OPTIMAL PORTFOLIO CHOICE:
AN INTEGRATED EXTREME RISK
MANAGEMENT APPROACH

Daniel Mantilla Garcia
18451

Thesis advisor: Prof. Jean-Christophe Meyfredi
Assessor: Dr. Fabrice Tahar
Thesis Msc. Risk and Asset Management
November 2007

EDHEC Business School does not express approval or disapproval concerning the opinions given in this paper which are the sole responsibility of the author.
Abstract

The portfolio choice problem is perhaps the most important but also among the hardest ones to solve in the so-called Market Finance. Much research has been done since its formal introduction by Harry Markowitz and several alternatives and variants have been proposed to overcome its practical and theoretical problems. The present work proposes an approach that steps away from the initial model of portfolio choice, focusing on extreme risk and aiming to avoid unrealistic assumptions. We integrate recent methods that claim to better represent the “reality” of asset’s returns. We assess the gain of using these new technologies by means of out-of-sample Backtesting exercises and comparison with the classic model.

We integrate models that take into account fat tails, asymmetry and stochastic volatility and employ a general framework to model dependence. These models derive from recent but popular methods in Finance such as Extreme Value Theory, GARCH models and Copulas. Using those tools we put together a model which allows us to optimize a portfolio minimizing extreme risk. We construct a portfolio using five equity industry indexes and find there is a gain on using a more sophisticated approach in terms of risk measurement accuracy. We also find that the allocation differs from the classic model. Although we focused on risk minimization, we also corroborate the strong relation between risk and return on the observed out-of-sample results.
ACKNOWLEDGEMENTS

First, I would like to thank my thesis adviser Prof. Jean-Cristophe Meyfredi for his support and advice throughout the whole process, giving the main directions for this thesis and providing methodology suggestions and specific literature. I am very grateful to Dr. Fabrice Tahar for providing lots of helpful comments at different stages and for assessing my thesis.

In addition, I would like to thank Prof. François-Serge Lhabitant whose comments also played an important role at the definition stage of this thesis and also to Prof. Lionel Martellini and Dr. Devraj Basu for their advice on the research process. Finally I would like to thank Prof. Murad S. Taqqu and Dr. Brendan O. Bradley who facilitate me their extreme dependence software.

Daniel Mantilla

Nice, December 2007
A mis padres, Álvaro y María Clemencia
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>2</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>3</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF FIGURES AND TABLES</td>
<td>7</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>10</td>
</tr>
<tr>
<td>2 Theoretical Background</td>
<td>15</td>
</tr>
<tr>
<td>2.1 Risk</td>
<td>15</td>
</tr>
<tr>
<td>2.1.1 Market Risk</td>
<td>15</td>
</tr>
<tr>
<td>2.2 Portfolio Optimization and Risk Management</td>
<td>16</td>
</tr>
<tr>
<td>2.2.1 Modern Portfolio theory: Markowitz, Sharpe, Tobin and more</td>
<td>16</td>
</tr>
<tr>
<td>2.2.2 Risk Management. Why models?</td>
<td>19</td>
</tr>
<tr>
<td>2.2.3 Do investors care about average or extreme risk?</td>
<td>20</td>
</tr>
<tr>
<td>2.3 Extreme Risk Measures</td>
<td>21</td>
</tr>
<tr>
<td>2.3.1 Value at Risk</td>
<td>22</td>
</tr>
<tr>
<td>2.3.2 Expected Shortfall</td>
<td>22</td>
</tr>
<tr>
<td>2.4 Extreme Value Theory - why is it used?</td>
<td>22</td>
</tr>
<tr>
<td>2.5 Time Series Analysis</td>
<td>23</td>
</tr>
<tr>
<td>2.6 Dependence</td>
<td>25</td>
</tr>
<tr>
<td>2.6.1 Linear Correlation</td>
<td>25</td>
</tr>
<tr>
<td>2.6.2 Copulas</td>
<td>26</td>
</tr>
<tr>
<td>2.6.3 Asymptotic Dependence</td>
<td>27</td>
</tr>
<tr>
<td>3 Methodology and Results</td>
<td>30</td>
</tr>
<tr>
<td>3.1 Data description</td>
<td>30</td>
</tr>
<tr>
<td>3.2 Motivation and formal procedure</td>
<td>32</td>
</tr>
<tr>
<td>3.2.1 EVT and Time Series Analysis</td>
<td>33</td>
</tr>
<tr>
<td>3.2.2 EVT and Copulas</td>
<td>33</td>
</tr>
<tr>
<td>3.2.2.1 The t-student Copula</td>
<td>34</td>
</tr>
<tr>
<td>3.2.3 EVT, Extreme Dependence and Asset Allocation</td>
<td>35</td>
</tr>
<tr>
<td>3.2.4 EVT, Copulas and Portfolio Optimization</td>
<td>37</td>
</tr>
<tr>
<td>3.3 Time Series modeling</td>
<td>40</td>
</tr>
<tr>
<td>3.3.1 ARMA and GARCH models</td>
<td>42</td>
</tr>
<tr>
<td>3.3.2 Model selection</td>
<td>44</td>
</tr>
<tr>
<td>3.3.3 Time Series Results</td>
<td>50</td>
</tr>
<tr>
<td>3.4 Extreme Value Theory</td>
<td>55</td>
</tr>
<tr>
<td>3.4.1 The Peaks over a Threshold method</td>
<td>58</td>
</tr>
<tr>
<td>3.4.1.1 Theoretical Background</td>
<td>58</td>
</tr>
<tr>
<td>3.4.1.2 The Threshold choice and Parameter estimation</td>
<td>61</td>
</tr>
<tr>
<td>3.4.2 EVT VaR and ES</td>
<td>64</td>
</tr>
<tr>
<td>3.4.3 Backtesting of VaR</td>
<td>66</td>
</tr>
<tr>
<td>3.4.4 Fitting the semi-parametric distribution</td>
<td>69</td>
</tr>
<tr>
<td>3.5 Dependency</td>
<td>72</td>
</tr>
<tr>
<td>3.5.1 Asymptotic Dependence</td>
<td>73</td>
</tr>
<tr>
<td>3.5.1.1 Asymptotic Dependence Estimation</td>
<td>75</td>
</tr>
<tr>
<td>3.5.2 Dependence with Copulas</td>
<td>79</td>
</tr>
<tr>
<td>3.5.2.1 T-student Copula</td>
<td>81</td>
</tr>
<tr>
<td>3.5.3 Parameter Estimation</td>
<td>84</td>
</tr>
<tr>
<td>3.5.4 Monte Carlo Simulation</td>
<td>86</td>
</tr>
<tr>
<td>3.6 Portfolio Optimization</td>
<td>88</td>
</tr>
</tbody>
</table>
LIST OF FIGURES AND TABLES

