis of Terrorism, Defence & Peace Economics*, 20, pp. 193-213.

Figure 5: Optimal portfolio choice for a single investor



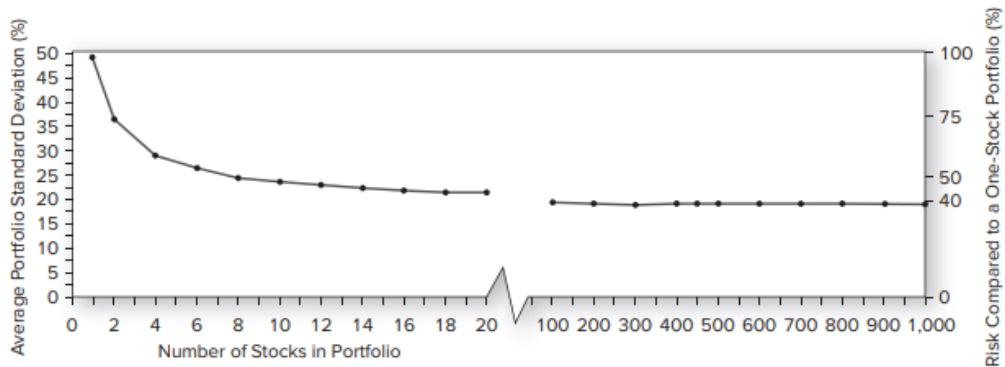
Source: Phillips P. J. (2009)⁸

1.2. The effects of diversification

As said in the first part of the chapter, the diversification is one of the most important characteristics of this model due to the fact that all the securities are subject to various sources of uncertainty. First, there is the risk caused by conditions in the general economy, such as business cycle, inflation, interest rates, and exchange rates. It is exceedingly difficult to precisely predict these factors, that jointly affect the rate of return of each stock in the market. In addition to these macroeconomic factors, there are the firm-specific influences. These asset-specific risks (idiosyncratic) can be compensated through diversification in the portfolio. Diversification consists in including additional securities in the portfolio. Obviously, when common sources of risk affect all securities, even extensive diversification cannot eliminate risk at all. However, it can be proved that the portfolio volatility falls as the number of securities increases, even though it cannot be reduced to zero. The risk that remains even after extensive diversification is called market risk, which is attributable to market-wide risk sources. This risk is also called systemic risk, or nondiversifiable risk. Conversely, the risk that can be eliminated by diversification is called unique risk, firm-specific risk, non-systematic risk, or diversifiable risk.

⁸ Phillips, P.J. (2009), *Applying Modern Portfolio Theory to the Analysis of Terrorism, Defence & Peace Economics*, 20, pp. 193-213.

Figure 6: Portfolio diversification



Source: Statman M. (1987), How Many Stocks Make a Diversified Portfolio?.

It is of interest to figure out how many assets are necessary to build a portfolio affected just by the market risk. Such analysis has been empirically implemented through data on NYSE stocks (Meir Statman⁹). Figure 6 shows that the average standard deviation of an equally weighted portfolio constructed by selecting stocks at random as a function of the number of stocks in the portfolio. On average, portfolio risk decreases as diversification takes place, but the ability of diversification to reduce risk is limited by systematic risk.

Given an introduction of the concept of diversification and the limits to the benefits of diversification, it is possible to show how this effect affects analytically the variance of a portfolio¹⁰. Recalling the general formula for the variance of a portfolio composed of n assets as

$$\sigma^2 = \sum_{i=1}^n \sum_{j=1}^n \sigma_{i,j} x_i x_j \quad (1.9)$$

Consider now the diversification strategy in which an equally weighted portfolio is constructed, meaning that $x_i = \frac{1}{n}$ for each security. In this case Equation 1.9 may be rewritten as¹¹

⁹ Statman M. (1987), *How Many Stocks Make a Diversified Portfolio?*, Journal of Financial and Quantitative Analysis 22.

¹⁰ Bodie, Z., Kane, A., Marcus, A. J. (2018), *Investments, Eleventh Edition*, McGraw-Hill Education, New York.

¹¹ Noting that when $i=j$, $Cov(r_i, r_j) = \sigma_{i,j} = \sigma_i^2$, i.e., the covariance equals the variance.

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n \frac{1}{n} \sigma_i^2 + \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{i=1}^n \frac{1}{n^2} \sigma_{i,j} \quad (1.10)$$

In Equation 1.10 there are n variance terms and $n(n-1)$ covariance terms.

Defining the average variance and average covariance of the assets as

$$\bar{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \sigma_i^2 \quad (1.11)$$

$$\overline{Cov} = \frac{1}{n(n-1)} \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{i=1}^n \sigma_{i,j} \quad (1.12)$$

The portfolio variance can be expressed as

$$\sigma^2 = \frac{1}{n} \bar{\sigma}^2 + \frac{n-1}{n} \overline{Cov} \quad (1.13)$$

Now it is possible to examine analytically the effect of the diversification. Equation 1.13 shows easily that when the average covariance (\overline{Cov}) among asset returns is zero, the portfolio variance only depends on the risk firm-specific, meaning that it can be arbitrarily reduced and therefore driven to zero. In fact, the second term of the equation will be zero in this case, while the first term approaches to zero as n increases. Therefore, when the security returns are uncorrelated, the power of diversification to reduce portfolio risk is unlimited.

However, zero correlation among assets is hard to be obtained and therefore the most important case study for diversification is when risk factors induce positive correlation among asset's returns. In this case, as the number of assets in the portfolio increases, portfolio variance remains positive. Firm-specific risk approaches to zero as diversification becomes higher (n gets larger), it is the first term in Equation 1.13, but the second term tends to the average covariance. This means that the systematic risk of a diversified portfolio depends on the average covariance of the returns.

Equation 1.13 can be expressed in terms of correlation (ρ)

$$\sigma_{i,j} = \rho_{i,j} \sigma_i \sigma_j \quad (1.14)$$

Plugging Equation 1.14 in Equation 1.13, the resulting formula is

$$\sigma^2 = \frac{1}{n} \bar{\sigma}^2 + \frac{n-1}{n} \rho_{i,j} \sigma_i \sigma_j \quad (1.15)$$

In Equation 1.15 the effect of correlation is explicit. As before, when $\rho = 0$, the portfolio variance approaches zero as the number of assets increases.

1.3. Role of correlation

Correlation, like diversification, plays a crucial role in this model. It has been shown that there is a relation between diversification and correlation useful to reduce the portfolio variance when $\rho = 0$, but not often the correlation between assets equals zero. In fact, the coefficient of correlation ranges between -1 and +1 ($-1 \leq \rho \leq +1$), and there are two extreme cases in which this value affects the shape of the efficient frontier, and they are when $\rho = \pm 1$.

To show the effect on the efficient frontier, it is necessary that the portfolio is composed by only two assets. In addition, it is necessary to hypothesize that:

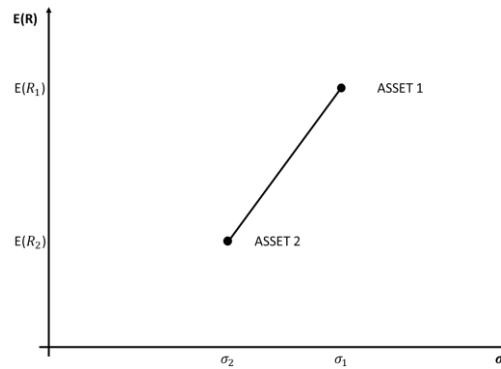
- R_1 and R_2 are respectively the expected returns of asset 1 and asset 2.
- r_1 and σ_1^2 are the mean and the variance of R_1 .
- r_2 and σ_2^2 are the mean and the variance of R_2 .
- $\sigma_{1,2} = \rho_{1,2} \sigma_1 \sigma_2$ is the covariance between R_1 and R_2 .
- $r_1 < r_2$ and $\sigma_1^2 < \sigma_2^2$ (or $r_1 > r_2$ and $\sigma_1^2 > \sigma_2^2$).

The last hypothesis is necessary to not have an asset that dominates the other. In fact, if $r_1 > r_2$ and $\sigma_1^2 < \sigma_2^2$ in this case asset 1 dominates the other based on the mean-variance criteria.

1.3.1. Coefficient of correlation between R_1 and R_2 equals +1 ($\rho_{1,2} = +1$)

When the coefficient of correlation is equal to +1, it indicates a perfect positive correlation between the two assets: as Asset 1 increases, asset 2 increases. As Asset 1 decreases, Asset 2 decreases.

Figure 7: Efficient frontier with perfect positive correlation



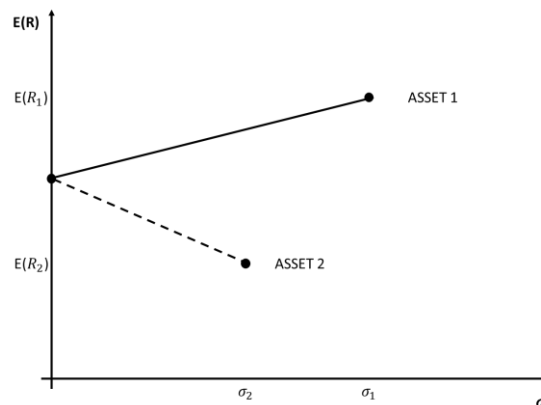
Source: Own production.

In Figure 7, it can be seen as the efficient frontier becomes a line that connects the two different assets. Is it possible, in case of perfect correlation, to reduce the risk? All portfolios represented in the efficient frontier have no less risk than the less risky asset; which means that diversification does not lead to a lower-risk portfolio.

1.3.2. Coefficient of correlation between R_1 and R_2 equals -1 ($\rho_{1,2} = -1$)

This second extreme case is the opposite situation of the perfect positive correlation, in fact, this is perfect negative correlation. As Asset 1 increases, Asset 2 decreases and vice versa.

Figure 8: Efficient frontier with perfect negative correlation



Source: Own production.

As can be observe in Figure 8, it is possible to distinguish the efficient frontier (continuous line) and the inefficient frontier (dashed line), and as said in the previous paragraphs all the efficient portfolios lie on the efficient part. The intercept is a theoretical portfolio composition, it does not exist, and it is risk-free.

Differently from the previous one, in this case, it is possible to obtain a reduction of the risks, thanks to diversification. In fact, any portfolio composition that lies on the efficient frontier and has volatility between 0 and σ_2 , has lower risk than asset 2 and a higher return.

1.4. Criticisms of the Modern Portfolio Theory

Despite its theoretical importance, critics of Modern Portfolio Theory argue whether it is an ideal investment tool since this model of financial markets does not match the real world in numerous ways.

Each individual has his/her own utility function that should be maximized through a quadratic utility function since, as Markowitz suggests, riskiness should only be described with the expected return and variance, i.e., the first and the second moments of the Gaussian distribution. Therefore, the implicit assumption is that the Skewness, the Kurtosis and all the subsequent moments have practical no impact on decision making. This is not credible since empirical evidence, such as the Jarque-Bera test to check whether a series is normally distributed, discards this hypothesis really often. In fact, in presence of Skewness and Kurtosis the non-normality is likely. The test is based on these two moments, and the test statistics reads as follows:

$$JB = \frac{n}{6} \left(S^2 + \frac{(K - 3)^2}{4} \right) \quad (1.16)$$

where n is the number of observations, or degrees of freedom, S is the Skewness and K is the Kurtosis.

When the test results different from zero means that there is non-normality.

It is worth noting that non-normality implies that the quadratic utility function is misspecified, due to the significant presence of Skewness and Kurtosis, which is omitted. Therefore, the decisions cannot be obtained by only using the first two moments of the return's distribution.

The second issue is mostly empirical. The estimate of the risk is never equal to its corresponding true value. In fact, even under unbiasedness, an estimator is affected by sampling errors (or estimation errors), that might be large with small samples. So, when applying the statistical formulas, the variance-covariance matrix might not be the correct

or precise one since it is an estimated matrix in which it is also included noise. In particular, empirical evidence suggested in this case that the mean-variance model tends to overweight assets that are characterized by:

- High expected returns.
- Negative correlation.
- Small value of variance.

When the choice is close to the optimal portfolio (which is on the efficient frontier), there are many statistically equivalent portfolios with similar expected return and variance (or standard deviation), but with vastly different weights. In case of revision of the portfolio, even a minimal change in parameters can lead to a heavy revision of the portfolio and, consequently, large transaction costs.