Figure 1: Indices created using the five industry portfolios’ returns ........................................... 31
Figure 2: Daily returns of the HiTec index .................................................................................... 32
Figure 3: Sample ACF of the ‘Other’ industry index returns .......................................................... 46
Figure 4: Sample PACF of the ‘Cnsmr’ industry index returns ...................................................... 47
Figure 5: Sample ACF of the ‘Cnsmr’ industry index squared returns ............................................ 47
Figure 6: Sample ACF of the ‘HiTec’ industry index squared returns ........................................... 48
Figure 7: Residuals and GARCH estimated Volatility for the ‘Manuf’ industry index ............... 51
Figure 8: Residuals and GARCH estimated Volatility for the ‘HiTec’ industry index ............... 52
Figure 9: Standardized residuals for the ‘Cnsmr’ industry index ................................................... 52
Figure 10: Standardized residuals for the ‘Hlth’ industry index ..................................................... 53
Figure 11: Sample ACF for the residuals of the ‘Manuf’ index ...................................................... 53
Figure 12: Sample ACF for the squared residuals of the ‘HiTec’ index ........................................ 54
Figure 13: QQ plot of Normal distribution vs. the residuals of the ‘Cnsmr’ index ...................... 56
Figure 14: QQ plot of Normal distribution vs. the residuals of the ‘Other’ index ..................... 57
Figure 15: Mean excess function for the losses of the ‘Manuf’ index ......................................... 62
Figure 16: Mean excess function for the losses of the ‘Hlth’ index ............................................. 62
Figure 17: Mean excess function for the losses of the ‘Hlth’ index ............................................. 63
Figure 18: VaR estimation for the conditional normal, conditional EVT and unconditional EVT methods for the ‘Cnsmr’ index .......................................................... 68
Figure 19: VaR estimation for the conditional normal and conditional EVT with negative returns of the ‘Cnsmr’ index .......................................................... 68
Figure 20: Left tail fit of using EVT, Normal and t distributions. ‘Other’ index ....................... 70
Figure 21: Right tail fit of using EVT, Normal and t distributions. ‘Other’ index ....................... 70
Figure 22: Semi-parametric cdf fit for the Cnsmr index .............................................................. 72
Figure 23: Scatter plot of two pairs: HiTec vs. Cnsmr and Manuf vs. Cnsmr ......................... 77
Figure 24: Scatter plot of two pairs: US vs. Japan and France vs. Germany ......................... 78
Figure 25: Empirical Copula for Hlth and Hitec indices ............................................................. 86
Figure 26: Empirical Copula for Cnsmr and Other indices ........................................................ 86
Figure 27: Dependence of simulated residuals corresponding to the Cnsmr ,Manuf and HiTec indices .......................................................... 87
Figure 28: Efficient Frontier: ES vs. Expected Returns of min-ES portfolios .......................... 93
Figure 29: VaR vs. Expected Returns of min-ES portfolios ..................................................... 94
Figure 30: Efficient Frontier: Volatility vs. Expected Returns of min-Vol portfolios .......... 94
Figure 31: ACF of returns, ‘Cnsmr’ index .................................................................................. 105
Figure 32: ACF of returns, ‘Manuf’ index ................................................................................. 105
Figure 33: ACF of returns, ‘HiTec’ index .................................................................................. 106
Figure 34: ACF of returns, ‘Hlth’ index ................................................................................... 106
Figure 35: PACF of returns, ‘Manuf’ index .............................................................................. 106
Figure 36: PACF of returns, ‘HiTec’ index .............................................................................. 107
Figure 37: PACF of returns, ‘Hlth’ index ................................................................................ 107
Figure 38: PACF of returns, ‘Other’ index ................................................................................ 107
Figure 39: ACF of squared returns, ‘Manuf’ index ................................................................. 108
Figure 40: ACF of squared returns, ‘Hlth’ index ................................................................... 108
Figure 41: ACF of squared returns, ‘Other’ index ................................................................. 108
Figure 42: residuals and standard deviation GARCH forecasts, ‘Cnsmr’ index .................... 109
Figure 43: residuals and standard deviation GARCH forecasts, ‘Hlth’ index ..................... 109
Figure 44: residuals and standard deviation GARCH forecasts, ‘Other’ index .................... 110
Figure 45: ACF of Standardized Residuals, ‘Cnsmr’ index ................................................... 110
Figure 46: ACF of Squared Standardized Residuals, ‘Cnsmr’ index ..................................... 111
Figure 47: ACF of Squared Standardized Residuals, ‘Manuf’ index ..................................... 111
Figure 48: ACF of Standardized Residuals, ‘HiTec’ index ................................................... 111
Figure 49: ACF of Standardized Residuals, ‘Hlth’ index ..................................................... 112
Figure 50: ACF of Squared Standardized Residuals, ‘Hlth’ index ......................................... 112
Figure 51: ACF of Standardized Residuals, ‘Other’ index .................................................. 112
Figure 52: ACF of Squared Standardized Residuals, ‘Other’ index ...................................... 113
Figure 53: Standardized Residuals, ‘Manuf’ index ............................................................... 113
Figure 54: Standardized Residuals, ‘HiTec’ index ............................................................... 113
Figure 55: Standardized Residuals, ‘Other’ index ............................................................... 114
Figure 56: QQ plot of Standardized Residuals vs. the normal dist., ‘Manuf’ index .............. 114
Figure 57: QQ plot of Standardized Residuals vs. the normal dist., ‘HiTec’ index ............... 114
Figure 58: QQ plot of Standardized Residuals vs. the normal dist., ‘Hlth’ index ................. 115
Figure 59: Mean Excess function of Negative Standardized Residuals, ‘Cnsmr’ index ....... 115
Figure 60: Mean Excess function of Negative Standardized Residuals, ‘HiTec’ index ..... 115
Figure 61: Mean Excess function of Negative Standardized Residuals, ‘Other’ index ....... 116
Figure 62: Mean Excess function of Positive Standardized Residuals, ‘Cnsmr’ index ...... 116
Figure 63: Mean Excess function of Positive Standardized Residuals, ‘Manuf’ index ...... 116
Figure 64: Mean Excess function of Positive Standardized Residuals, ‘HiTec’ index ...... 117
Figure 65: Mean Excess function of Positive Standardized Residuals, ‘Hlth’ index ...... 117
Figure 66: Pareto, Normal, t and empirical lower tail, ‘Cnsmr’ index ................................... 118
Figure 67: Pareto, Normal, t and empirical upper tail, ‘Cnsmr’ index ................................... 118
Figure 68: Pareto, Normal, t and empirical lower tail, ‘Manuf’ index .................................... 119
Figure 69: Pareto, Normal, t and empirical upper tail, ‘Manuf’ index .................................... 119
Figure 70: Pareto, Normal, t and empirical lower tail, ‘HiTec’ index .................................... 120
Figure 71: Pareto, Normal, t and empirical upper tail, ‘Hlth’ index ..................................... 120
Figure 72: Pareto, Normal, t and empirical lower tail, ‘Hlth’ index ..................................... 121
Figure 73: Pareto, Normal, t and empirical upper tail, ‘Hlth’ index ..................................... 121
Figure 74: VaR estimation for the conditional normal, conditional EVT and unconditional EVT methods for the ‘Manuf’ index ............................................................... 122
Figure 75: VaR estimation for the conditional normal and conditional EVT methods for the ‘HiTec’ index ............................................................... 123
Figure 76: VaR estimation for the conditional normal, conditional EVT and unconditional EVT methods for the ‘Hlth’ index ............................................................... 124
Figure 77: VaR estimation for the conditional normal and conditional EVT methods for the ‘Other’ index ............................................................... 125
Figure 78: VaR estimation for the conditional normal, conditional EVT and unconditional EVT methods for the ‘Cnsmr’ index ............................................................... 125
Figure 79: VaR estimation for the conditional normal and conditional EVT methods for the ‘Manuf’ index ............................................................... 126
Figure 80: VaR estimation for the conditional normal, conditional EVT and unconditional EVT methods for the ‘HiTec’ index ............................................................... 127
Figure 81: VaR estimation for the conditional normal and conditional EVT methods for the ‘Hlth’ index ............................................................... 128
Figure 82: VaR estimation for the conditional normal, conditional EVT and unconditional EVT methods for the ‘Other’ index ............................................................... 129
Figure 83: Return Series ‘Cnsmr’ index ............................................................................... 126
Figure 84: Return Series ‘HiTec’ index ................................................................................. 126
Figure 85: Return Series ‘Hlth’ index ................................................................................... 127
Figure 86: Return Series ‘Other’ index .................................................................................. 127
Table 1: Descriptive Statistics of the five industry portfolios ................................. 32
Table 2: BIC criterion for AR(1), MA(1) and ARMA(1,1) ............................................. 49
Table 3: BIC criterion for AR(1), MA(1) and ARMA(1,1) with GARCH (1,1) ............... 49
Table 4: T-stats of the estimated parameters of AR(1), MA(1) and ARMA(1,1) ............ 50
Table 5: T-stats of the estimated parameters of AR(1), MA(1) and ARMA(1,1) with GARCH (1,1) .................................................................................................................. 50
Table 6: Estimated parameters of AR(1), MA(1) and ARMA(1,1) with GARCH (1,1) .... 50
Table 7: Ljung-Box test on each of the standardized residuals ..................................... 54
Table 8: Jarque-Bera test for the returns and residuals .................................................. 58
Table 9: Threshold of left and right tails using the 95% and 5% criterion ....................... 63
Table 10: GPD parameters Maximum Likelihood estimates ......................................... 64
Table 11: Expected and actual number of failures of the 3 VaR estimates ........................ 69
Table 12: Extreme dependence measures and asymptotic hypothesis test. ..................... 77
Table 13: Chi-bar for filtered returns ............................................................................. 79
Table 14: Copula Maximum Likelihood parameters estimates .................................... 85


1 Introduction

- *Everything should be made as simple as possible, but not simpler.*
  
  Albert Einstein

The main “law” of Finance is the relation between risk and reward; this principle is the base and the core of the so-called Market Finance. The study of investments choice, which derives to asset allocation and portfolio construction, is perhaps one of the most important topics in this area. A portfolio can be interpreted as an investor’s choice to balance the risk and rewards that she/he is taking. However, this decision is not an easy one; the investor faces virtually innumerable investable choices and more importantly an infinite number of combinations of them. His choice might involve beliefs, objectives, preferences, expectations, risk aversion, time and budget constrains, estimations among other. In addition, the choice will probably also be affected by external factors and changes that are taking place (e.g. in the economy) during the life of its investments. Hence, the investor also faces a dynamic decision problem that may change in time.

In this thesis we focus on the portfolio construction process, which starts with the assessment of the risks and rewards of the set of possible investment opportunities and their relations and ends with a way to balance the assets in the portfolio. A common practice in economics is the ‘need’ to diminish the complexity of the problem and to make some simplifications and assumptions thereof. However, we approach the problem in a way where we try to minimize both the distance with “reality” and the amount of assumptions made. Although, these two purposes would not seem to have a tension in between, when it comes to the implementation
and analysis of the problem, we find that there is actually a trade-off between the amount of “reality” we want to have and the complexity of the models used.

This trade-off between reality and model complexity is derived from the fact that a model with more realistic assumptions about the behavior of the variables that analyses tends to come with a higher number of assumptions as well. Furthermore, more complicated models that aim to capture a higher portion of the reality’s complexity often bear a risk related to the estimation of a presumably higher number of parameters that compose it. On the other hand, a more ‘parsimonious’ model tends to have a lower number of parameters to estimate and should be easier to understand and manipulate.

During the definition of the specific objectives of this thesis, the author in his seek for “reality” looked for models able to emulate in a closer way the features observed in “real” assets with less unrealistic assumptions concerning their behavior. On the other hand, the ‘parsimony’ principle was taken into account when choosing the models. The choice of each model used in this thesis was done based on this trade-off principle and hence was sufficiently motivated by the sake of completeness, coherence and “reality”.

The “investment science”, although in formal terms is relatively new, has experienced vertiginous developments in the recent years, motivated by the needs of more sophisticated investors and the increasing number of scientists that have been looking at the problem since its formal inception by Harry Markowitz in 1952. The amount of new models and methods to estimate the risk and performance of assets, and the different ways to construct a portfolio with those assets is increasing rapidly. This is also provoked by the “reality check” that investors have experienced after taking Markowitz first theoretical model into practice; clear
evidence exists that the benefits promised by portfolio optimization in its conception are still not fully achieved, and hence the demand for improvements is strong.

The aim of this thesis is to investigate the problem of portfolio construction, by looking at an alternative to what classic portfolio choice proposes. Although in principle we also look at the balance between risk and rewards, there are two main differences with the classic approach: (1) the definition and (2) the estimation of risk and hence the optimal balance is achieved in a different way. We decided to stick to a different definition of risk for two reasons: first we believe that a risk metric should focus only on adverse variations of the value of assets which is not the case of the classic risk metric (standard deviation) and second, we motivate the use of ‘Extreme risk’ metrics given their importance to Banks for regulatory purposes; which is the case of the Value-at-Risk: used by the Capital Adequacy Directive of the Bank of International Settlement (BIS) in Basle.

The main reason for the second difference with the classic model is because in this thesis we aim to employ some of the “new technologies” available nowadays to assess those risks. All of these new technologies or models that we use have become very popular among Finance practitioners because they have been designed (or are able) to overcome the unrealistic assumptions made by the earlier models. The promoters of the new models argue that the assumptions made by classic portfolio theory are “unrealistic” because they don’t reflect the empirical observations of historical real data and that the new models take into account what they call ‘stylized facts of financial time series’.

In this sense in this thesis we try to step away from the classic assumption that the returns on assets (in particular equities) follow a multivariate normal distribution. This assumption was
very convenient in developing the first portfolio optimization model, however this hypothesis has been rejected on its empirical tests. The so-called ‘stylized facts’ of financial returns refer indeed to the departures from normality presented by real returns data, and are commonly listed as: fat(ter) tails (with respect to the normal distribution), asymmetry, autocorrelation and stochastic volatility and more recently extreme and non-linear (non normal) dependency between the different assets’ returns. The presence of these stylized facts, when the risk management is based on the classic ‘normal’ assumptions, might translate into a dangerous miss-estimation of risks which may result in financial failures.

Therefore, our motivation to departure from the classic model and more importantly from its assumptions is to avoid the risk that we take when we decide not to step away from it. This risk is what is known as ‘model risk’ which comes in when particular characteristics of the factors analysed by the model are not properly captured. In order to avoid the model risk as much as possible, we try to choose models with the ability to incorporate the stylized facts previously mentioned, while still keeping in mind the ‘parsimony’ principle (i.e. keep it as simple as possible).

The models used are the following: ‘Extreme Value Theory’, used to capture the fat tails of return distributions (and asymmetry as we will see), which are very suitable to measure extreme risk; Time Series Analysis (ARMA and GARCH in particular) to account for serial correlation and stochastic volatility, Copulas and some Extreme Dependency measures to model dependency among assets within a more robust and general framework (and also coherent with our departure from normality). We also use a ‘Kernel smoothing’ method to model the ‘body’ of the marginal distributions, mostly to be consistent with the sake to avoid normality assumptions. All the implementation of these models is perfectly integrated in a coherent way, complementing each other and sufficiently motivated. These set of methods
allow us to have an integrated model useful to perform simulations, risk estimations and finally portfolio construction. The purpose is to base the procedure on a suitable model that aim to replicate in a more “realistic way” the risks that the investor may take. Our approach is a ‘risk management’ one because we believe the future performance estimation has not been satisfactorily documented yet in the academic literature and in any case the performance is not left out of the way given it is closely tied to risk.

We decided to narrow the problem by performing the portfolio construction exercise with US equities using a factor approach. However our factors are in a sense ‘model free’, since they are simply industry indices (although, there is a weak assumption of completeness of the indices when using them as factors). The choice to limit the investment universe to the equity asset class was done for model simplicity purposes; remark the set of suitable models for a given asset class might not be the same for another asset class given it may have different stylized features. Likewise, we choose the US for data availability and completeness reasons.

The rest of the thesis is organized as follows: The second part gives a theoretical background of the different concepts and methods used throughout the thesis and also provide some motivations for their use (in particular simulation methods) in a Risk Management context; the third part briefly present the main results and conclusions of some of the research literature that motivated the work done in this thesis. The set of academic papers summarized in that section exhibit reasons to use the models implemented, the relation among them and some of their most common applications; the fourth part present in a higher level of detail the models and tools used and the results from the empirical work; The fifth section concludes and give some suggestions for further research followed by references and an appendix.
2 Theoretical Background

The purpose of this section is to briefly introduce the main theoretical concepts used throughout the thesis, in order to establish a conceptual ground from where we build the subsequent work. It also aims to give some motivation for the use of the models applied and expose the relations between them.

2.1 Risk

Several definition have been given to this term, among the earliest is the one given by Knight(1921) who defined risk as measurable uncertainty. However, lots of different definitions came along after, some of them who criticized it. Holton (2004) makes an attempt to summarize the most relevant definitions and give a more general concept of risk: the exposure to a proposition of which one is uncertain.

Other definitions address the concept in a closer to a Finance relevant context such as Jorion (2000) who defines it as the volatility of the expected results on the value of assets and liabilities of interest. Depending on the type of the factors that produce it, for firms, the risk he groups it in four categories: Market Risk, Credit Risk, Strategic or Business Risk and Operational Risk. Another definition of risk in a portfolio management context that we consider more appropriate is the one given by Cool (1999) who defines risk as the absolute value of probable loss.

2.1.1 Market Risk

According to Kaplanski and Kroll (2001) Market Risk is defined as a decrease in the value of a position due to changes in the financial market prices. As we can see from these first definitions, there is not a complete agreement or clarity on what is considered as risk for an
investor. Jorion’s (2000) definition of risk makes reference to the term “volatility” which is very often used as synonymous of the standard deviation of the returns\(^1\) of a given asset. The standard deviation is a measure of the average departure of the returns from its own historical mean. On the other hand, Cool (1999) definition describes the risk as the probable decrease in value of (presumably) a portfolio exposed to market prices. The latter definition is closer to what is known as *extreme risk* while the former is a measure of what we will refer as *average risk*. The definition of risk that the portfolio (and/or risk) manager will stick to, may have important implications on the practices used by the manager and consequently on the outcomes of the portfolio as we will illustrate in this thesis.

2.2 Portfolio Optimization and Risk Management

In this section we will briefly refer to some of the most outstanding pieces in the literature concerning portfolio optimization and its relation with (Financial) Risk Management. We will also present some more recent results and findings on the outcomes of portfolio optimization actual implementation.

2.2.1 Modern Portfolio theory: Markowitz, Sharpe, Tobin and more.

*Modern portfolio theory* (MPT)—or *portfolio theory*—was introduced by Harry Markowitz with his paper "Portfolio Selection," which appeared in the 1952 Journal of Finance. Thirty-eight years later, he shared a Nobel Prize with Merton Miller and William Sharpe for what has become a broad theory for portfolio selection.\(^2\)

---

\(^1\) The return of an asset is usually calculated as the percentage change of its value plus adjustments if any.

Prior to Markowitz's work, the assessment of the risks and rewards of portfolios was carried out through the analysis of individual securities independently. By formalizing the concept of diversification, he proposed that investors should focus on selecting portfolios based on their joint risk-reward features instead of merely compiling individually attractive securities regardless of their relation to the other assets on their portfolios.

Using the historical returns of each asset on a portfolio and statistical measures such as average (return), standard deviation and linear correlation is possible to estimate the expected return and volatility of any portfolio constructed with those assets. Markowitz used volatility and expected return as proxies for risk and reward. Within the infinite possible alternatives that an investor has to construct a portfolio, Markowitz defined an “optimal” way of doing so by balancing the risk and reward features of the portfolio. The set of portfolios constructed in this optimal manner conform what he called the efficient frontier. He concludes that an investor should select a portfolio that lies on the efficient frontier.

James Tobin (1958) expanded on Markowitz's work by adding a risk-free asset to the analysis. He pointed out that by using leverage or deleverage on the portfolios on the efficient frontier it was possible to outperform them in terms of their risk and reward relation. By doing so he introduced the notions of “Capital Market Line” and “super-efficient portfolio”.

Sharpe (1964) formalized the Capital Asset Pricing Model (CAPM). Using strong assumptions over investors and market behaviour he created a model that lead to interesting conclusions. He finds that the “market portfolio” sits on the efficient frontier, and is also actually Tobin's super-efficient portfolio. According to CAPM, all investors should hold the market portfolio, leveraged or de-leveraged with positions in the risk-free asset according to their risk aversion profile. CAPM also introduced the concept of “beta” and relates an asset's expected return to its beta.
Portfolio theory provides a broad framework to understand the interactions of risk and reward. It has profoundly influenced the way institutional portfolios are managed, and motivated what is known as “passive management” investment techniques. The mathematics of portfolio theory is widely used in Financial Risk Management and is a theoretical predecessor for more recent risk measures.

Although the important insights that these models have given to portfolio managers, unfortunately when it comes to practice, every assumption explicit on the model or used in the implementation of it will manifest and have deep implications on the actual risk and reward that the portfolio’s holders will bear/obtain.

Even disregarding the important debate on whether the theoretical assumptions that the fathers of MPT hold in reality or not, recent research has shown that other practical problems can kill the gains promised by optimal portfolio choice. DeMiguel, Garlappi and Uppal (2007) showed that the error explicit on the most common model’s parameters estimators can offset the gains from optimal diversification. They show this by demonstrating how a naively diversified portfolio with equal weights in every asset, can out-performs out-of-sample on a risk adjusted basis (Sharpe-ratio in this case) an “optimally” diversified portfolio.