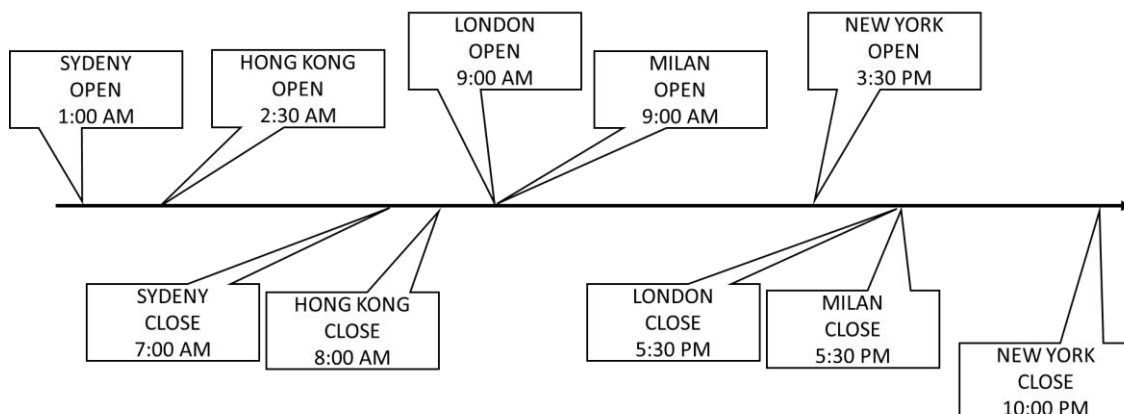
The alternative to the optimal portfolio is to take the equally-weighted portfolio and the Global Minimum-Variance (*GMV*) portfolio. Both portfolios have attractive characteristics; in fact, the *GMV* portfolio is efficient and the least risky. On the other side, the equally-weighted portfolio, generally, turns to be a good solution because of its expected return, which is higher than the *GMV* portfolio's expected return, moreover it is easy to construct. Furthermore, it has the advantage of being characterized by low transaction costs in the event of a revision of the portfolio and may not outperform the worst performing asset.

Chapter II Synchronization of the Returns

Given the importance of correlations among assets in the Markowitz model, it is necessary to analyze how correlation changes when the returns of the individual securities in the portfolio are not synchronized. Since the assets in the portfolio, as it will be shown in this thesis, but very often it is also found in reality, come from different stock exchanges, the daily prices, and consequently the returns, are asynchronous and this is due to the fact that when, for example, the New York Stock Exchange is open, the Tokyo Stock Exchange is closed. This leads not only to a distortion of the portfolio's value, but also to an incorrect calculation of risk measures and hedging strategies. And it is for this reason that in this thesis the problem of the synchronization of returns is discussed, in order to obtain an “*adjusted*” correlation matrix, not sensitive to asynchronous observations, between the different stocks that should be used to get correct estimates.

For this work, the chosen assets come from five Stock Exchanges, precisely from the exchanges of Sydney, Hong Kong, London, Milan, and New York. As can be seen in Figure 9, each exchange has its own opening and closing time, for example, when it is 10:00 AM in Sydney and the stock exchange opens, it is 1:00 AM in Italy (Milan); at 4:00 PM in Sydney the exchange closes, that corresponds to 7:00 AM in Italy, where markets are not open yet. Consider, for example, a day when the Italian Market falls 1% after Sydney closes. Valuing Sydney stocks at the closing price is highly inaccurate. It will result in the Italian share of the portfolio declining today, while the Sydney share might receive the impact of bad news in Milan tomorrow. The same reasoning can be applied with respect to other countries.

Figure 9: Asynchronous dynamics of closing prices in Stock Exchanges



Source: Own production.

Aiming at obtaining reliable estimates, the general approach is to recognize that even when markets are closed, the asset value may change and that new values can be estimated for use before the market reopens. Value at risk calculations, asset allocations and hedging decisions should all be considered based on the estimated value of portfolio. Considering the information available to the market, the estimated prices should have the property that they are unbiased; the average returns are unchanged, only the timing is different.

Following Burns, Engle and Mezrich (1998), the synchronized log-price for asset j is defined as

$$\log S_{t,j}^s = E \left[\log S_{t,j} \mid F_t \right] \quad (2.1)$$

where t_j is the closing price for market j and where F_t includes all the prices $S_{t_j,j}$ observed *at or before* instant t ¹². Clearly, if S is observed at t , then its expectation is just this value. Equation 2.1 explains that if the market closed before t , the past prices and all the market that have subsequently closed, are potentially useful in predicting S at time t .

It is assumed that synchronized prices are unbiased predictors for the next recorded price

$$\log S_{t,j}^s = E \left[\log S_{t,j} \mid F_t \right] = E \left[\log S_{t_{j+1},j} \mid F_t \right] \quad (2.2)$$

Generally, returns for day t for n assets can be calculated as

$$r_t^0 = \begin{pmatrix} \log \left(\frac{S_{t,1}}{S_{t-1,1}} \right) \\ \vdots \\ \log \left(\frac{S_{t,n}}{S_{t-1,n}} \right) \end{pmatrix} \quad (2.3)$$

Synchronized returns are defined as

$$r_t^s = \begin{pmatrix} \log S_{t,1}^s - \log S_{t-1,1}^s \\ \vdots \\ \log S_{t,n}^s - \log S_{t-1,n}^s \end{pmatrix} \quad (2.4)$$

¹² F_t is defined as $F_t = \{S_{t_i,i} \mid t_i \leq t, \forall i\}$

After several mathematical passages and simplification that can be found in Burns, Engle and Mezrich (1998)¹³, synchronized returns can be defined in matrix notation as

$$r_t^s = r_t^0 + E_t[r_{t+1}^0] - E_{t-1}[r_t^0] \quad (2.5)$$

Equation 2.5 proves that synchronized returns are equivalent to asynchronous returns plus a correction term. To make the synchronization possible, there is the need to define a statistical model able to estimate $E_t[r_{t+1}^0]$. As Burns, Engle and Mezrich (1998) in their Discussion paper *Correlations and Volatilities of Asynchronous Data* suggested, it can be done using a Moving Average Model, to account for the lag 1 autocorrelation implied by the asynchronicity effect. In this work, the MA model is represented by a State-Space Model.

2.1. State-Space Model

The State Space Model (SSM) refers to a probabilistic model class which describes the probabilistic dependence between the latent state variable and the observed measurement. The SSM offers a general framework for the analysis of deterministic and stochastic systems which are observed through a sequence of observations, consistent with a stochastic process trajectory. In fact, SSM is used to gain a computational advantage in estimating the Maximum Likelihood function by defining a stochastic model in a more complex way that will result more efficient for the computational issues.

A State Space Model for a multivariate time series y_t consists of a measurement equation relating the observed data to a non-observable state vector α_t , and a transition equation that defines the evolution of the state vector over time.

The general measurement equation of a State Space Model can be written as

$$y_t = Z_t \alpha_t + d_t + \epsilon_t \text{ with } \epsilon_t \sim N(0, \sigma^2) \quad (2.6)$$

The transition equation for the state vector α_t is

$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t \text{ with } \eta_t \sim N(0, \sigma^2) \quad (2.7)$$

¹³ Burns, P., Engle, R., Mezrich, J., (1998), *Correlations and Volatilities of Asynchronous Data*, University of California, Discussion Paper 97-30R.

It is assumed that the measurement equation errors ϵ_t are independent of the transition equation errors. The state space representation is completed by specifying the behaviour of the initial state $\alpha_0 \sim N(\alpha_0, P_0)$. If y_t is covariance stationary, then the state space system will be time invariant.

Consider the standard 1st order Moving Average, MA(1)

$$y_t = \mu + \eta_t + \theta\eta_{t-1}, \quad \eta_t \sim N(0, \sigma^2) \quad (2.8)$$

Define $\alpha_t = (y_t, \eta_t)$, so that the transition equation is

$$\alpha_t = \begin{pmatrix} y_t \\ \eta_t \end{pmatrix} = \begin{bmatrix} 0 & \theta \\ 0 & 0 \end{bmatrix} \begin{pmatrix} y_{t-1} \\ \eta_{t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \eta_t \quad (2.9)$$

In this way, it is possible to define the measurement equation as

$$y_t = (1, 0)\alpha_t + \mu \quad (2.10)$$

Finally, this is the Model in state space form that has been used to obtain the synchronized returns and the adjusted correlation matrix for the calculation of the measures of risk. An important tool, which can be used to determine which are the optimal estimates of the state vector α_t and to evaluate the Likelihood function, is the Kalman filter. This filter is a set of recursive formulas.

Chapter III Extreme Value Theory

After completing the analysis and the synchronization of the returns, it is possible to apply the Extreme Value Theory (EVT) to the portfolio of stocks. This theory is useful for measuring the magnitudes and the probabilities of significant potential losses due to extreme events such as currency crisis, trading scandals or stock market crashes. The methods of Extreme Value Theory concentrate on modelling the tail behaviour of a loss distribution using only extreme values rather than all of the data.

In this chapter are discussed the two main types of models used for the extreme value theory. The first group of models refer to the *Block Maxima* models, while a more recent approach is labelled as *Peaks-Over-Threshold (POT)*. In addition to the definition of these models, the way in which they can be estimated through the maximum likelihood function will also be analyzed. Finally, thanks to the estimation of these models, it will be possible to calculate the main risk measures such as *Value-at-Risk (VaR)* and *Expected Shortfall (ES)*.

3.1. Modelling Block Maxima

Before describing the family of Block Maxima models, it is necessary to state the main result underlying this theory, i.e., the Fisher-Tippet Theorem. This theorem suggests a suitable transformation of the maxima that converges to a particular distribution called *Generalized Extreme Value (GEV)* distribution, useful for the description of the Block Maxima models.

3.1.1. The Fisher-Tippet Theorem and the GEV Distribution

Considering a sequence of independent and identically distributed (*i.i.d.*) random variables X_1, X_2, \dots, X_n representing risks or losses with an unknown cumulative distribution function (CDF) of the type

$$F(x) = \Pr\{X_i \leq x\} \quad (3.1)$$

In general, for financial application the random variables X_i describe negative returns or losses on a financial asset or portfolio. Note that in this thesis losses are treated as a positive numbers and extreme events occur when losses take values in the right tail of the loss distribution F . The worst-case loss in a sample of n losses can be defined as

$$M_n = \max(X_1, \dots, X_n) \quad (3.2)$$

From the *i.i.d.* assumption, the CDF of M_n is given by

$$\Pr\{M_n \leq x\} = \Pr\{X_1 \leq x, \dots, X_n \leq x\} = \prod_{i=1}^n F(x) = F^n(x) \quad (3.3)$$

Since F^n is unknown and the empirical distribution function is an extremely poor estimator of $F^n(x)$, the Fisher-Tippet Theorem¹⁴ provides a solution to make inference on M_n . In addition, since $F^n(x) \rightarrow 0$ or 1 as $n \rightarrow \infty$ and x is fixed, since it describes a probability raised to power n , it is convenient to study the asymptotic approximation of F^n is based on the following standardization

$$Z_n = \frac{M_n - \mu_n}{\sigma_n} \quad (3.4)$$

where μ_n and $\sigma_n > 0$ are sequences of real numbers, and they can be interpreted, respectively, as a location measure and a scale measure, such that

$$\lim_{n \rightarrow \infty} \Pr(Z_n \leq x) = \lim_{n \rightarrow \infty} F^n(\sigma_n x + \mu_n) = H(x) \quad (3.5)$$

where $H(x)$ is a non-degenerate distribution function. If the standardized value expressed in Equation 3.4 converges to some non-degenerate distribution function, it must be a *Generalized Extreme Value (GEV)* distribution of the form

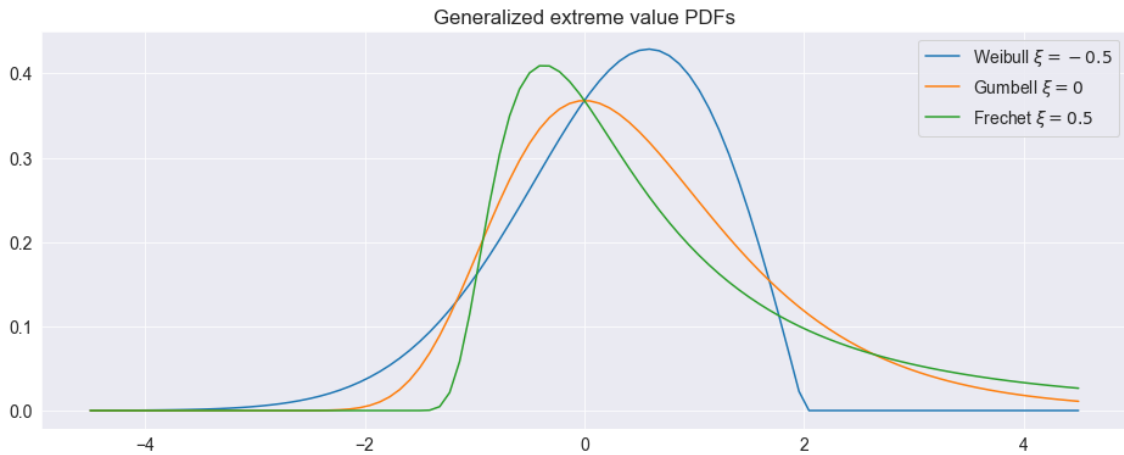
$$H_\xi(z) = \begin{cases} \exp\{-(1 + \xi z)^{-1/\xi}\} & \xi \neq 0, 1 + \xi z > 0 \\ \exp\{-\exp(-z)\} & \xi = 0, -\infty \leq z \leq \infty \end{cases} \quad (3.6)$$