However, several studies have given alternatives to improve some parameters of Markowitz’s mean-variance model. An example of this is the paper by the same Sharpe in 1963 where he uses the market portfolio in order to improve the estimations of the expected return and covariance matrix. Other examples are in Elton and Gruber (1973) and Ledoit and Wolf (2004) where they find that “imposing” a structure to the covariance matrix as opposed to its sample estimator, can improve the output of optimally constructed portfolios.

---

3 The covariance matrix is nothing but the combination of the standard deviations and the correlations of all assets in a portfolio.
Even though this recent research gives some hope for the mean-variance approach, the fact that there is not a well-known robust estimate for the expected return of most assets, evidence the situation that the benefits promised by portfolio optimization in its conception are still not fully achieved, and hence further research has to be done in this sense.

2.2.2 Risk Management. Why models?

Risk management is the process of analyzing the different risk factors that a portfolio is exposed to in order to determine how to best handle such exposures. The development of the markets and science has contributed significantly to the current risk management practices. In particular the use of derivatives instruments, the development of quantitative models applied to finance and computer science have been the factors with major influence.

A lot of the research related to the Risk management practices has focus in modeling the uncertainty of the value of assets on a portfolio (among other economic variables) and the relations between them, often heavily relying in probability theory and statistics. Through extensive exploit of computer programs, these models are frequently used to simulate possible future scenarios. Although the use of a model constitutes another risk by itself, it may enable portfolio and risk managers to explicitly take into consideration some of the uncertainty they face and to quantify and estimate as accurately as possible the risks they take (or consider to take). The use of such models may play an important roll on the decision process of a risk manager, however is important to understand that they constitute solely another tool of analysis.
In this thesis we integrate different probabilistic and statistical models that allow us to estimate and emulate the risk features of a given portfolio and to use simulation techniques that enable us to perform portfolio optimization taking into account explicitly the uncertainty in the value of the assets we invest in.

2.2.3 Do investors care about average or extreme risk?

As we mentioned in section 2.2.1, MPT uses the standard deviation of asset returns as the measure of risk, and focuses on minimizing it when constructing a portfolio. The standard deviation simply measures the average distance of the returns from its historical mean. This can be interpreted as a measure of how “volatile” the price of an asset is.

However, we should still think if this measure of risk is the closest to investors concerns and ask the question: does “volatility” capture what an investor perceives as risk? We may argue that volatility is a measure of uncertainty of the value of an asset rather than it’s risk from an investor perspective. For instance, if we use volatility as the risk-metric and we want to analyze a share that suddenly starts having very high (positive) returns, perhaps generated by the company’s new investment opportunities, the conclusion would be that this share is a more risky asset, due to the bigger distances between the new returns and their historical average. However an investor might not see this volatility as a bad signal or risk but actually as a good opportunity to invest.

In this sense, instead of looking at both, positive and negative variation, we may want to consider risk-metrics focused only in adverse variations on the value of a portfolio. Hence we shall think about measures like the Negative Semi-deviation (similar to standard deviation but
only considers observations below the mean or target return) or measures concerning the possible worse losses of a portfolio during a defined period of time. The latter is related to what we call extreme risk measures. One may argue that this kind of measures can reflect closer an investor’s perception of risk. During the rest of the thesis we will focus on attempts to assess extreme risk and its management.

2.3 Extreme Risk Measures

Extreme risk is usually related to the (joint) losses in the value of an asset (or portfolio of assets) bore during “extreme” situations. Extreme situations can be tough as the worst case scenarios that one could expect with respect to the value of the assets held. Various measures of extreme risk are around, but the most renown is the Value-at-Risk (VaR). This measure is used by the Capital Adequacy Directive of the Bank of International Settlement (BIS) in Basle (Basle Committee 1996), that determined that banks must have a capital cushion sufficient to cover losses on the bank's trading portfolio over a ten-day holding period in 99% of occasions.

However, Artzner, Delbaen, Eber and Heath. (1997, 1999) pointed out some drawbacks of the VaR as a market risk measure. First they show that VaR is not necessarily ‘subadditive’ and explain why this may cause serious problems if the risk-management system of a financial institution is based in VaR-limits for individual books. Furthermore, VaR gives only an upper bound on the losses that occur with a given frequency; VaR tells us nothing about the potential size of the loss given that a loss exceeding this upper bound has occurred. Artzner et al. (1997, 1999) propose the use of the so-called Expected Shortfall instead of VaR, which addresses the issue of the potential size of a loss in extreme scenarios.
2.3.1 Value at Risk

According to Jorion (2000), the Value at Risk (VaR) estimates the maximum loss (or worst loss) that a portfolio can have within a determined time horizon and a given confidence level. For a horizon of N days and a confidence level α %, the VaR is the loss corresponding to the (100- α) quantile in the distribution of the variations of a portfolio’s value during the next N days.

In this thesis we estimate the VaR of a portfolio using Extreme Value Theory and compare it to estimates based on Gaussian assumptions. However, we perform portfolio optimization using as our risk measure the Expected Shortfall, as suggested by Artzner et. al. (1997, 1999).

2.3.2 Expected Shortfall

The expected shortfall (ES) or tail conditional expectation, according to McNeil and Frey (2001) is a risk measure that gives some information about the size of the potential losses given that a loss bigger than VaR has occurred. The tail conditional expectation measures the expected loss given that the loss L exceeds VaR; in mathematical terms it is given by $E[L | L > VaR]$. This risk-metric has the advantage over the VaR that is a coherent measure under the Artzner et. al. (1997, 1999). criteria.

2.4 Extreme Value Theory - why is it used?

Extreme Value Theory (EVT) is a branch of statistics that studies “rare” or extreme events. It has been proofed for modeling catastrophic events in insurance and finance. It is especially well suited to describe the “fat-tails” of the profit and losses distributions typically found in stock returns.
EVT has experienced nowadays a boom in the financial field, especially with respect to risk management, particularly with its application to VaR estimation. It provides theoretically robust models with accessible application methods and constitutes an alternative to models using the assumption of normality of financial markets, which do not reflect the reality of the situation and can produce particularly wrong estimates for the VaR.

Furthermore, if we want to focus on what we call extreme risk it is of particular interest to use appropriate models to estimate extreme events as opposed to “normal” situations. McNeil et al (2001) points out two features that make EVT-based methods attractive for tail estimation: They are based on a sound statistical theory, and they offer a parametric form for the tail of a distribution, allowing for some extrapolation beyond the range of the data, even when special care is required at this point. Another attractive of EVT is that it does not assume a particular distribution for the returns and therefore the model risk is considerably reduced.

In this thesis we use EVT-methods in order to model the tails of the returns’ distributions of the stock indices on which we base the empirical work. We compare its outputs to alternative methods using the normality assumption and arrived to the same conclusion that various authors have come to: EVT is a good alternative to model extreme risks and stresses the fact that relying on a Gaussian assumption when dealing with fat-tailed events can lead to very dangerous miss-estimations of risks.

2.5 Time Series Analysis

The motivation to use time series analysis (TSA) in this thesis is primarily to acknowledge the fact that the variables of analysis, which in this case are equity returns, are constantly
changing in time, not only in the sense of not being constant values, but also recognize that the factors that may govern or affect their behavior may transform in time.

A time series is simply a sequence of values usually recorded at regular increasing intervals (yearly, monthly, weekly, ... secondly). According to Campbell, Lo and MacKinlay (1997) - Tsay (2005) page 2- In financial econometrics it is more common to work with series of returns as opposed to series of prices for two reasons: first, for average investors, the return of an asset is a complete and scale-free summary of the investment opportunity. Second, return series are easier to handle than prices given its “better” statistical properties.

Time series analysis can be useful to see how a given asset, security or economic variable changes over time or how it changes compared to other variables over the same time period. In particular, TSA is used for two main purposes: (a) identifying and characterizing the nature of the analyzed variables and (b) forecasting; which means predicting future values of the time series variable. Both of these goals require identifying certain patterns in the time series data and then try to formally describe and model them.

Besides the motivation to use TSA ‘per se’ of including the time variation of the factors influencing equity returns, there is also an important relation between TSA and EVT which motivates their joint implementation. The EVT methods that we apply in this thesis for modeling the tails of the return’s distributions assume that the series of returns are independent and identically distributed (i.i.d.) random variables; however market data returns usually present the so-called stylized facts such as: heavy tails, clustered extremes and random volatility, which means the i.i.d. assumption does not hold in general. In order to correct for this features we first filter the returns using TSA techniques to get i.i.d. time series
and then apply the EVT method to model the tails of the new filtered series. This joint use of TSA and EVT methods is motivated by the results on McNeil and Frey (2001). For further details on this refer to section 3.2.1.

2.6 Dependence

The concept of dependence is a very intuitive and common one. The presence of dependence can be described as the state of being determined, influenced, or controlled by something else. In portfolio optimization is of great interest to take into account the dependence among the assets used to construct the portfolio, since the benefits of diversification come specifically from the joint variation on the assets’ values in time.

In order to integrate the dependence on the portfolio optimization process we first need to determine the best way of quantifying this dependence. The most common measure of dependence is the one used by Markowitz mean-variance model called Linear Correlation (or Pearson Correlation). However, given that this measure has certain drawbacks that will be pointed out, we will use a more complete method to model dependence called copulas.