If Z_n converges to the limit in equation 3.6, then the CDF F of the underlying data is in the domain of attraction of H_ξ . The parameter ξ is a shape parameter and determines the tail behavior of H_ξ . When $\xi = 0$ then H_ξ is of the *Gumbel type*, distributions in the domain of attraction of the Gumbel type are thin tailed distributions such as the normal, log-normal, exponential, and gamma, and for all these distributions it is possible to estimate all the moments. When $\xi > 0$, H_ξ is of the *Fréchet type*, that are thick tailed such as the Pareto, and Student-t distributions. When $\xi < 0$, H_ξ is of the *Weibull type* and the tail is finite. Distribution in the domain of attraction of the Weibull type are uniform and

¹⁴ Fisher, R.A., Tippet, L.H.C., (1928), *Limiting Forms of Frequency Distribution of the Largest or Smallest Members of a Sample*, Proceedings of the Cambridge Philosophical Society, 24, 180-190

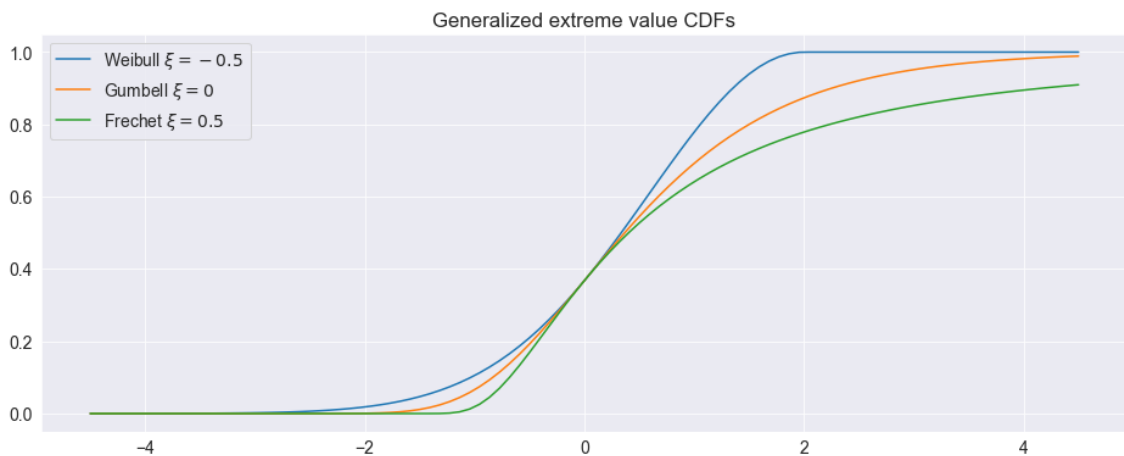
beta distributions. Typically, when dealing with financial losses or negative returns, the distribution is expected to be of the Fréchet type.

Figure 10: Generalized Extreme Value PDFs H_ξ for Fréchet ($\xi = 0.5$), Weibull ($\xi = -0.5$) and Gumbell ($\xi = 0$) with $\mu = 0$ and $\sigma = 1$



Source: Own production (Python code)

Figure 11: Generalized Extreme Value CDFs H_ξ for Fréchet ($\xi = 0.5$), Weibull ($\xi = -0.5$) and Gumbell ($\xi = 0$) with $\mu = 0$ and $\sigma = 1$



Source: Own production from Python

A nice property of the GEV distribution states it is invariant with respect to linear transformations, and in fact

$$H_\xi(z) = H_\xi\left(\frac{x - \mu}{\sigma}\right) = H_{\xi, \mu, \sigma}(x) \quad (3.7)$$

Therefore, the Fisher-Tippett Theorem states that, for large n , the distribution of M_n is a GEV, and this is given by

$$\Pr(Z_n < z) = \Pr\left(\frac{M_n - \mu_n}{\sigma_n} < z\right) \approx H_\xi(z) \quad (3.8)$$

3.1.2. Estimating the GEV distribution by the Maximum Likelihood

Once the Fisher-Tippett has been described and analyzed, it can be used practically by estimating the model's parameters using parametric methods such as the maximum likelihood.

Let X_1, \dots, X_T be identically distributed losses from a sample of size T with unknown CDF F and let M_T denote the sample maximum. For inference on M_T the parameters ξ , σ_T and μ_T must be estimated. Since there is only one value of M_T for the entire sample, it is not possible to use a likelihood function for computing the parameters' estimates. However, if interest is on the maximum of X over a large finite subsample or block size $n < T$, M_n , implementing the Block Maxima, it is possible to use the likelihood function for the parameters ξ , σ_n and μ_n . The general idea behind Block Maxima models is to split the whole data sample in blocks of equal size and find the maximum value on each block. In this way, a sequence of *extreme* observations is obtained, and can be used as a sample to evaluate the GEV's parameters. For the application of this model, it is assumed that the block size n is sufficiently large so that the Fisher-Tippett Theorem holds.

The sample is divided into m non-overlapping blocks of equal size $n = \frac{T}{m}$

$$[X_1, \dots, X_n | X_{n+1}, \dots, X_{2n} | \dots | X_{(m-1)n+1}, \dots, X_{mn}] \quad (3.9)$$

and $M_n^{(j)}$ is defined as the maximum value of X_i in block $j = 1, \dots, m$.

The log likelihood function assuming *i.i.d.* observations from a GEV distribution with $\xi \neq 0$ is

$$\begin{aligned}
l(\mu, \sigma, \xi) = & -m \ln(\sigma) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^m \ln \left[1 + \xi \left(\frac{M_n^{(i)} - \mu}{\sigma} \right) \right] \\
& - \sum_{i=1}^m \left[1 + \xi \left(\frac{M_n^{(i)} - \mu}{\sigma} \right) \right]^{-1/\xi}
\end{aligned} \tag{3.10}$$

such that

$$1 + \xi \left(\frac{M_n^{(i)} - \mu}{\sigma} \right) > 0 \tag{3.11}$$

The log-likelihood for the case $\xi = 0$ (Gumbel family) is

$$\begin{aligned}
l(\mu, \sigma, \xi) = & -m \ln(\sigma) - \sum_{i=1}^m \left(\frac{M_n^{(i)} - \mu}{\sigma} \right) \\
& - \sum_{i=1}^m \exp \left[- \left(\frac{M_n^{(i)} - \mu}{\sigma} \right) \right]
\end{aligned} \tag{3.12}$$

For $\xi > -0.5$ the Maximum Likelihood Estimates for the location, scale and shape parameters are consistent and asymptotically normally distributed with asymptotic variance given by the inverse of the observed information matrix.

The properties of the MLE will depend on the number of blocks (sample size) and the block size (convergence achieved by the Fisher-Tippett theorem). There is a trade-off between bias and variance. It can be proved that the bias of the MLE is reduced by increasing the block size, and the variance of the MLE is reduced increasing the number of blocks, i.e., the number of observations in the sample.

3.1.3. Return Level

A useful risk measure for block maxima that is related to a high quantile is the *return level*. First, for $\alpha \in (0,1)$ a quantile of a continuous distribution with distribution function F is the value q_0 such that

$$q_0 = F^{-1}(\alpha) \tag{3.13}$$

The k n -block return level, $R_{n,k}$, can be defined as the level which exceeded in one out of every k block of size n on average. Formally, the k n -block return level is the loss value such that

$$\Pr\{M_n > R_{n,k}\} = \frac{1}{k} \quad (3.14)$$

The n -block in which the return level is exceeded is called *stress period*. Note that, by construction, $R_{n,k}$ is simply the $1 - \frac{1}{k}$ quantile of $H_\xi(\mu, \sigma)$ and can be analytically computed from

$$R_{n,k} \approx H_{\xi, \mu, \sigma}^{-1}\left(1 - \frac{1}{k}\right) = \mu - \frac{\sigma}{\xi} \left(1 - \left(-\ln\left(1 - \frac{1}{k}\right)\right)^{-\xi}\right) \quad (3.15)$$

Using the Maximum Likelihood Estimators for the parameters ξ, μ and σ , the MLE for $R_{n,k}$ is given by

$$\hat{R}_{n,k} = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left(1 - \left(-\log\left(1 - \frac{1}{k}\right)\right)^{-\hat{\xi}}\right) \quad (3.16)$$

3.1.4. Return Period

The average number of blocks passed before observing an extreme event $M_n > u$, for a given u , is the k n -block *return period*, $k_{n,u}$.

Given that $H_{\xi, \mu, \sigma} = \Pr(M_n < u)$, the event $Y = (M_n > u)$ is observed for the first time after k blocks has probability

$$\Pr(Y = k) = H_{\xi, \mu, \sigma}^{k-1} (1 - H_{\xi, \mu, \sigma}) \quad (3.17)$$

This is a geometric distribution with average $\frac{1}{1 - H_{\xi, \mu, \sigma}}$, and therefore the k n -block return period is given by

$$k_{n,u} = \frac{1}{1 - H_{\xi, \mu, \sigma}(u)} \quad (3.18)$$

3.2. Peaks Over Thresholds (POT)

Modelling data only using Block Maxima is inefficient when other data on extreme values are available, since the sample size corresponds to the number of blocks, and therefore the method requires a huge time series to be effective. A more efficient alternative approach that uses more data is to model the behaviour of extreme values above some pre-defined high threshold. This approach is known as *Peaks Over Thresholds (POT)*. An advantage of this approach, in addition of being a more efficient

model, is that common risk measures as *Value-at-Risk* and *Expected Shortfall* may be easily computed.

3.2.1. The Limiting Distribution of Extremes Over High Threshold and the GPD distribution

As for the analysis of the Block Maxima, let X_1, X_2, \dots, X_n be a sequence of *i.i.d.*¹⁵ random variables representing risks or losses with an unknown CDF F and let, as in the previous chapter, M_n be the sample maximum¹⁶. A measure of extreme events are values of the X_i that exceed a high threshold u . The *excess distribution* above the threshold u can be defined as the conditional probability

$$F_u(y) = \Pr\{X - u \leq y | X > u\} = \frac{F(y + u) - F(u)}{1 - F(u)}, y > 0 \quad (3.19)$$

It can be shown that, for the family of distribution F such that the CDF of the standardized value of M_n converges to a GEV distribution, for large enough u there exists a positive parameter $\beta(u)$, that depends on the value u , such that the excess distribution is given in equation 3.19 can be well approximated by the *Generalized Pareto Distribution (GPD)*

$$G_{\xi, \beta(u)}(y) = \begin{cases} 1 - \left(1 + \left(\frac{\xi y}{\beta(u)}\right)^{-\frac{1}{\xi}}\right) & \text{for } \xi \neq 0, \beta(u) > 0 \\ 1 - \exp\left(-\frac{y}{\beta(u)}\right) & \text{for } \xi = 0 \end{cases} \quad (3.20)$$

this is defined for $y \geq 0$ when $\xi \geq 0$ and $0 \leq y \leq -\frac{\beta(u)}{\xi}$ when $\xi < 0$. This means that $F_u(y) \approx G_{\xi, \beta(u)}(y)$ for a sufficiently large values of u . To achieve this result, the value of the threshold must be defined and the estimates of the unknown parameters ξ and $\beta(u)$ must be computed.

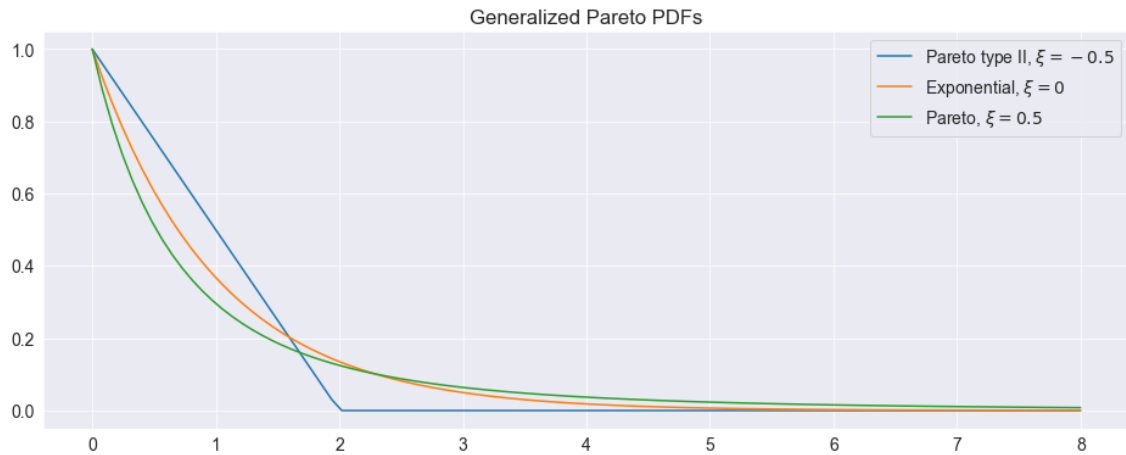
It can be shown that there is a close connection between the GEV distribution for block maxima and the GPD for the threshold excesses. For a given value of the threshold u , the parameters ξ , μ and σ of the GEV distribution determine the parameters ξ and $\beta(u)$. Specifically, ξ , or rather the shape parameter of the GEV distribution, is the same shape parameter ξ in the GPD, independently from the threshold value u . Hence, when

¹⁵ Remind that *i.i.d.* means Independent and Identically Distributed

¹⁶ $M_n = \max\{X_1, X_2, \dots, X_n\}$

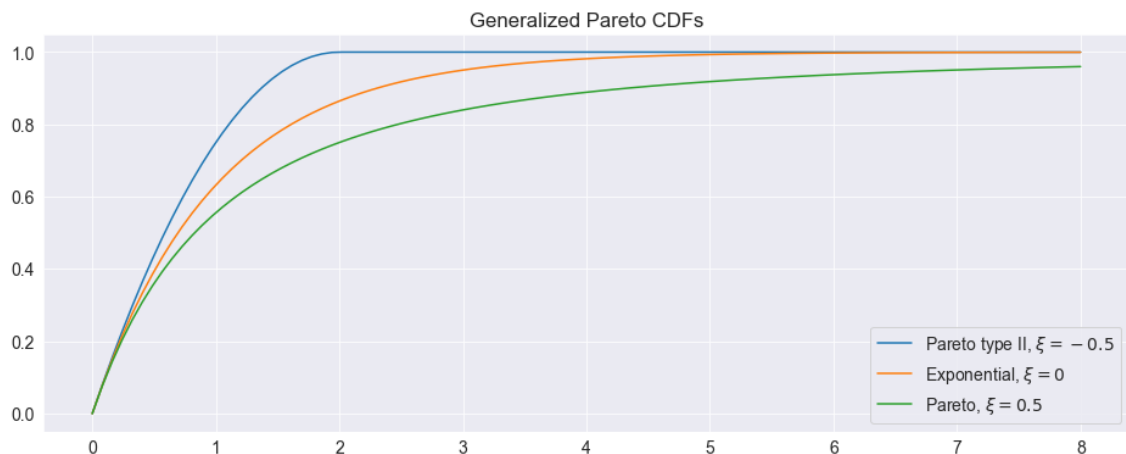
$\xi < 0$ then F is of the Weibull type and $G_{\xi,\beta(u)}$ is a *Pareto type II distribution*; when $\xi = 0$ then F is of the Gumbel type and $G_{\xi,\beta(u)}$ is an *exponential distribution*; and the most interesting case for risk management purposes is when $\xi > 0$ then F is in the Fréchet family and $G_{\xi,\beta(u)}$ is a *Pareto distribution*. Furthermore, it can be shown that $E[X^k] = \infty$ for $k \geq \alpha = \frac{1}{\xi}$.

Figure 12: Generalized Pareto PDFs for Pareto ($\xi = 0.5$), Pareto Type II ($\xi = -0.5$) and exponential ($\xi = 0$) with $\mu = 0$ and $\sigma = 1$



Source: Own production (Python Code)

Figure 13: Generalized Pareto CDFs for Pareto ($\xi = 0.5$), Pareto Type II ($\xi = -0.5$) and exponential ($\xi = 0$) with $\mu = 0$ and $\sigma = 1$



Source: Own production (Python Code)

The evaluation of the threshold value u , from which the asymptotic Pareto approximation should work, is one of the key issues in the POT method. Some empirical

research¹⁷, suggested to select the value of the threshold u such that $P(X > u) \approx 10\% - 15\%$.

An alternative method is to study the *mean excess function*. Suppose the threshold excess $X - u_0$ follows a GPD with parameters $\xi < 1$ and $\beta(u_0)$. The mean excess over the threshold u_0 is defined as

$$E[X - u_0 | X > u_0] = \frac{\beta(u_0)}{1 - \xi} \quad (3.21)$$

Furthermore, for any $u > u_0$, the *mean excess function* $e(u)$ can be defined as

$$e(u) = E[X - u | X > u] = \frac{\beta(u_0) + \xi(u - u_0)}{1 - \xi} \quad (3.22)$$

Alternatively, for any $y > 0$

$$e(u_0 + y) = E[X - (u_0 + y) | X > u_0 + y] = \frac{\beta(u_0) + \xi y}{1 - \xi} \quad (3.23)$$

Note that the mean excess function $e(u)$ is a *linear function* of $y = u - u_0$, for a given value of ξ . Therefore, a graphical method can be used to make inference on the optimal value for u . The *empirical* mean excess function is defined as

$$e_n(u) = \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) \quad (3.24)$$

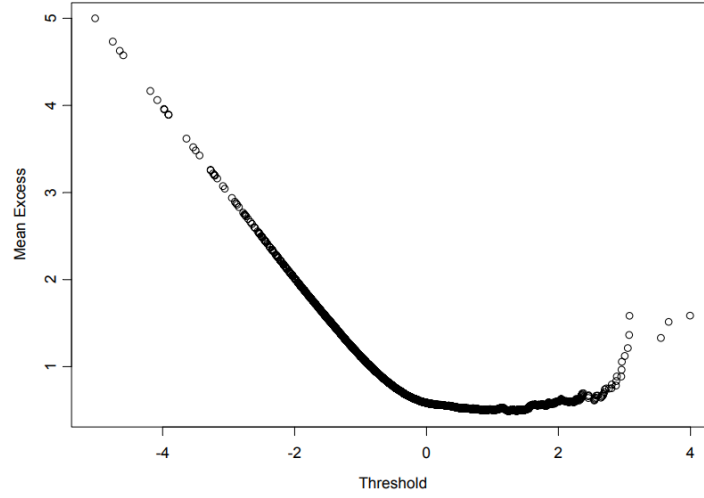
where $x_{(i)}$ ¹⁸ are the values of x_i such that $x_i > u$. Figure 14 shows the mean excess plot that is a plot of $e_n(u)$ against u and should be linear in u for $u > u_0$. An upward sloping plot indicates heavy-tailed behavior. Especially, a sign of Pareto behaviour in tail is designed by a straight line with positive slope above u_0 . Therefore, the threshold u can be chosen as the smaller value from which $e_n(u)$ is linear (and possibly increasing).

Conversely, a downward trend shows thin-tailed behaviour, while a line with zero slope shows an exponential tail.

¹⁷ Zivot, E., Wang, J., (2006), *Modelling Financial Time Series with S-PLUS*, Ch. 5, Springer.

¹⁸ With $i = (1, \dots, n_u)$

Figure 14: Example of empirical mean excess plot for the S&P 500 negative returns.



Source: Zivot, E., Wang, J., (2006), Modelling Financial Time Series with S-Plus.

In Figure 14, it can be seen as for the S&P 500 negative returns the empirical mean excess plot is linear and with positive slope for $u > 1$, i.e., Pareto tail behaviour.

3.2.2. Estimating the GPD by the Maximum Likelihood

Once u is set, it is possible to estimate the GPD using Maximum Likelihood. Let x_1, \dots, x_T be an i.i.d. sequence of losses with unknown CDF F . Extremes are those x_i value for which $x_i - u > 0$. These *ordered* extreme values can be defined as $x_{(1)}, \dots, x_{(k)}$ and the threshold excess as $y_i = x_{(i)} - u$ for $i = 1, \dots, k$. The GPD log-likelihood function, for $\xi \neq 0$, is defined as

$$l(\xi, \beta(u)) = -k \ln(\beta(u)) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^k \ln\left(1 + \frac{\xi y_i}{\beta(u)}\right) \quad (3.25)$$

on condition that $y_i \geq 0$ when $\xi > 0$ and $0 \leq y_i \leq -\frac{\beta(u)}{\xi}$ when $\xi < 0$.

3.2.3. Estimating the Tails of the Loss Distribution

Note that for sufficiently high threshold u , $F_u(y) \approx G_{\xi, \beta(u)}(y)$, and suppose u is set, as described above. Setting $x = u + y$ and using the equation 3.19, the tails of the unknown loss distribution $F(x)$ for $x > u$ can be approximated as

$$F(x) = (1 - F(u))G_{\xi, \beta(u)}(y) + F_u(y) \quad (3.26)$$

Using the empirical CDF, it is possible to estimate non parametrically the value of $F(u)$

$$\hat{F}(u) = \frac{(n - k)}{n} \quad (3.27)$$

where n is the sample size and k denotes the number of exceedences over the threshold u . Combining the formulas, the result gives an estimator for the tail of the loss distribution $F(x)$

$$\hat{F}(x) = 1 - \frac{k}{n} \left(1 + \hat{\xi} \frac{x - u}{\hat{\beta}(u)} \right)^{-\frac{1}{\hat{\xi}}} \quad (3.28)$$

where $\hat{\xi}$ and $\hat{\beta}(u)$ are the Maximum Likelihood Estimators of ξ and $\beta(u)$, respectively. It is worth noting that $\hat{F}(x)$ is an estimate of the negative returns and it can be used to compute risk measures as VaR and ES .

3.3. Risk Measures

The extremes over high threshold methods are really useful to evaluate two common risk measures: Value-at-Risk (VaR) and Expected Shortfall (ES). VaR is a high quantile of the loss distribution. That is, for $0.95 \leq \alpha \leq 1$, VaR_α is the α th quantile of the distribution F

$$VaR_\alpha = F^{-1}(\alpha) \quad (3.29)$$

where F^{-1} is the inverse of F . An estimate of equation 3.29 based on inverting the tail estimation equation 3.28, for a given probability $\alpha > F(u)$, is

$$\widehat{VaR}_\alpha = u + \frac{\hat{\beta}(u)}{\hat{\xi}} \left(\left(\frac{n}{k} (1 - \alpha) \right)^{-\hat{\xi}} - 1 \right) \quad (3.30)$$

Expected shortfall is the expected loss size, given that VaR_α is exceeded

$$ES_\alpha = E[X|X > VaR_\alpha] \quad (3.31)$$

The VaR_α and the ES_α are strictly related by

$$ES_\alpha = VaR_\alpha + E[X - VaR_\alpha | X > VaR_\alpha] \quad (3.32)$$

where the second term in equation 3.32 is the mean of the excess distribution $F_{VaR_\alpha}(y)$ over the threshold VaR_α . By the properties of the GPD distribution, the approximation of GPD to $F_{VaR_\alpha}(y)$ has shape parameter ξ and scale parameter $\beta(u) + \xi(VaR_\alpha - u)$. Then, provided $\xi < 1$, gives

$$E[X - VaR_\alpha | X > VaR_\alpha] = \frac{\beta(u) + \xi(VaR_\alpha - u)}{1 - \xi} \quad (3.33)$$

Combining the previous results, the GPD approximation to ES_α is given by

$$\widehat{ES}_\alpha = \frac{\widehat{VaR}_\alpha}{1 - \hat{\xi}} + \frac{\hat{\beta}(u) + \hat{\xi}u}{1 - \hat{\xi}} \quad (3.34)$$

Chapter IV Real Data Application

In this chapter, it will be shown the empirical analysis based on some software written through Python and taking advantage of the libraries `Pandas` and `Statsmodels`. A portfolio, composed by 10 stocks from all over the world with a time horizon of 7-years is considered, i.e., from January 2015 to December 2021, on which it will be applied the Extreme Value Theory to compute the VaR and the ES.

The stocks considered come mainly from the Australian, Chinese, Italian, British and American stocks exchanges. In particular, BHP Billiton Ltd and Sonic Healthcare Ltd are Australian firms; Wuliangye Yibing Co Ltd and Wanhua Chemical Group are listed in the Chinese stock exchange; Leonardo S.p.A. and Hera S.p.A. are Italian firms; Phoenix Group Holdings PLC and Barclays PLC are British companies; and Intel Corporation and Tesla Inc are listed in the American stock exchange.

Table 1: Descriptive statistics

	BHP Billiton Ltd (BHP)	Sonic Healthcare Ltd (SHL)	Leonardo Spa (LDOF)	Hera Spa (HRA)	Phoenix Group Holdings PLC (PHNX)	Barclays PLC (BARC)	Intel Corporation (INTC)	Tesla Inc (TSLA)	Wuliangye Yibin Co Ltd (000858)	Wanhua Chemical Group Co Ltd (600309)
Mean	0.0004620	0.0006900	0.0002150	0.0005300	0.0001710	0.0001320	0.0004470	0.0026820	0.0017390	0.0013820
Variance	0.0004060	0.0002330	0.0006300	0.0002500	0.0002720	0.0006590	0.0004350	0.0013250	0.0006540	0.0007170
St. Deviation	0.0201494	0.0152643	0.0250998	0.0158114	0.0164924	0.0256710	0.0208567	0.0364005	0.0255734	0.0267769

Source: Own production

Table 1 provides the mean, the variance and the standard deviation of the different stocks computed using Python.