2.6.1 Linear Correlation

Embrechts, Lindskog and McNeil (2001) give the following definition:

Let \((X, Y)^T\) be a vector of random variables with nonzero finite variances. The linear correlation coefficient for \((X, Y)^T\) is

\[
\rho(Y, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}},
\]

where \(\text{Cov}(X, Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y)\) is the covariance of \((X, Y)^T\), and \(\text{Var}(X)\) and \(\text{Var}(Y)\) are the variances of \(X\) and \(Y\).
Linear correlation is a measure of linear dependence. In the case of perfect linear dependence, i.e., $Y = aX + b$, we have $|\rho(X, Y)| = 1$ and the converse holds. Otherwise, $-1 < \rho(X, Y) < 1$.

Linear correlation is invariant under strictly increasing transformations and easily manipulated under linear operations. For $\alpha \in \mathbb{R}^n$, 

$$Var(\alpha^T X) = \alpha^T \text{Cov}(X) \alpha,$$

where $\text{Cov}(X) := \text{Cov}(X, X)$. Hence the variance of a linear combination is fully determined by pair wise covariance of the components. This is a very important property in MPT.

Pearson correlation is a z-score based dependence measure and has the underlying assumption that the two variables analyzed follow a bivariate normal distribution (see Cohen 2001 for details). We find that this assumption is strongly violated by the equity returns analyzed; hence we use a more general method to model dependence called copulas.

### 2.6.2 Copulas

According to Meucci (2005), the copula represents the true interdependence structure of a (multivariate) random variable, which in our application is the market. Intuitively, the copula is a standardized version of the purely joint features of a multivariate distribution, which is obtained by filtering out all the purely one-dimensional features, namely the marginal distribution on each equity index $X_n$.

The popularity of Copulas as a dependence model in finance has experienced an important growth. As Mikosch (2007) points out, we can find various Copula application in numerous finance related academic papers and software applications for the main statistical and quantitative packages (including MatLab, Splus, R, Mathematica). In addition, the
International Actuarial Association (2004) in its paper on Solvency II\(^4\) recommends using copulas for modeling dependence in insurance portfolios.

Moody's also uses Gaussian copulas for modeling credit risk and provides software for it which is used in many financial institutions. According to Mikosch (2007), since Basle II copulas are now standard tools in credit risk management.

### 2.6.3 Asymptotic Dependence

The conventional dependence measure, Pearson correlation, is appropriate only for linear association. As we can see from its definition, this “average risk” measure is constructed as an average of deviations from the mean, which implies that the weight assigned to extreme realizations is the same as for all of the other observations in the sample.

Hence it is important to notice if the dependence characteristics among extreme observations differ from the rest in the sample, otherwise the conclusions drawn from the Pearson correlation could result in a financial institution risking bankruptcy (Poon, Rockinger and Tawn (2003). Poon et. al. argue that the Pearson correlation is not a good measure of dependency in cases where extreme realizations are important and propose the use of alternative “extreme dependence” measures based on statistical developments in extreme value theory. They also describe the dependence between to variables in a more complete way, describing not only the degree of dependence but also the type of dependence.

---

\(^4\) Solvency II will be a treaty for insurers similar to the Basle I and II treaties for banks.
Extreme dependence can be understood as the chance of obtaining large values of the two variables analyzed. It is important to understand extreme dependence, to appreciate the form and degree of such dependence because this will determine the way joint movements should be modeled.

Poon, Rockinger and Tawn (2003) present two classes of extreme (value) dependence, asymptotic dependence and asymptotic independence, for which the characteristics of events behave differently as the events become more extreme. Both forms of extremal dependence permit dependence between moderately large values of each variable, but the very largest values from each variable can occur together only when the variables exhibit asymptotic dependence.

The reason we decided to analyze extreme dependence in some degree of detail, by looking at alternative ways of assessing it, is to determine an appropriate way to model the dependence among assets. This point is particularly stressed by Poon, Rockinger and Tawn (2003): They show that conventional multivariate extreme value theory has emphasize is a particular dependence class (the asymptotically dependent class) resulting in its wide use in all the finance applications. However, if the series do not correspond to this kind of dependence, such an approach will result in the over-estimation of extreme value dependence, and consequently of the financial risk.

Furthermore, they show empirically that both kinds of dependence are present in financial data, suggesting that very often the variables are found to be asymptotically independent but not exactly independent.
Given the importance of dependence characteristics amongst the assets that will conform the portfolio, we will look at the extreme dependence tools proposed by Poon, Rockinger and Tawn (2003) and used them to select an appropriate way to model this dependence. A more detailed definition of the extreme dependence measures will be given in section 4.5.1
3 Methodology and Results

This section has two purposes: first to briefly present the main results and conclusions of some of the research literature that motivated the procedures implemented in this thesis. Secondly we present to some extent the theory behind the methods used and report the results of the empirical protocol implemented. As mentioned before, we perform a portfolio optimization exercise using a stochastic scenario generation techniques that take into account for the stylized facts of equity return series – such as fat-tails, volatility clustering and non-linear dependence.

When creating a portfolio of assets, we face a high-dimensionality problem, because the universe of assets available to invest is virtually infinite. In order to overcome this issue, we restrict the problem in two ways. First, for data availability reasons, we decided to build up our portfolio using equity of the US market. Second, as Broadin and Klüppelberg (2006) suggest, instead of looking at individual stocks, we use a few corresponding risk factors, by grouping them into economic sectors.

3.1 Data description

We use five industry portfolios for the US equity markets. These portfolios are constructed on a value-weighting scheme using the stocks composing the NYSE, AMEX, and NASDAQ indices. We took the data from Kenneth R. French data library web page⁵. They group the stocks within NYSE, AMEX, and NASDAQ in industry portfolios at the end of June of each year based on their four-digit SIC code at that time.

The five industry portfolios correspond to the following sectors:

⁵ http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
1. Cnsmr: Consumer Durables, Non-Durables, Wholesale, Retail and Some Services (Laundries, Repair Shops).
4. Hlth: Healthcare, Medical Equipment and Drugs.

We used all the daily data available for the five industry indices at the beginning of this research project. There are 10951 trading days starting the 1st of July 1963 to the 29th of December 2006. In Figure 1 there is an index for each industry portfolio of returns with an initial nominal value of 100. This represents the evolution of the nominal value of 100 dollars invested at the beginning of the period in each of the industry portfolios.

![Index Cumulative Returns](image)

Figure 1: Indices created using the five industry portfolios’ returns.

In figure 2 we see the daily returns of the HiTec portfolio. In Appendix A. we can see the daily returns of the other four portfolios.
Figure 2: Daily returns of the HiTec index

Table 1 shows some descriptive statistics of the five series of returns. We can see that all of them present a common feature: they are all leptokurtic and negatively skewed, which suggest a clear departure from “normality”; a formal test of the departure from normality will be presented later on.

<table>
<thead>
<tr>
<th>Stats</th>
<th>Csmr</th>
<th>Manuf</th>
<th>HiTec</th>
<th>Hlth</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0473%</td>
<td>0.0466%</td>
<td>0.0430%</td>
<td>0.0520%</td>
<td>0.0479%</td>
</tr>
<tr>
<td>Volatility</td>
<td>0.90%</td>
<td>0.84%</td>
<td>1.20%</td>
<td>1.06%</td>
<td>0.94%</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>19.271</td>
<td>27.507</td>
<td>16.017</td>
<td>14.877</td>
<td>15.080</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.636</td>
<td>-0.985</td>
<td>-0.119</td>
<td>-0.470</td>
<td>-0.474</td>
</tr>
<tr>
<td>Min</td>
<td>-16.84%</td>
<td>-18.22%</td>
<td>-19.33%</td>
<td>-17.89%</td>
<td>-15.63%</td>
</tr>
<tr>
<td>Max</td>
<td>9.12%</td>
<td>8.74%</td>
<td>14.24%</td>
<td>9.96%</td>
<td>8.24%</td>
</tr>
<tr>
<td>Median</td>
<td>0.06%</td>
<td>0.06%</td>
<td>0.05%</td>
<td>0.06%</td>
<td>0.08%</td>
</tr>
</tbody>
</table>

Table 1: Descriptive Statistics of the five industry portfolios.

3.2 Motivation and formal procedure

In this thesis we integrate different models coming from different branches of probability and statistics and put them at the service of portfolio managers in a model that allow us to estimate and minimize risk. In what follows, we present several papers that motivate the use of the different procedures that we implement here and show the relations and natural integration of the different methodologies used.
3.2.1 EVT and Time Series Analysis

As we mention in section 2.5, we could argue that there is a motivation to use time series models when we want to use EVT-methods. This is related to the assumption of i.i.d. series made by some EVT-based methods and the possibility to filter equity returns to (approximately) turn them into i.i.d. series. This is clearly shown by McNeil and Frey (2001); they propose a method for estimating VaR and ES modeling the tail of the conditional distribution of a heteroskedastic\(^6\) financial return series (including equity returns). They combine quasi maximum likelihood fitting of GARCH models to estimate the current volatility and extreme value theory (EVT) for estimating the tail of the innovation distribution of the GARCH model. They find that this procedure gives better estimates than methods which ignore the heavy tails of the innovations or the stochastic nature of the volatility. They mention than their “two-stage-method" can be extended to multivariate series, just as we did in this thesis.

Analogously, Nyström and Skoglund (2002) combine ARMA-(asymmetric) GARCH and EVT methods, applying them to the estimation of extreme quantiles univariate portfolio risk factors and obtain VaR and ES. They find that for the higher quantiles (97-98%) the use of EVT does indeed give a substantial contribution and the generalized Pareto distributions are more able than the normal distribution to accurately model the empirically observed tail.