The idea is to evaluate, in terms of portfolio selection, if synchronization is actually an issue. For this reason, some Python routines have been created to compare the different behaviours of the portfolio as the initial assumption change: a first case in which the returns of the stocks are not synchronized, while in the following two the synchronization of the returns has been considered¹⁹, but with a small difference in the imputation of data between the two cases (that refers to problem of dealing with missing data). In the first code of synchronized returns, the *'inner'* parameter is used within the Python *'pd.merge'* function with the aim of creating a database that only includes the days in which all the stocks have a price/observation. In the second case of

¹⁹ In this code, it is used the *'inner'* parameter in the *'pd.merge'* function.

synchronization, instead, 'outer' is used, i.e., with the possibility of replacing a missing value with the previous value, doing this it is possible to account for all the available days in the sample. A missing value is observed, for example, when a particular stock exchange is closed because of a national holiday, for instance, while all the others are open. A real example is December 31st, when the Italian stock exchange is closed, while the Chinese exchange is open. In fact, as it can be noted from the number of rows of the prices' database using the parameter 'inner', 1598 daily prices are used for the calculation of the returns, while in the case of 'outer' the number of prices used are 1821.

For clarity, these three different codes will be identified as *Case 1*, *Case 2*, and *Case 3*, respectively²⁰.

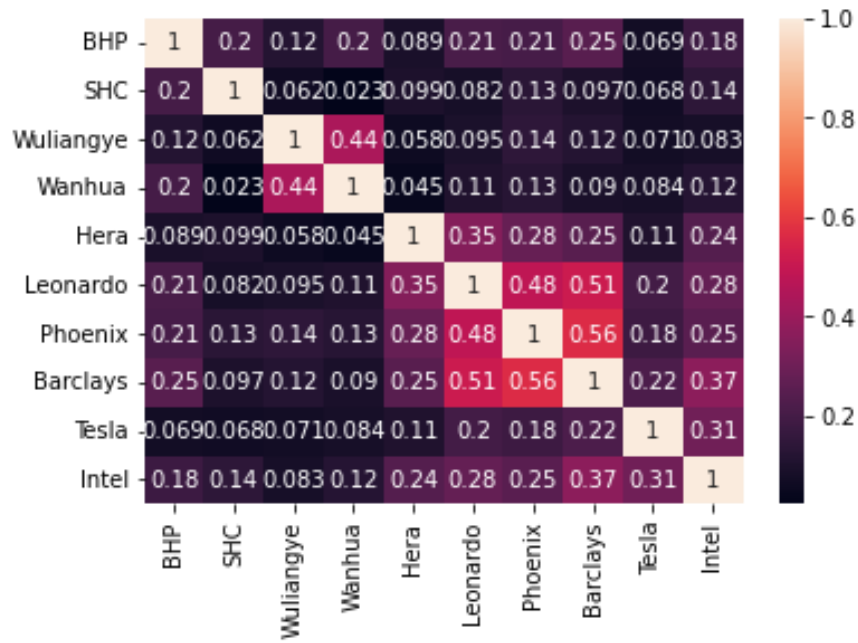
After importing the prices of the different stocks and creating a single database, the daily returns of the stocks are calculated using the following formula

$$R_t = \left(\frac{P_t - P_{t-1}}{P_{t-1}} \right) * 100 \quad (4.1)$$

Subsequently, after calculating the returns, the correlation matrix of the returns was created, thus finding the first differences between the different cases studied in this thesis.

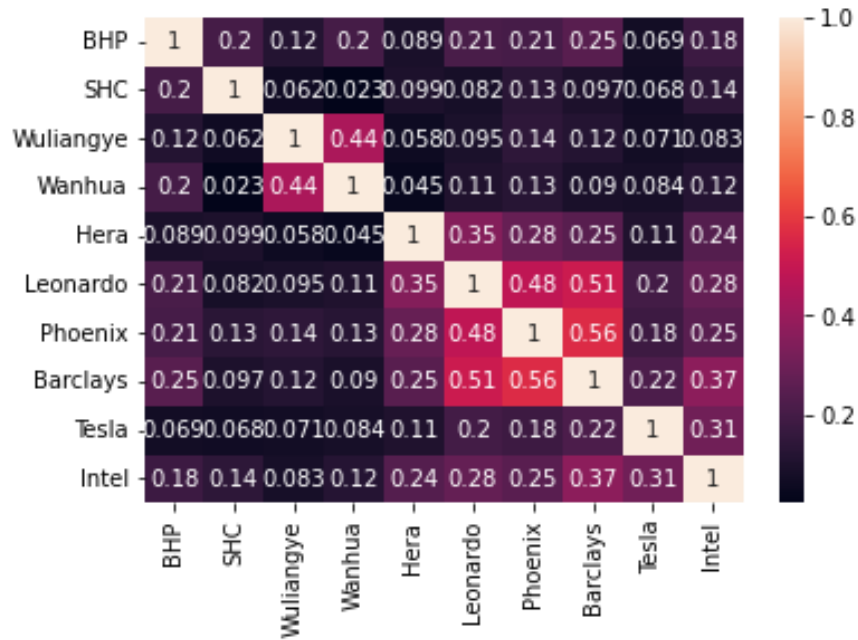
²⁰ *Case 1*: not synchronized returns; *Case 2*: synchronized returns with 'inner'; *Case 3*: synchronized returns with 'outer'.

Figure 15: Case 1 – Correlation Matrix of Returns



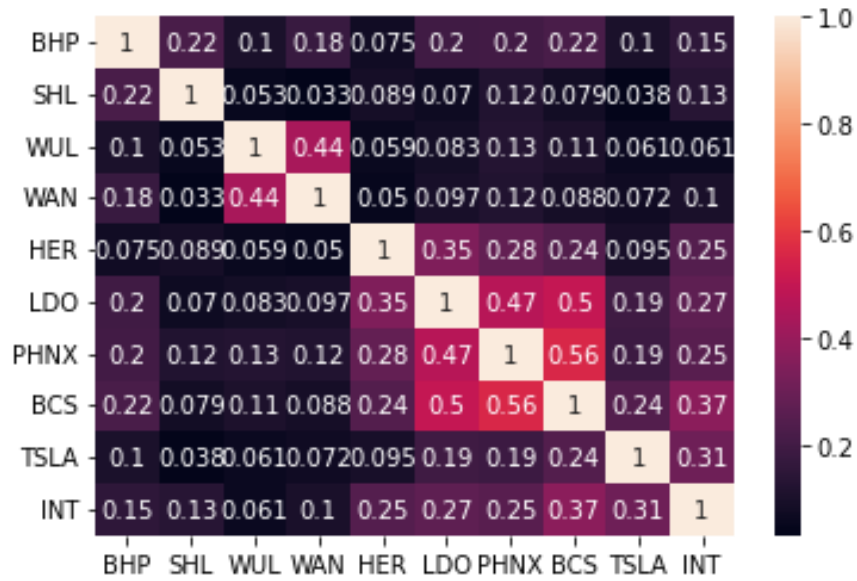
Source: Own production (Python Code)

Figure 16: Case 2 – Correlation Matrix of Return with 'inner' parameter



Source: Own production (Python Code)

Figure 17: *Case 3* – Correlation Matrix of Returns with ‘outer’ parameter



Source: Own production (Python Code)

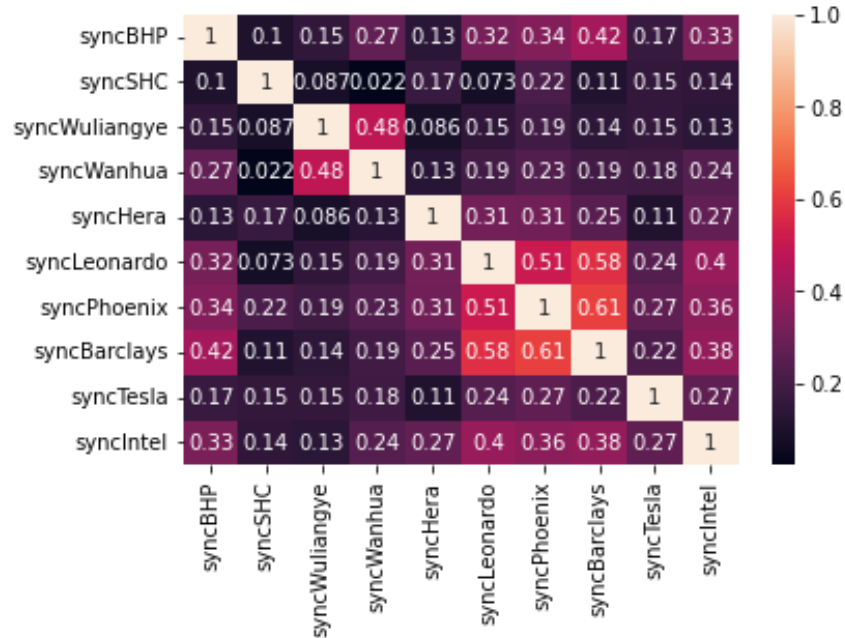
Correlation matrices of *Case 1* and *Case 2*, as can be noticed, respectively, from Figure 15 and Figure 16, are identical and this is due to the same ‘inner’ parameter used for the imputation of the data. Instead, comparing the first two matrices with the third one, it is possible to note that in some cases the correlation between returns, observable in Figure 17, is greater or lower. Consider the correlation, for example, between *Tesla* and *Barclays*, in cases 1 and 2 it is equal to 0.22, while in *Case 3* the value of the correlation is greater and equal to 0.24. conversely, comparing the correlation between *Phoenix (PHNX)* and *Sonic Healthcare (SHL)*, in *Case 3* it turns out to be lower than in the first two cases.

In *Case 1*, remembering that in this code there is not synchronization of the returns, it is possible to directly build the portfolio through the Markowitz model. In this first case, the portfolio composition gives an expected return of 32.30% with a volatility of 21.145% and the Sharpe ratio is 1.528.

In *Case 2* and *Case 3*, on the other hand, firstly, the returns are synchronized following the *MA(1)* model described in the previous chapters. It is necessary to estimate the parameter of the model, and since the thesis is based on ten stocks, there are 145 parameters to be estimated, but only the first 45 parameters are the necessary ones to compute the synchronized returns, while the remaining 100 are the noise parameters.

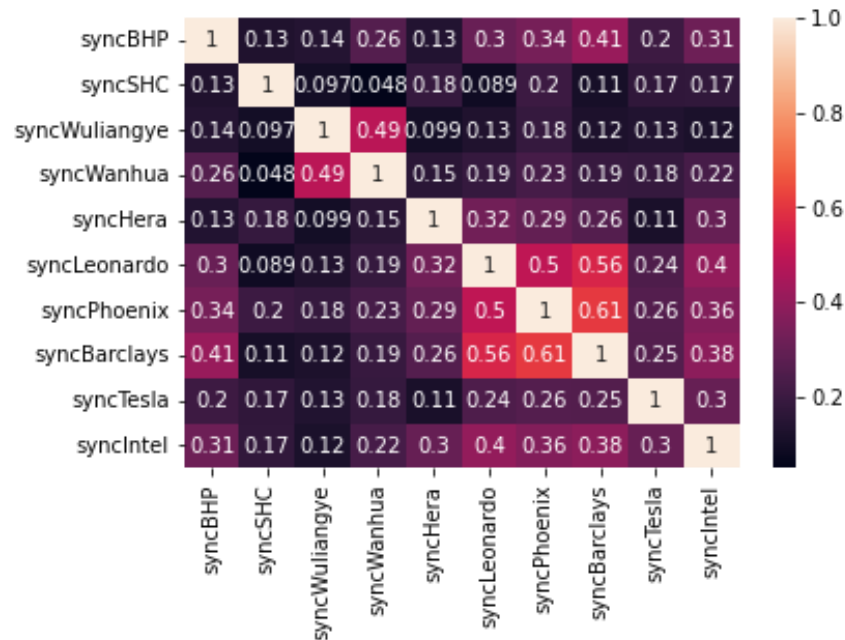
Once the parameters of the model are estimated, for both *Case 2* and *Case 3*, the synchronized returns can be calculated, and subsequently it is created, as before, the correlation matrix, but in this cases using the synchronized returns.