3.2.2 EVT and Copulas

According to Broadin and Klüppelberg (2006), if one is concerned to estimate large joint losses, one should model a portfolio as a multivariate random vector. Unfortunately, this tends

\(^6\) See section 4.2 for the explanation of this term.
easily to a very high-dimensional problem, and modeling of the dependence structure could be very difficult or even impossible. They said that one way out is to use comparably few selected risk factors, or to group assets in different sectors or geographical regions. In this thesis we use five sectors portfolios for the US equity markets (using the stocks composing the NYSE, AMEX, and NASDAQ\(^7\) indices).

Mourany and Mukherji (2007) find that the traditional VaR models, assuming a multivariate conditional normal distribution for risk factor returns, underestimate portfolio VaR, and motivate the use of copulas and EVT to model equity returns when estimating a portfolio’s VaR. They come to the conclusion that the copula/EVT-based approach outperforms the traditional VaR models assuming a conditional normal multivariate distribution for risk factor log-returns.

### 3.2.2.1 The t-student Copula

Embrechts, Lindskog and McNeil (2001) are keen to emphasize the potential usefulness of t-copulas as an alternative to Gaussian copulas. They argue that although Gaussian and t-copulas are easily parameterized by the linear correlation matrix, only t-copulas yield dependence structures with tail dependence. Besides the analytical illustration presented of this situation, they contrast real data with simulations from both copulas. They note that the Gaussian copula does not get the extreme joint tail observations clearly present in the real data. Furthermore, they highlight the fact that the t2-copula\(^8\) seems to be able to do a much better job in that respect. They also conclude that the scatter plot of the t2 simulation shows most of the graphical features in the real data.

---

\(^7\) See Kenneth R. French data library web page for details: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

\(^8\) The t2 stands for a t-student with 2 degrees of freedom.
Martellini and Meyfredi (2006) implement a multi-variate copula approach to Value-at-Risk estimation for fixed-income portfolios. They use a parsimonious model to extract time-varying parameters for the factors affecting the yield curve, and a t-student copula to model the dependence structure of these factors. They find that their VaR estimates strongly dominate standard VaR ones.

Mashal and Zeevi (2002) investigate the potential for extreme co-movements between financial assets by directly testing the underlying dependence structure. In particular, a t-copula is used as a proxy to test for this extremal behavior. They show that the presence of extreme co-movements is statistically significant in three asset markets (equities, currencies, and commodities), as well as across international (G5) equity markets. Their results indicate that the “correlation-based” Gaussian dependence structure is not supported on the basis of observed asset co-movements.

The t-student copula is classified as an elliptical copula. In this sense, Klüppelberg, Kuhn and Peng (2006) point out that there has been an increasing interest in applying elliptical distributions to risk management; this situation is in line with the result by Hult and Lindskog (2002) who show that under some conditions, a random vector with an elliptical distribution is in the domain of attraction of a multivariate extreme value distribution.

3.2.3 EVT, Extreme Dependence and Asset Allocation

According to Poon, Rockinger and Tawn (2003) conventional multivariate extreme value theory has emphasized the asymptotically dependent class resulting in its wide use in all the finance applications. If the series are truly asymptotically independent, such an approach will result in the over-estimation of extreme value dependence, and consequently of the financial
risk. The degree of this over-estimation depends on the degree of asymptotic independence. They find left-tail dependence to be usually stronger than right-tail dependence. In addition, they demonstrate that many of these stock index returns do not exhibit asymptotic dependence, suggesting that much of the extreme value dependence reported in previous studies is likely to be over-estimated. With the use of volatility filters, they find that most of the extreme value dependence is caused by changing stock market volatility, but detect little difference in the results produced by the three volatility filters.

Bradley and Taqqu (2004) investigate the portfolio construction problem for risk-averse investors seeking to minimize quantile based measures of risk. Using the dependence measures from EVT $\chi$ and $\bar{\chi}$, they find that most international equity markets are asymptotically independent. They also find that the few cases of asymptotic dependence occur mostly in markets which are in close geographic proximity. They point out that when markets are asymptotically independent, there is a greater benefit to diversification at the highest confidence levels $\alpha$ than there is at lower $\alpha$.

Bradley and Taqqu develop a methodology for asset allocation where the goal is to guard against catastrophic losses. The examination of the two market allocation problem relied on an optimization over a dense grid of possible allocations. When considering many assets, this approach becomes numerically intractable. For the asset allocation procedure, we also minimize extreme risk but we use a Monte Carlo simulation-based optimization instead.

In another paper, Bradley and Taqqu (2005) consider investors attempting to guard against large losses and use techniques from extreme value theory to compare univariate and multivariate approaches to the asset allocation problem. They conclude that the multivariate
EVT approach, with the “standard choice” of the logistic distribution to model the multivariate extremes will overstate the risk of a diversified portfolio. This is because the inference stage of this model almost always leads to a model of the tail with asymptotically dependent marginal components.

According to this, it is then important to determine if the variables under analysis present asymptotic dependency or not. In case we find that the asymptotic dependency is present amongst all equity indices, there would be enough motivation to use a multivariate extreme value theory approach. On the other hand, if the variables turn out to be asymptotically independent we may rather opt for a more conventional dependence structure, because the multivariate EVT way might lead to risk overestimation.

As we will show in the following sections, our analysis of asymptotic dependence is in line with the findings of Poon et al. (2002) and Bradley and Taqqu (2004): most financial series in our case are asymptotically independent. Furthermore, as in Poon et al. (2002) we find that this asymptotic independence becomes more evident when filtering out the time variation of volatility. Following the conclusion of Poon et. al (2002) and Bradley and Taqqu (2005), a multivariate EVT approach does not appear to be the right way to model the equity assets we are looking at. Instead, we choose the t-copula copula, very popular in risk management, to model the dependence among the assets of the portfolio.

3.2.4 EVT, Copulas and Portfolio Optimization

Di Clemente and Romano (2003) explore how the optimal portfolio composition with respect to ES may change when assuming different hypotheses for generating asset return scenarios. In order to achieve this purpose, primarily they generate scenarios for portfolio asset returns
assuming the traditional hypothesis of multivariate conditional normal distribution. Successively, they generate the scenarios from the empirical distribution using Filtered Historical Simulation (FHS). Finally, they employ Monte Carlo asset return scenarios by using the Student’s t-copula and marginal distributions normal in the centre and EVT distributed in the tails.

Di Clemente and Romano follow the technique described by Rockafellar and Uryasev (1999) to solve portfolio optimization problems when minimizing the $ES$. Rockafellar and Uryasev demonstrated how the minimization of portfolio’s $ES$ is obtained by solving a stochastic optimization problem. This minimization produces a relevant VaR reduction too. When the uncertainty is modeled through the generation of a finite number of Monte Carlo or historical scenarios for the portfolio asset log-returns, the problem above may be solved using a linear programming technique.

They conclude that when considering the fact that asset return distributions are far from normality, in particular due to the fat tails, the minimization of the portfolio’s variance produces misleading results in terms of capital allocation. In particular, the 99% $ES$ minimization produces a different capital allocation with respect to the one obtained by Markowitz model.

In this thesis we implemented the same portfolio optimization method to minimize extreme risk (estimated with $ES$). We compare the results with a Markowitz kind of portfolio and also find a different allocation when taking into account departure from normality and minimizing extreme risk as opposed to average risk.
Our general procedure is close to the methodology presented by Di Clemente and Romano, however we use TSA to model the time variation of the risk factors and use an kernel non-parametric distribution instead of the normal distribution for the center of the marginals, in an attempt to be consistent with the non-normality assumption.

Given the motivation that papers cited give to put together the models, we can implement them and aligne them towards the purpose of constructing a portfolio minimizing extreme risk and using better estimates for the risk. We summarize the procedure we use to generate the stochastic scenarios in the following steps:

1. Filter the returns by autocorrelation and heteroskedasticity, if any, using TSA models. With this procedure we obtain approximately i.i.d. series, as required by the EVT model.

2. Fit a semi-parametric distribution to the standardized residuals. We use an EVT distribution for the tails and an empirical distribution (Kernel smoothing approximation) for the center of the distribution of each asset.

3. Transform the standardized residuals to uniform distributed variables using the c.d.f. of the semi-parametric distribution fitted in step 2.

4. Fit a t-student copula to the uniform variables.

5. Simulate a high number of random scenarios uniformly distributed for the five variables with the t-student dependence structure fitted in step 4.

6. Transform the uniform variables using the inverse c.d.f. of the semi-parametric distribution adjusted to each asset in step 2.

7. Re-introduce the autocorrelation and heteroscedasticity to the random variables using the TSA models of step 1.
Finally we use the random scenarios generated in the simulation process as an input for the portfolio optimization. We construct the portfolio minimizing $ES$ using the approach proposed by Rockafellar and Uryasev (1999).

Now we summarize the theory behind the models we use to generate possible future scenarios of the assets in the portfolio and risk measures for them. We use the output of this models as a mean to take into account the uncertainty of the future value of the assets as accurately as possible to construct a portfolio with minimum risk.

### 3.3 Time Series modeling

In what follow we apply the same procedure to each of the five assets (the industry portfolios) that will compose the final portfolio. Although they are all different variables, they share several features due to the fact that all of them come from equity return series.

According to Tsay (2005) the foundation of TSA is stationarity. A time series $\{r_t\}$ is strictly stationary if the joint distribution of $\{r_{t-1}, \ldots, r_t\}$ is identical to that of $\{r_{t-k}, \ldots, r_{t-1}\}$ for all $t$, where $k$ is an arbitrary positive number. This is a very strong assumption that is hard to verify empirically. A weaker version of stationarity is often assumed. A time series $\{r_t\}$ is weakly stationary if both the mean of $r_t$ and the covariance between $r_t$ and $r_{t-l}$ are time-invariant, where $l$ is an arbitrary integer. The weak stationarity implies that the time plot of the data would show that the $T$ values fluctuate with constant variation around a fixed level.
Tsay (2005) define a time series $r_t$ as white noise if \( \{r_t\} \) is a sequence of independent and identically distributed (i.i.d.) random variables with finite mean and variance. In practice, if all sample ACFs are close to zero, then the series is a white noise series.

Let \((X_i^t, t \in \mathbb{Z})\) where \(i = \{1, 2, 3, 4, 5\}\), be strictly stationary time series representing daily observations of the log return on the \(i\)th equity industry portfolio. We assume that the dynamics of each \(X^i\) are given by

\[
X_i^t = \mu_i + \sigma_i Z_i^t,
\]

where \(\mu_i\) and \(\sigma_i\) are the mean and standard deviation the process \(X^i\) and the innovations \(Z_i\) are strict white noise process with zero mean, unit variance and marginal distribution function \(F_Z(z)\). We assume that \(\mu_i\) and \(\sigma_i\) are measurable with respect to \(G_{t-1}\), the information about the return process available up to time \(t-1\).

To implement an estimation procedure for quintile-based risk measures such as VaR and ES we will choose a specific model for the dynamics of the conditional mean and volatility. As McNeil and Frey (2000) point out, many different models for volatility dynamics have been proposed in the econometric literature including models from the ARCH/GARCH family (Bollerslev, Chou, and Kroner 1992), HARCH processes (Müller, Dararogna, Davé, Olsen, Pictet, and von Weizsäcker 1997) and stochastic volatility models (Shephard 1996). In this thesis we use the parsimonious but effective GARCH(1,1) process for the volatility and look at ARMA type models for the conditional mean.

When estimating quintile-based risk measures with GARCH-type models it is commonly assumed that the innovation distribution is standard normal. A GARCH-type model with
normal innovations can be fitted by maximum likelihood (ML) and $\mu_{t+1}$ and $\sigma_{t+1}$ can be estimated using standard 1-step forecasts, so that an estimate of the quintile $x_q^t$ is easily constructed. This is close in spirit to the approach advocated in J.P. Morgan's Riskmetrics, but according to McNeil and Frey (2001) empirical finding, that approach often underestimates the conditional quantile for $q > 0.95$; the distribution of the innovations seems generally to be heavier-tailed or more leptokurtic than the normal.

Another standard approach is to assume that the innovations have a leptokurtic distribution such as Student's t distribution (scaled to have variance 1). GARCH-type models with t-innovations can also be fitted with maximum likelihood and the additional parameter $\nu$ (degrees of freedom) can be estimated. McNeil and Frey (2001) point out that this method can be viewed as a special case of the approach we implemented here (the same as McNeil and Frey); they argue that it yields quite satisfactory results as long as the positive and the negative tail of the return distribution are (roughly) equal.

### 3.3.1 ARMA and GARCH models

Consider the ARMA(p,q) model

$$
\mu_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \xi_j \epsilon_{t-j} + \epsilon_t
$$

where $\epsilon_t$ is independent and identically distributed with mean zero and variance $\sigma^2$. According to Nyström and Skoglund (2002) ARMA models are traditionally employed in time series analysis to capture (linear) serial dependence. They allow conditioning of the mean process on past realizations and are generally successful for the short-term prediction of time series.
The assumption of conditional homoskedasticity is however too restrictive for financial data where we typically observe volatility clusters, implying that a large absolute return is often followed by more large absolute returns. A GARCH model for the conditional variance process extends the simple ARMA model by assuming that

\[ \epsilon_t = z_t \sigma_t \]

where \( z_t \) is independent and identically distributed with mean zero and unit variance and \( z_t, \sigma_t \) are stochastically independent. GARCH models try to capture the dynamics of \( \sigma_t \), the conditional variance at time \( t \). The GARCH model cannot only capture the volatility clustering of financial data but also to some extent excess kurtosis, since

\[ k_4 = \frac{E \epsilon_t^4}{(E \epsilon_t^2)^2} = v_4 \frac{E \sigma_t^4}{(E \sigma_t^2)^2} \geq v_4 \]

where \( v_4 \) is the kurtosis of \( z_t \).

The GARCH model more commonly used in financial applications in its most simple form, the GARCH (1,1) model, in which the conditional variance is given by

\[ \sigma_t^2 = \sigma_0 + a_1 \epsilon_{t-1}^2 + b \sigma_{t-1}^2 \]

\[ = \sigma_0 + (a_1 + b) \sigma_{t-1}^2 + a_1 (\epsilon_{t-1}^2 - \sigma_{t-1}^2) \]

the term

\[ (\epsilon_{t-1}^2 - \sigma_{t-1}^2) = \sigma_{t-1}^2 (z_{t-1}^2 - 1) \]

has zero mean conditional on past information and can be interpreted as the shock to volatility. The coefficient \( a_1 \) therefore measures the extent to which a volatility shock in period \( j \) feeds through into the volatility in period \( j+1 \), while \( (a_1 + b) \) measures the rate at which this effect dies out in time.
The GARCH(1,1) model can also be written in terms of the squared errors, in that form is clear that GARCH(1,1) is really an ARMA(1,1) for the squared returns, but with heteroskedastic shocks. We refer the reader to Nyström and Skoglund (2002) and Tsay (2005) for an extensive presentation and interpretation on GARCH models.

According to McNeil and Frey (2001) the mean-adjusted series \((\varepsilon_t)\) is strictly stationary if

\[ E[\log(b + a_t \varepsilon_i^2)] < 0. \]

by using Jensen’s inequality and the convexity of \(-\log(x)\) it is seen that at sufficient condition for this to hold is that \(b + a_t < 1\), which moreover ensures that the marginal distribution \(F_x(x)\) has a finite second moment.

We fit the ARMA/GARCH models using the pseudo-maximum-likelihood (PML) method as implemented in McNeil and Frey. This means that to obtain the parameter estimates we maximize the likelihood for a GARCH(1,1) model with normal innovations. Whilst this resembles to fitting a model using an assumption over the distribution, we do not necessarily believe, the PML method delivers reasonable parameter estimates. As McNeil and Frey (2001) point out, in fact, it can be shown that the PML method yields a consistent and asymptotically normal estimator; see for instance Chapter 4 of Gouriéroux (1997).

### 3.3.2 Model selection

As mentioned before, one key assumption in EVT is that extreme returns are i.i.d. series. Hence, before using an EVT method we may want to check that our observations are approximately i.i.d. In case we find that they don’t fulfill this hypothesis we would have to
apply a filter before fitting them to the EVT model. In particular, it is very common to find some degree of serial correlation and more importantly, heteroskedasticity in financial return series, which leads to pursue a test of these stylized facts on the different industry return series.

In this sense, we will check if the i.i.d. hypothesis holds and determine the adequate filter to apply in case we are looking at returns suffering from autocorrelation and heteroskedasticity. These two characteristics can be filtered away by applying Autoregressive (AR) and/or Moving Average (MA) models for the conditional mean and Generalized Autoregressive Conditional Heteroskedastic (GARCH) models for the conditional variance.

An examination of the sample autocorrelation functions (ACF) of the index returns and squared returns and the partial autocorrelation functions (PACF) of the returns can help to choose adequate models to filter the data.

The ACF is a related concept to linear correlation. The ACF is simply the set of correlation coefficients between \( r_t \) and \( r_{t-l} \); called the \( l \)-lag autocorrelations of \( r_t \) and are commonly denoted by \( \rho_l \). The \( l \)-lag sample PACF of \( r_t \) can be interpreted as the measure of the added contribution of \( r_{t-l} \) to \( r_t \) over the AR\((t-l)\) model\(^9\).

In this sense, we can use the ACF and PACF to determine the order of the models for the conditional mean and the ACF of squared returns to see if there is evidence of persistence in variance. In theory, for an AR\((p)\) model, the lag-\(p\) sample PACF should not be zero, but the

\(^9\) we refer the reader to Tsay (2005) for precise definitions of ACF and PACF
\( \hat{\phi}_{j,j} \) should be close to zero for all \( j > p \). We make use of this property to determine the order \( p \).

For a MA(\( q \)) model the ACF is useful to determine the order \( q \). For a time series \( r_t \) with ACF \( \rho_l \), if \( \rho_q \neq 0 \), but \( \rho_l = 0 \) for \( l > q \), then \( r_t \) follows an MA(\( q \)) model.

By looking at the ACF and PACF of the five industry returns (see Figures 3 and 4 below and figures... in Appendix A) we can see that in all indices there is evidence of serial correlation on the first lag. However, it is not clear if we should use an AR(1), MA(1) or ARMA(1,1) for the conditional mean.

Figure 3: Sample ACF of the ‘Other’ industry index returns.
Additionally, the sample ACF of the squared returns (see figures 5 and 6 below and Appendix A) suggest that all industry indices exhibit heteroskedasticity, and implies that a GARCH modeling may significantly condition the data used in the subsequent tail estimation process.
After observing the ACF of the returns and squared returns and trying to keep a certain parsimony on the modeling (keep it as simple as possible) we decided to test and compare AR(1), MA(1) and ARMA (1,1) for the conditional means plus a GARCH(1,1) model for the conditional volatility.