Figure 18: *Case 2* – Correlation Matrix of Synchronized Returns with ‘inner’ parameter



Source: Own production (Python Code)

Figure 19: *Case 3* – Correlation Matrix of Synchronized Returns with ‘outer’ parameter

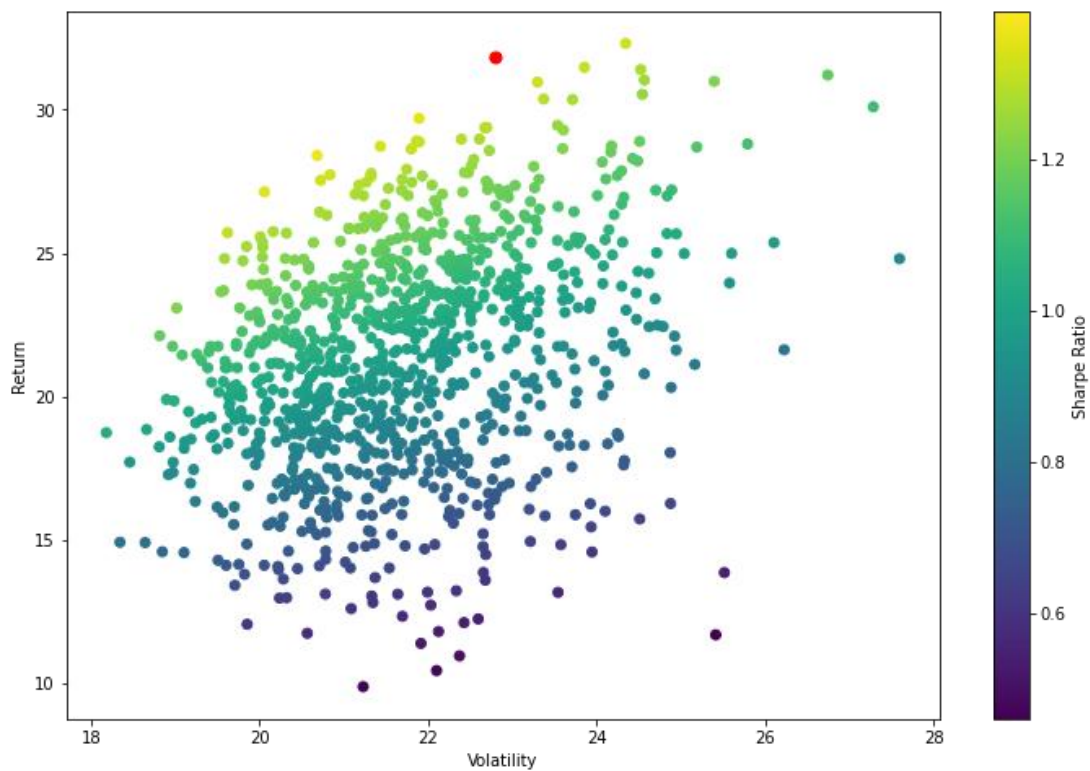


Source: Own production (Python Code)

Comparing the correlation matrices in Figures 16 and 17 with those in Figure 18 and Figure 19, it is worth noting how most of the correlation values increase in the synchronized case, while only in few cases, such as for example the correlation between BHP Billiton (*BHP*) and Sonic Healthcare (*SHL*) (both stocks listed on the Australian stock exchange), decreases.

Afterwards, as in *Case 1*, the best portfolio composition is selected, the one with the highest Sharpe ratio among 1000 possibilities, with expected return of 31.81% and 28.04%, and volatility 22.78% and 21.57%, respectively for *Case 2* and *Case 3*.

Figure 20: *Case 2* – Portfolio with the highest Sharpe ratio



Source: Own production (Python Code)

Figure 20 shows all the possible portfolio compositions and the best one, i.e., the one with the highest Sharpe ratio, can be identified by a red dot. Similar results can be found for *Case 1* and *3*.

Table 2: Weights of the optimal portfolios

	Case 1	Case 2	Case 3
BHP	15.29%	8.80%	8.80%
SHC	4.28%	15.44%	15.44%
Wuliangye	2.51%	11.98%	11.98%
Wanhua	5.77%	11.21%	11.21%
Hera	16.18%	12.06%	12.06%
Leonardo	12.11%	16.72%	16.72%
Phoenix	8.74%	11.10%	11.10%
Barclays	16.69%	1.82%	1.82%
Tesla	0.52%	9.92%	9.92%
Intel	17.90%	0.95%	0.95%

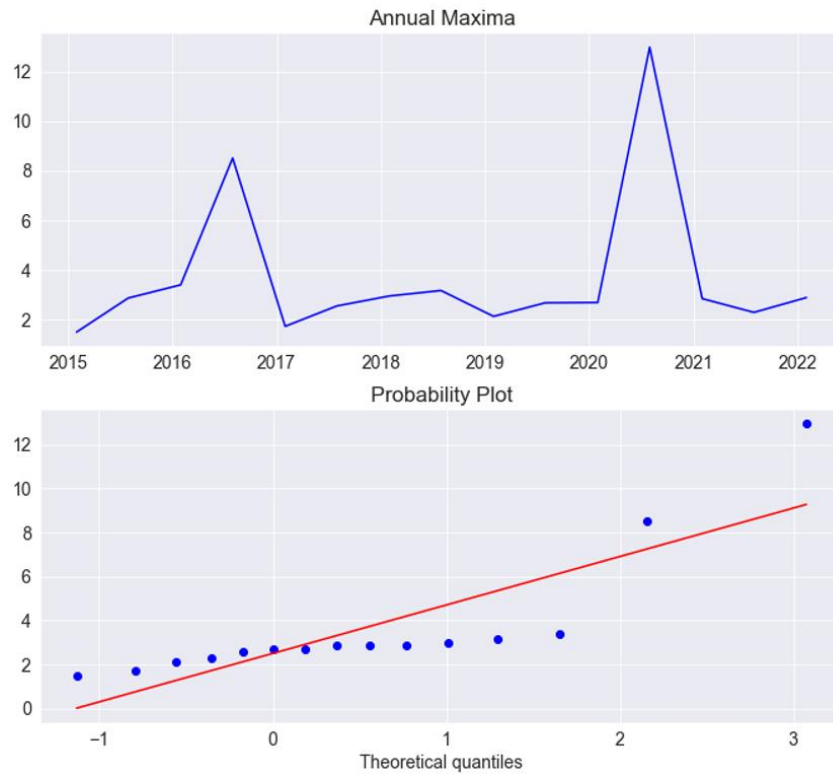
Source: Own production

Table 2 shows the different weights of the stocks that compose the three portfolios calculated using Markowitz. It is worth noting that, despite the different way in which missing values in *Case 2* and *Case 3* are computed, the weights associated to each stocks are the same.

After the selection of the best portfolio composition, it is possible to recover the time series of the portfolio's returns and on that sequence apply the Extreme Value Theory, starting with the Block Maxima model and then moving on to the Peak Over Threshold, and, finally, calculating the risk measures (VaR and ES).

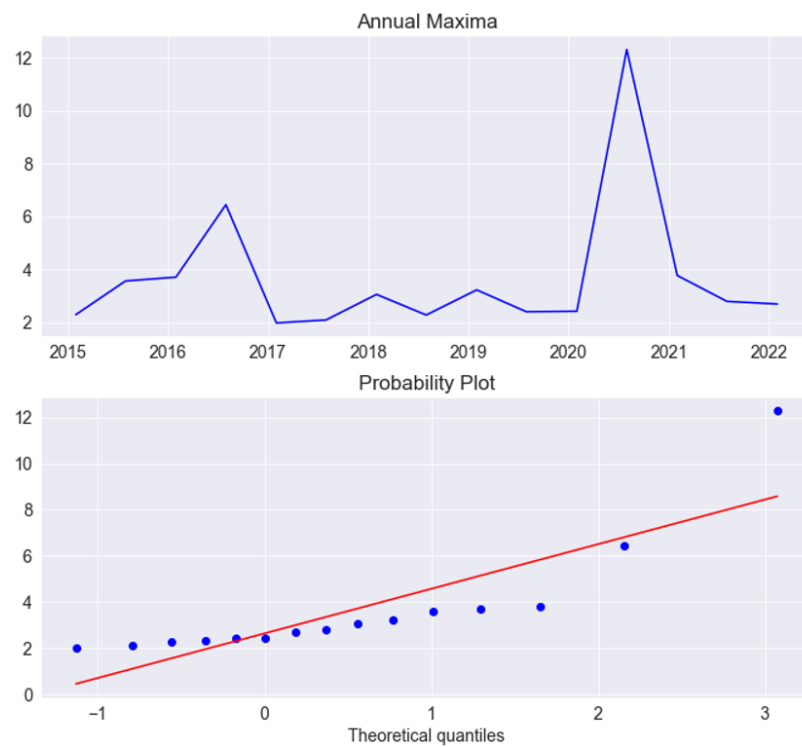
Firstly, for the construction of the Block Maxima model, it is necessary to define the size of the blocks, and, in this thesis, the size has been defined as six months. After the blocks are built, it is possible to measure the maximum for each block, analyze how each maximum behaves.

Figure 21: *Case 1* – Identification of the maximums and comparison with a quantile



Source: Own production (Python Code)

Figure 22: *Case 2* and *Case 3*– Identification of the maximums and comparison with a quantile



Source: Own production (Python Code)

The first chart identifies the different maximums over time and in all the three cases it is possible to observe that the highest maximum is between years 2020 and 2021. The second graph, instead, describes the behavior of the maximums with respect to a quantile of an Exponential distribution. It is used this particular distribution because is the one that behaves most similarly to the GEV distribution. As it can be observed, from the graph in Figure 21 and Figure 22, all dots are in line with the quantile, apart the last located in the top-right corner.

The next step is to estimate the GEV distribution using the Maximum Likelihood Function, as seen in Chapter 3²¹. For this particular distribution, estimates of the shape, location and scale parameters are needed.

Once estimated the parameters and therefore the GEV distribution, it is possible to compare the *Crude Residuals* of the GEV with the quantile of an Exponential distribution. Crude residuals allows to evaluate if the fit of the model is appropriate, and in particular, if the crude residuals closely follow an exponential distribution, it implies the model is correctly specified. Crude residuals are built in a way that their correct distribution is actually exponential, exploiting some properties of the transformation of random variables. In fact, a continuous CDF F , it is known that

$$F(x) = P(X \leq x) = P(F(X) \leq F(x)) = F(x) \quad (4.2)$$

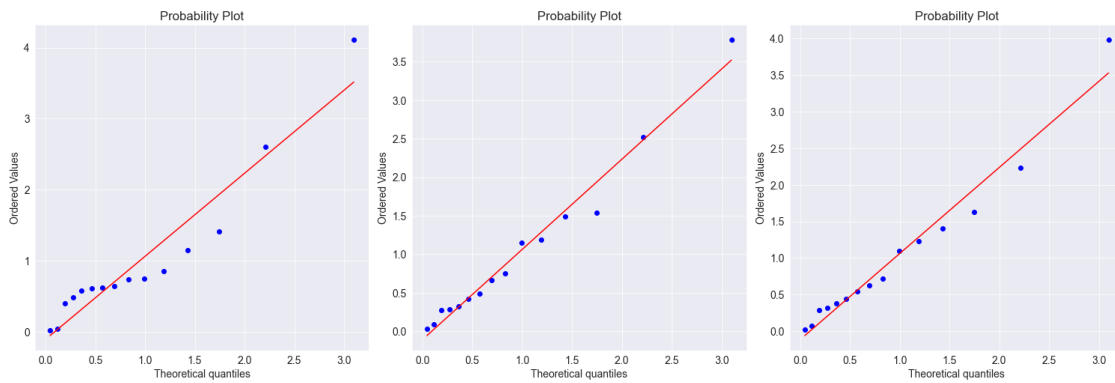
Therefore, the transformation $U = F(X)$ gives a Uniform random variable. Applying the *crude* transformation, namely, $-\log F(X) = -\log(U)$. And noting that if U is Uniform, also $1 - U$ is Uniform, proves that crude residuals are actually $Exp(1)$.

In fact, note that $-\log(1 - U)$ is the inverse of F^{-1} for an $Exp(1)$. So, $-\log(U)$ is $Exp(1)$. Consequently, applying the $-\log(F(X))$ transformation, it is expected to obtain an exponential distribution.

For this reason it is possible to compare the crude residuals with the quantile of an exponential distribution.

²¹ See Chapter 3.1.2. *Estimating the GEV distribution by the Maximum Likelihood*

Figure 23: QQ – Plots of *Case 1*, *Case 2* and *Case 3*



Source: Own production (Python Code)

It can be noticed that it seems reasonable compare the crude residuals and an exponential quantile, since the majority of the maximum lie on the quantile. There are two dots, for each graph, that are far from the other and they represent exactly the abnormal dot in the second graph of Figure 21 and Figure 22.

With the estimates made, it is possible to calculate the risk measures seen before such as the return level²² and the return period²³.

A possible example is to answer to the questions: it is possible to observe an extreme loss larger than u ? This corresponds to compute the probability of $M_n > u$. For instance, what happens if $u = 3\%$?

Based on the assumption and the data, for *Case 2*, it takes almost 2.44 years on average to observe a loss larger than 3%. This question has also been applied to *Case 1* and *3*, and respectively 2.30 years and 2.38 years are the results obtained.

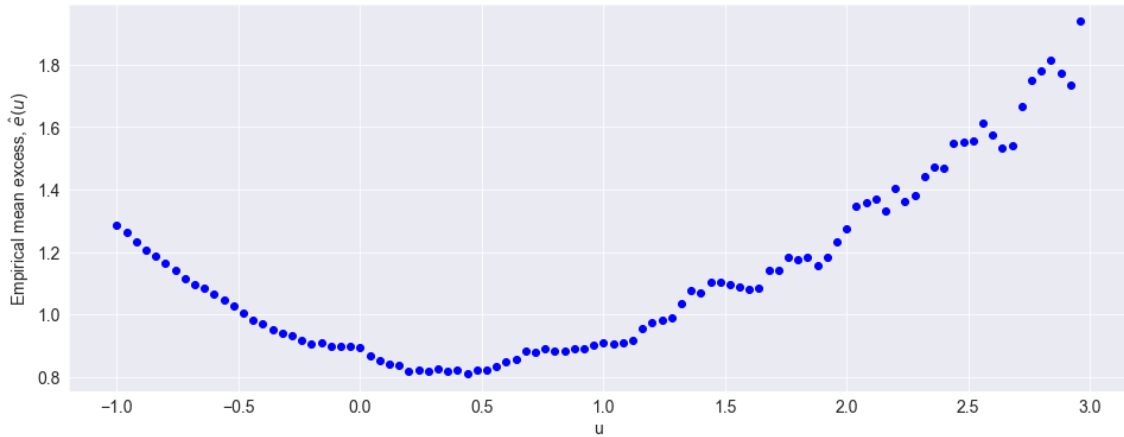
A criticism of the Block Maxima model is the rigidity because assuming that in a year there have been many significant maximum peaks, while in another year there have not been any. Using the Block Maxima model and looking for, for example, an annual maximum, in the first year the greatest peak will be taken excluding the other significant peaks, while in the year in which there are no significant peaks, it may result as a maximum an extremely low value, such as 1%.

²² See Chapter 3.1.3. *Return Level*

²³ See Chapter 3.1.4. *Return Period*

For this reason, the Peaks Over Threshold (*POT*) model is used. This model is closely related to the Block Maxima model, but here the tail of the distribution of a very large sample is directly analyzed. In the POT model, a threshold is defined above which an observation can be interpreted as a maximum. To define the threshold value, it is possible to use a graphical method.

Figure 24: *Case 2* – Definition of the threshold value



Source: Own production (Python Code)

Figure 24 shows the empirical excess mean for the definition of the threshold of *Case 2*, the graph of *Case 1* and *Case 3* behave similarly.

As seen in Chapter 3, when the empirical excess mean becomes linear and positively sloped, it means that the approximation works, and consequently the threshold is identified as the curve begins to increase. In the three cases, the threshold assumes values 0.7, 0.5, and 0.6, respectively for *Case 1*, *Case 2* and *Case 3*.

As for the GEV distribution, the GDP's parameters, for each case, are estimated through Maximum Likelihood using as threshold the value identified before. In the GPD the parameters estimated are $\hat{\xi}$ and $\hat{\beta}(u)$.

Similarly, than for Block Maxima, crude residuals might provide hints on the goodness of the GPD model. In this case, crude residuals are defined as

$$w_i = \frac{1}{\hat{\xi}} \log \left(1 + \frac{\hat{\xi} z_i}{\hat{\beta}} \right) \quad (4.2)$$

where z_i are excess losses.

It is possible to compare the crude residuals with respect to the quantile of an exponential distribution, and also in this case the approximation given by the crude residuals seems to be reasonable in comparison with an exponential distribution.

Once the parameters estimates are known, it is possible to use them to calculate the Value-At-Risk and the Expected Shortfall, using formulas seen in Chapter 3²⁴.

The results of the different VaR and ES for each code are summarized in Table 1.

Table 3: Results of VaR and ES at 99% and 95% level of confidence

	No Synchronization	Synchronization with 'Inner'	Synchronization with 'Outer'
VaR 99%	3.495%	3.521%	3.318%
ES 99%	5.544%	5.128%	4.748%
VaR 95%	1.673%	1.782%	1.711%
ES 95%	2.933%	2.931%	2.765%

Source: Own production

Table 3 provides results about the two main risk measures VaR and ES with 99% and 95% level of confidence. As seen in Chapter 3, generally, the $ES_{\alpha} > VaR_{\alpha}$ and here can be also see that $VaR_{0.99} > VaR_{0.95}$.

Notable as *Case 3*, i.e., the code with synchronization using parameter 'Outer' provides lower values for VaR and ES than the other two cases, that have closed results. In fact, it is possible to see how *Case 1* with respect to *Case 2* has a lower VaR, but a higher ES for both, 99% and 95%, levels of confidence.

Analyzing the results of *Case 3*, it can be stated that with 1% probability the daily return could be as low as -3.318% and, given that the return is less than -3.318% , the average return value is -4.748% . Similarly, with 5% probability the daily return could be as low as -1.711% with an average return of -2.765% , given that the return is less than -1.711% .

²⁴ See Chapter 3.3. *Risk Measures*

Conclusion

The main purpose of this thesis has been that of exploiting the statistical procedures deriving from extreme value theory and the synchronization of the markets, in order to model the behavior of extreme events. In particular, it was decided to study the observations of a portfolio composed of ten stocks, from January 2015 to December 2021 to have sufficient observations.

Firstly, the correlation structure between returns of different stocks is different when the synchronization procedure is applied. As previously seen, in general, the correlation increases, and becomes a higher positive value, i.e., the stocks move in the same direction: as one increases, also the other increases. Only in a few cases, the correlation decreases, while, however remaining, positive.

Then, it has been proved that the composition of the best portfolio in the three cases analyzed changes as the initial assumptions of the code change. In fact, the quantity invested in the single stock varies and this directly affects the time series of the portfolio, which is used to implement the Extreme Value Theory. In fact, looking at the application of the Block Maxima model, the *six-months* blocks created have different maximum based on the time series of the portfolio.

Anyway, the most interesting results derive from the analysis carried out by comparing the values obtained for the Value-at-Risk and the Expected Shortfall. The two risks measures are affected by the synchronization, and this changes the probabilities to observe an extreme event.

To conclude, the extreme value analysis certainly represents the more suitable technique able to predict the occurrence of extreme events. However, as it tries to make inference outside of the range of available data, a critic view is always required when applying such methodology to real life examples.

Bibliography

- Bodie, Z., Kane, A., Marcus, A. J. (2018), *Investments, Eleventh Edition*, McGraw-Hill Education, New York.
- Burns, P., Engle, R., Mezrich, J., (1998), *Correlations and Volatilities of Asynchronous Data*, University of California, Discussion Paper 97-30R.
- Coles, S.G. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer, London.
- Commandeur, J.J.F., Koopman, S.J. (2007), *An introduction to the State Space Time Series Analysis*, Oxford University Press.
- Fisher, R.A., Tippett, L.H.C., (1928), *Limiting Forms of Frequency Distribution of the Largest or Smallest Members of a Sample*, Proceedings of the Cambridge Philosophical Society, 24, 180-190.
- Fulton, C. (2017), *Estimating time series models by state space methods in Python: Statsmodels*, Working Paper.
- Haan, L., Ferreira, A. (2006), *Extreme Value Theory: An Introduction*, Springer, New York.
- Markowitz, H. (1952), *Portfolio Selection*, The Journal of Finance, Vol. 7, No. 1, pp. 77-91.
- McNeil, A.J. (1998), *Calculating Quantile Risk Measures for Financial Returns using Financial Returns using Extreme Value Theory*, ETH E-Collection, Zurich.
- McNeil, A.J. (1999), *Extreme Value Theory for Risk Managers*, ETH Zentrum, Zurich.
- Meinhold, R.J., Singpurwalla, N.D. (1983), *Understanding the Kalman Filter*, The American Statistician, Vol. 37, pp. 123-127.
- Phillips, P.J. (2009), *Applying Modern Portfolio Theory to the Analysis of Terrorism*, Defence & Peace Economics, 20, pp. 193-213.
- Reiss, R.-D., Thomas, M. (2007), *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*, Birkhäuser Basel, Berlin.

Sheppard, K. (2020), *Notes on Financial Econometrics*, Chapter 9, University of Oxford.

Statman M. (1987), *How Many Stocks Make a Diversified Portfolio?*, *Journal of Financial and Quantitative Analysis* 22.

Zivot, A., Wang, J. (2006), *Modeling Financial Time Series with S-PLUS*, Second Edition, Springer.

Annex A Code

The following code is one of the three codes that have been used to produce all the figures and the results presented in the previous chapters.

```
#libraries
import pandas as pd
import numpy as np
import scipy
import statsmodels.api as sm
import matplotlib.pyplot as plt
import seaborn as sn
from functools import reduce
import scipy.stats
import datetime as dt

#uploading the prices of the stocks
INT=pd.DataFrame(pd.read_excel('Intel Corp.xlsx',
index_col=0))
BCS=pd.DataFrame(pd.read_excel('Barclays.xlsx',
index_col=0))
BHP=pd.DataFrame(pd.read_excel('BHP Billiton.xlsx',
index_col=0))
HER=pd.DataFrame(pd.read_excel('Hera SpA.xlsx',
index_col=0))
LDO=pd.DataFrame(pd.read_excel('Leonardo SpA.xlsx',
index_col=0))
PHNX=pd.DataFrame(pd.read_excel('Phoenix Group.xlsx',
index_col=0))
SHC=pd.DataFrame(pd.read_excel('Sonic HealthCare.xlsx',
index_col=0))
TSLA=pd.DataFrame(pd.read_excel('Tesla.xlsx', index_col=0))
WAN=pd.DataFrame(pd.read_excel('Wanhua Chemical.xlsx',
index_col=0))
WUL=pd.DataFrame(pd.read_excel('Wuliangye Yibin.xlsx',
index_col=0))

#database of the prices
prezzi=pd.merge(BHP, SHC, on='Data', how='inner')
prezzi=pd.merge(prezzi, WUL, on='Data', how='inner')
prezzi=pd.merge(prezzi, WAN, on='Data', how='inner')
prezzi=pd.merge(prezzi, HER, on='Data', how='inner')
prezzi=pd.merge(prezzi, LDO, on='Data', how='inner')
prezzi=pd.merge(prezzi, PHNX, on='Data', how='inner')
prezzi=pd.merge(prezzi, BCS, on='Data', how='inner')
prezzi=pd.merge(prezzi, TSLA, on='Data', how='inner')
prezzi=pd.merge(prezzi, INT, on='Data', how='inner')
prezzi

#returns of the stocks
k=len(prezzi.columns)
```

```

returns=(prezzi.iloc[:, :k].pct_change(1)*100)
returns=returns.dropna(axis=0)
returns

#correlation matrix
corr_mat=returns.corr()
sn.heatmap(corr_mat, annot=True)
plt.show()

# Construct the model for an MA(1) with diagonal MA matrix
class MA1d(sm.tsa.statespace.MLEModel):
    def __init__(self, endog):
        # Initialize the state space model
        super(MA1d, self).__init__(endog, k_states=20,
k_posdef=10,

initialization='stationary')
        # Setup the fixed components of the state space
representation
        self['design'] = np.hstack((np.eye(10),
np.zeros((10,10))))
        self['transition'] = np.zeros((20,20))
        self['selection'] = np.vstack((np.eye(10),
np.eye(10)))
# Describe how parameters enter the model
    def update(self, params, transformed=True, **kwargs):
        params = super(MA1d, self).update(params,
transformed, **kwargs)
#self['design', 0, 1] = params[0]
        self['transition', 0, 11] = params[0]
        self['transition', 0, 12] = params[1]
        self['transition', 0, 13] = params[2]
        self['transition', 0, 14] = params[3]
        self['transition', 0, 15] = params[4]
        self['transition', 0, 16] = params[5]
        self['transition', 0, 17] = params[6]
        self['transition', 0, 18] = params[7]
        self['transition', 0, 19] = params[8]
        self['transition', 1, 12] = params[9]
        self['transition', 1, 13] = params[10]
        self['transition', 1, 14] = params[11]
        self['transition', 1, 15] = params[12]
        self['transition', 1, 16] = params[13]
        self['transition', 1, 17] = params[14]
        self['transition', 1, 18] = params[15]
        self['transition', 1, 19] = params[16]
        self['transition', 2, 13] = params[17]
        self['transition', 2, 14] = params[18]
        self['transition', 2, 15] = params[19]
        self['transition', 2, 16] = params[20]
        self['transition', 2, 17] = params[21]

```



```
self['transition', 2, 18] = params[22]
self['transition', 2, 19] = params[23]
self['transition', 3, 14] = params[24]
self['transition', 3, 15] = params[25]
self['transition', 3, 16] = params[26]
self['transition', 3, 17] = params[27]
self['transition', 3, 18] = params[28]
self['transition', 3, 19] = params[29]
self['transition', 4, 15] = params[30]
self['transition', 4, 16] = params[31]
self['transition', 4, 17] = params[32]
self['transition', 4, 18] = params[33]
self['transition', 4, 19] = params[34]
self['transition', 5, 16] = params[35]
self['transition', 5, 17] = params[36]
self['transition', 5, 18] = params[37]
self['transition', 5, 19] = params[38]
self['transition', 6, 17] = params[39]
self['transition', 6, 18] = params[40]
self['transition', 6, 19] = params[41]
self['transition', 7, 18] = params[42]
self['transition', 7, 19] = params[43]
self['transition', 8, 19] = params[44]
```

```
self['state_cov', 0, 0] = params[45]
self['state_cov', 0, 1] = params[46]
self['state_cov', 0, 2] = params[47]
self['state_cov', 0, 3] = params[48]
self['state_cov', 0, 4] = params[49]
self['state_cov', 0, 5] = params[50]
self['state_cov', 0, 6] = params[51]
self['state_cov', 0, 7] = params[52]
self['state_cov', 0, 8] = params[53]
self['state_cov', 0, 9] = params[54]
self['state_cov', 1, 0] = params[55]
self['state_cov', 1, 1] = params[56]
self['state_cov', 1, 2] = params[57]
self['state_cov', 1, 3] = params[58]
self['state_cov', 1, 4] = params[59]
self['state_cov', 1, 5] = params[60]
self['state_cov', 1, 6] = params[61]
self['state_cov', 1, 7] = params[62]
self['state_cov', 1, 8] = params[63]
self['state_cov', 1, 9] = params[64]
self['state_cov', 2, 0] = params[65]
self['state_cov', 2, 1] = params[66]
self['state_cov', 2, 2] = params[67]
self['state_cov', 2, 3] = params[68]
self['state_cov', 2, 4] = params[69]
self['state_cov', 2, 5] = params[70]
self['state_cov', 2, 6] = params[71]
```

```
self['state_cov', 2, 7] = params[72]
self['state_cov', 2, 8] = params[73]
self['state_cov', 2, 9] = params[74]
self['state_cov', 3, 0] = params[75]
self['state_cov', 3, 1] = params[76]
self['state_cov', 3, 2] = params[77]
self['state_cov', 3, 3] = params[78]
self['state_cov', 3, 4] = params[79]
self['state_cov', 3, 5] = params[80]
self['state_cov', 3, 6] = params[81]
self['state_cov', 3, 7] = params[82]
self['state_cov', 3, 8] = params[83]
self['state_cov', 3, 9] = params[84]
self['state_cov', 4, 0] = params[85]
self['state_cov', 4, 1] = params[86]
self['state_cov', 4, 2] = params[87]
self['state_cov', 4, 3] = params[88]
self['state_cov', 4, 4] = params[89]
self['state_cov', 4, 5] = params[90]
self['state_cov', 4, 6] = params[91]
self['state_cov', 4, 7] = params[92]
self['state_cov', 4, 8] = params[93]
self['state_cov', 4, 9] = params[94]
self['state_cov', 5, 0] = params[95]
self['state_cov', 5, 1] = params[96]
self['state_cov', 5, 2] = params[97]
self['state_cov', 5, 3] = params[98]
self['state_cov', 5, 4] = params[99]
self['state_cov', 5, 5] = params[100]
self['state_cov', 5, 6] = params[101]
self['state_cov', 5, 7] = params[102]
self['state_cov', 5, 8] = params[103]
self['state_cov', 5, 9] = params[104]
self['state_cov', 6, 0] = params[105]
self['state_cov', 6, 1] = params[106]
self['state_cov', 6, 2] = params[107]
self['state_cov', 6, 3] = params[108]
self['state_cov', 6, 4] = params[109]
self['state_cov', 6, 5] = params[110]
self['state_cov', 6, 6] = params[111]
self['state_cov', 6, 7] = params[112]
self['state_cov', 6, 8] = params[113]
self['state_cov', 6, 9] = params[114]
self['state_cov', 7, 0] = params[115]
self['state_cov', 7, 1] = params[116]
self['state_cov', 7, 2] = params[117]
self['state_cov', 7, 3] = params[118]
self['state_cov', 7, 4] = params[119]
self['state_cov', 7, 5] = params[120]
self['state_cov', 7, 6] = params[121]
self['state_cov', 7, 7] = params[122]
```



```

#Theta matrix
Theta=np.zeros((10,10))
Theta[0,1]=res.params[0]
Theta[0,2]=res.params[1]
Theta[0,3]=res.params[2]
Theta[0,4]=res.params[3]
Theta[0,5]=res.params[4]
Theta[0,6]=res.params[5]
Theta[0,7]=res.params[6]
Theta[0,8]=res.params[7]
Theta[0,9]=res.params[8]
Theta[1,2]=res.params[9]
Theta[1,3]=res.params[10]
Theta[1,4]=res.params[11]
Theta[1,5]=res.params[12]
Theta[1,6]=res.params[13]
Theta[1,7]=res.params[14]
Theta[1,8]=res.params[15]
Theta[1,9]=res.params[16]
Theta[2,3]=res.params[17]
Theta[2,4]=res.params[18]
Theta[2,5]=res.params[19]
Theta[2,6]=res.params[20]
Theta[2,7]=res.params[21]
Theta[2,8]=res.params[22]
Theta[2,9]=res.params[23]
Theta[3,4]=res.params[24]
Theta[3,5]=res.params[25]
Theta[3,6]=res.params[26]
Theta[3,7]=res.params[27]
Theta[3,8]=res.params[28]
Theta[3,9]=res.params[29]
Theta[4,5]=res.params[30]
Theta[4,6]=res.params[31]
Theta[4,7]=res.params[32]
Theta[4,8]=res.params[33]
Theta[4,9]=res.params[34]
Theta[5,6]=res.params[35]
Theta[5,7]=res.params[36]
Theta[5,8]=res.params[37]
Theta[5,9]=res.params[38]
Theta[6,7]=res.params[39]
Theta[6,8]=res.params[40]
Theta[6,9]=res.params[41]
Theta[7,8]=res.params[42]
Theta[7,9]=res.params[43]
Theta[8,9]=res.params[44]

#synchronized returns
rsync=(np.identity(10)+Theta)@epsilon

```

```

#dataframe with synchronized returns
sync_returns=returns.iloc[:, -10:]
sync_returns

#correlation matrix of synchronized returns
sync_corr=sync_returns.corr()
sn.heatmap(sync_corr, annot=True)
plt.show()

#Markowitz portfolio
np.random.seed(40)
num_ports = 1000
all_weights = np.zeros((num_ports, k))
ret_arr = np.zeros(num_ports)
vol_arr = np.zeros(num_ports)
sharpe_arr = np.zeros(num_ports)

for x in range(num_ports):
    # Weights
    weights = np.array(np.random.random(10))
    weights = weights/np.sum(weights)
    #Save weights
    all_weights[x,:] = weights
    # Expected return
    ret_arr[x] = np.sum((sync_returns.mean() * weights *
252))
    # Expected volatility
    vol_arr[x] = np.sqrt(np.dot(weights.T,
np.dot(sync_returns.cov()*252, weights)))
    # Sharpe Ratio
    sharpe_arr[x] = ret_arr[x]/vol_arr[x]

print('Max Sharpe ratio:{}'.format(sharpe_arr.max()))
print('Location of the Max sharpe ratio
is:{}'.format(sharpe_arr.argmax()))

#return and volatility of the best portfoglio
ret_port=ret_arr[sharpe_arr.argmax()]
vol_port=vol_arr[sharpe_arr.argmax()]

#efficient frontier
plt.figure(figsize=(12,8))
plt.scatter(vol_arr, ret_arr, c=sharpe_arr, cmap='viridis')
plt.colorbar(label='Sharpe Ratio')
plt.xlabel('Volatility')
plt.ylabel('Return')
plt.scatter(vol_port, ret_port, c='red', s=50) # red dot
plt.show()

#time series of the portfolio
Port=pd.DataFrame(sync_returns.dot(weights))

```

```

Port.index = pd.to_datetime(Port.index)
Port.columns=['Returns']
Port['Loss']=-Port['Returns']
Port

#QQ plot vs. Normal quantiles
plt.rc("figure", figsize=(8, 8))
scipy.stats.probplot(Port['Returns'], dist="norm",
plot=plt)
plt.rc("figure", figsize=(16, 6))

#Block Maxima Method with Blocks of 6-months
BMax=Port.resample('6M').agg({'Loss': 'max'})
BMax

#compute the record development plot (see Zivot, p. 150)
cum_max= Port['Loss'].copy()
cum_max=cum_max.to_frame()
cum_max['cmax']= Port['Loss'].cummax()
#if returns =0, then no new max!
cum_max['changes']= cum_max['cmax'].pct_change()
cum_max=cum_max[cum_max['changes'] != 0].dropna()
cum_max

#descriptive statistics for the maxima
f,ax=plt.subplots(2,2, figsize=(25,10), sharex=False)
print(ax)

ax[0,0].plot(BMax['Loss'], color='b' )
ax[0,0].set(title='Annual Maxima' )

ax[0,1].hist(BMax['Loss'], bins=10)
ax[0,1].set(title='Annual Maxima histogram' )

ax[1,0] = scipy.stats.probplot(BMax['Loss'],
dist="gumbel_r", plot=ax[1,0])

ax[1,1].plot(cum_max['cmax'], 'o', color='b' , label='cum
max for losses')
ax[1,1].set(title='Record Developement Plot' )

plt.show()

#Calculate the log-likelihood
def MLE_GEV(x, xi, m, s):

    return np.sum(scipy.stats.genextreme.logpdf(x, xi, m,
s))

#set the parameters and compute the likelihood
from statsmodels.base.model import GenericLikelihoodModel

```

```

class GEVmle(GenericLikelihoodModel):
    def __init__(self, endog, **kws):
        super(GEVmle, self).__init__(endog, **kws)

    def nloglikeobs(self, params):
        xi= params[0]
        mu = params[1]
        sigma = params[2]
        ll = MLE_GEV(self.endog, xi, mu, sigma)
        return -ll

    def fit(self, start_params=None, maxiter=10000,
maxfun=5000, **kws):
        if start_params == None:
            # Reasonable starting values
            start_params = [-0.1, 4.032, 0.297]
        return super(GEVmle,
self).fit(start_params=start_params, maxiter=maxiter,
maxfun=maxfun,**kws)

#fitting the model
mod = GEVmle(BMax['Loss'])
res = mod.fit()

#estimates results
print(res.summary(xname=['shape', 'loc', 'scale']))

#QQ plot of crude residuals
xi=res.params[0]
loc=res.params[1]
scale=res.params[2]

#crude residuals
w=(1 - xi*(BMax['Loss']-loc)/scale)**(1/xi)
#is the model's fit good?

plt.rc("figure", figsize=(8, 8))
scipy.stats.probplot(w, dist="expon", plot=plt)
plt.rc("figure", figsize=(16, 6))

#QQ plot excess losses vs. Exponential, setting as threshold
u=1
EPort=Port[ Port['Loss']>1]
plt.rc("figure", figsize=(8, 8))
scipy.stats.probplot(EPort['Loss'], dist="expon", plot=plt)
plt.rc("figure", figsize=(16, 6))

#Optimal threshold
eu = [Port.Loss[ Port['Loss']>j ].mean() -j for j in
np.arange(-1,3,.04)]

```

```

#Plotting the threshold
plt.plot(np.arange(-1,3,.04), eu, 'bo')
plt.xlabel("u")
plt.ylabel("Empirical mean excess,  $\hat{e}(u)$ ")

plt.show()

#setting location parameter to 0
soglia=0.5
paretopar = scipy.stats.genpareto.fit(Port.Loss[
Port['Loss']>soglia ]-soglia, floc=0)
paretopar

#likelihood function
def MLE_pareto(x, xi, beta):
    return np.sum(scipy.stats.genpareto.logpdf(x, xi,
loc=0, scale=beta))

#set the parameters and compute the likelihood
class paretomle(GenericLikelihoodModel):
    def __init__(self, endog, **kwds):
        super(paretomle, self).__init__(endog, **kwds)

    def nloglikeobs(self, params):
        xi= params[0]
        beta = params[1]
        ll = MLE_pareto(self.endog, xi, beta)
        return -ll

    def fit(self, start_params=None, maxiter=10000,
maxfun=5000, **kwds):
        if start_params == None:
            # Reasonable starting values
            start_params = [0.5, 1]
        return super(paretomle,
self).fit(start_params=start_params, maxiter=maxiter,
maxfun=maxfun, **kwds)

#set the threshold
ts=0.5
modP = paretomle(Port.Loss[ Port['Loss']>ts ]-ts) #as input
exchess losses, namely, (loss-ts)
resP = modP.fit(start_params=[.1, 1])

print(resP.summary(xname=['xi', 'beta(u)']))

#crude residuals
resP.summary(xname=['xi', 'beta'])
xi=resP.params[0]

```



```

beta=resP.params[1]

#QQplot w vs exponential
plt.rc("figure", figsize=(8, 8))
scipy.stats.probplot(z, dist="expon", plot=plt)
plt.rc("figure", figsize=(16, 6))

#computing the value at risk and the expected shortfall
k= len(Port[ Port['Loss']>ts ])
n= len(Port)

def VaR_pareto(alpha, xi, beta, n, k, ts):
    VaR= ts + beta * ( (n*(1-alpha)/k)**(-xi)-1 ) / xi
    ES= (VaR + beta - xi*ts)/(1-xi)

    return VaR, ES

VaR99, ES99= VaR_pareto(.99, xi, beta, n, k, ts)
VaR95, ES95= VaR_pareto(.95, xi, beta, n, k, ts)

print(VaR99, VaR95)
print(ES99, ES95)

```