In particular the AR(1) model for the conditional mean is described by the following equation

$$\mu_t = \phi X_{t-1}$$

and the MA(1) is given by

$$\mu_t = c_0 + \epsilon_t + \theta_t \epsilon_{t-1}$$

In order to choose the models to for $\mu_t$ and $\sigma_t$ we estimated the parameters of each model by Pseudo Maximum Likelihood (PML) and then calculated the Bayesian Information Criteria (BIC). We used this criterion to choose the model, as opposed to the Akaike Information Criteria for instance, because this criterion penalizes stronger for the number of parameters that are need to be estimated. The reason for this penalization is to keep parsimony in the modeling. The BIC criteria is given by

$$BIC = -2 \cdot LLF + (NumParams \times \log(NumObs))$$
where the LLF, NumParams and NumObs are the Log-likelihood function, the number of parameters of the model and the number of observations used to estimate these parameters respectively. The better model according to this criterion is the one with lower BIC.

We estimated the parameter and calculated this criterion for the following models:

1. AR(1)
2. MA(1)
3. ARMA(1,1)
4. AR(1)+GARCH(1,1)
5. MA(1)+GARCH(1,1)
6. ARMA(1,1)+GARCH(1,1)

The result is showed in tables 2 and 3 below.

<table>
<thead>
<tr>
<th></th>
<th>'Cnsmr'</th>
<th>'Manuf'</th>
<th>'HiTec'</th>
<th>'Hlth'</th>
<th>'Other'</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>-7.22885616778440</td>
<td>-7.370144247584434</td>
<td>-6.573259628882429</td>
<td>-6.875365722269197</td>
<td>-7.147789887900512</td>
</tr>
<tr>
<td>MA</td>
<td>-7.229655918112579</td>
<td>-7.371670587713525</td>
<td>-6.573523621853084</td>
<td>-6.877052713849377</td>
<td>-7.148348653289962</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>-7.22875180130070</td>
<td>-7.371505071077546</td>
<td>-6.573226679033676</td>
<td>-6.876202919756019</td>
<td>-7.147458546487034</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>'Cnsmr'</th>
<th>'Manuf'</th>
<th>'HiTec'</th>
<th>'Hlth'</th>
<th>'Other'</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA</td>
<td>-7.504528934127991</td>
<td>-7.69154090482909</td>
<td>-7.004999392129530</td>
<td>-7.169018841176279</td>
<td>-7.491846765870602</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>-7.503803794991338</td>
<td>-7.69062246894817</td>
<td>-7.004231817405697</td>
<td>-7.168521516194150</td>
<td>-7.492251477520460</td>
</tr>
</tbody>
</table>

Table 2: BIC criterion for AR(1), MA(1) and ARMA(1,1)

According to the BIC criteria the models chosen are an AR(1)-GARCH(1,1) for the ‘Cnsmr’, ‘Hlth’ and ‘Other’ industry indices and a MA(1)-GARCH(1,1) for the ‘Manuf’ and ‘HiTec’ sector indices.
### 3.3.3 Time Series Results

We estimated the parameters of the models we tested using the PML method. As we can see from Tables 4 and 5, the t-stats of the estimated parameters further confirm the selection made by the criteria, given that the significance is stronger for the corresponding models.

<table>
<thead>
<tr>
<th>'Cnsmr'</th>
<th>'Manuf'</th>
<th>'HiTec'</th>
<th>'Hlth'</th>
<th>'Other'</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR 1</td>
<td>28.3536</td>
<td>28.4902</td>
<td>11.1463</td>
<td>28.9661</td>
</tr>
<tr>
<td>MA 1</td>
<td>28.8564</td>
<td>31.3068</td>
<td>11.8438</td>
<td>30.7182</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>-1.0631</td>
<td>-6.1607</td>
<td>-3.9146</td>
<td>-1.4615</td>
</tr>
<tr>
<td></td>
<td>6.1576</td>
<td>11.4663</td>
<td>4.8898</td>
<td>7.1043</td>
</tr>
</tbody>
</table>

*Table 4: T-stats of the estimated parameters of AR(1), MA(1) and ARMA(1,1)*

<table>
<thead>
<tr>
<th>'Cnsmr'</th>
<th>'Manuf'</th>
<th>'HiTec'</th>
<th>'Hlth'</th>
<th>'Other'</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR 1</td>
<td>16.9869</td>
<td>16.5653</td>
<td>8.8661</td>
<td>19.6031</td>
</tr>
<tr>
<td>MA 1</td>
<td>16.7949</td>
<td>17.1250</td>
<td>9.1540</td>
<td>19.5329</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>1.7664</td>
<td>0.2244</td>
<td>-1.5729</td>
<td>2.1487</td>
</tr>
<tr>
<td></td>
<td>1.1866</td>
<td>2.7257</td>
<td>2.5171</td>
<td>1.7809</td>
</tr>
</tbody>
</table>

*Table 5: T-stats of the estimated parameters of AR(1), MA(1) and ARMA(1,1) with GARCH (1,1)*

As mentioned before, we can check the stationarity of the fitted model by verifying that $\hat{b} + \hat{a}_1 < 1$. If we sum the last two columns of table 6 we find that this assumption holds for all equity indices.

<table>
<thead>
<tr>
<th>Index</th>
<th>Mean Model</th>
<th>Parameter</th>
<th>GARCH</th>
<th>ARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cnsmr</td>
<td>AR(1)</td>
<td>0.1671</td>
<td>0.9099</td>
<td>0.0793</td>
</tr>
<tr>
<td>Manuf</td>
<td>MA(1)</td>
<td>0.1680</td>
<td>0.9154</td>
<td>0.0818</td>
</tr>
<tr>
<td>HiTec</td>
<td>MA(1)</td>
<td>0.0930</td>
<td>0.9272</td>
<td>0.0682</td>
</tr>
<tr>
<td>Hlth</td>
<td>AR(1)</td>
<td>0.1920</td>
<td>0.9148</td>
<td>0.0788</td>
</tr>
<tr>
<td>Other</td>
<td>AR(1)</td>
<td>0.2311</td>
<td>0.8862</td>
<td>0.1018</td>
</tr>
</tbody>
</table>

*Table 6: Estimated parameters of AR(1), MA(1) and ARMA(1,1) with GARCH (1,1)*
Estimates of the conditional mean and standard deviation series can be calculated recursively with the AR or MA and GARCH models correspondingly. In figures 7 and 8 we show the residual series for two of the indices and the corresponding conditional standard deviations estimated from the GARCH model below them (in Appendix A are the same graphs for the rest of the indices).

Figure 7: Residuals and GARCH estimated Volatility for the "Manuf" industry index.
The standardized residuals are calculated both to check the adequacy of the GARCH modeling and to use in Stage 2 of the method. They are calculated as

$$ (z_{t-n+1}, \ldots, z_t) = \left( \frac{x_{t-n+1} - \hat{\mu}_{t-n+1}}{\hat{\sigma}_{t-n+1}}, \ldots, \frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t} \right), $$

and should be approximately i.i.d. if the fitter model did the job. A plot of the standardized residuals for two indices is shown in figures 9 and 10.
In order to investigate the i.i.d. hypothesis over the residuals of our model we examine the sample ACF of the residuals and squared residuals. The plots of these correlograms are in figures 11 and 12 and in appendix A. While the returns data are clearly not iid, by looking at the correlograms we can see that this assumption seems to hold for all the residuals of every index. We further explore this hypothesis with a formal test known as the Ljung-Box test.
The Ljung-Box lack-of-fit hypothesis test is based on the Q-statistic

\[ Q = N(N + 2) \sum_{k=1}^{L} r_k^2 / (N - k) \]

where \( N \) is the sample size, \( L \) the number of autocorrelation lags included in the statistic, and \( r_k^2 \) is the squared sample autocorrelation at lag \( k \). The Q-statistic is used as a lack-of-fit test for a departure from the white noise hypothesis. Under the null hypothesis that the model fit is adequate, the test statistic is asymptotically chi-square distributed.

We performed a Ljung-Box test on each of the standardized residuals of the indices for 22 lags (number of days in a month on the data). The results are presented in table 7.

<table>
<thead>
<tr>
<th>Ljung-Box Test</th>
<th>Cnsmr</th>
<th>Manuf</th>
<th>HiTec</th>
<th>Hlth</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td>0,1507394</td>
<td>0,9625063</td>
<td>0,25114962</td>
<td>0,25565453</td>
<td>0,00963784</td>
</tr>
<tr>
<td>Rejection at 5%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Rejection at 0.9%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Qstat</td>
<td>28,7973038</td>
<td>11,7406407</td>
<td>26,0124092</td>
<td>25,9079651</td>
<td>40,4255828</td>
</tr>
<tr>
<td>Critical Value</td>
<td>40,677537</td>
<td>40,677537</td>
<td>40,677537</td>
<td>40,677537</td>
<td>40,677537</td>
</tr>
<tr>
<td>Number of Lags</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Ljung-Box test on each of the standardized residuals

After this verification, we can now see that the i.i.d. test is satisfied reasonably well, which means that the purpose of our AR/MA-GARCH model in filtering the returns by
heteroskedasticity and autocorrelation was fairly accomplish. Given this, we end stage 1 by calculating estimates of the conditional mean and variance for day \( t + 1 \), which are the obvious 1-step forecasts

\[
\hat{\mu}_{t+1} = \hat{\phi}_t, \\
\hat{\sigma}_{t+1}^2 = \hat{a}_0 + \hat{a}_1 \hat{\varepsilon}_t^2 + \hat{b} \hat{\sigma}_t^2,
\]

where \( \hat{\varepsilon}_i = x_i - \hat{\mu}_i \).

3.4 Extreme Value Theory

Nyström and Skoglund (2002) emphasize the following general remarks about Extreme Value Theory:

1. EVT techniques concentrate on the behavior of the extreme observations. It can be used to estimate out of sample quantiles. Hence, being a theory of extrapolation.
2. The extreme value method does not assume a particular distribution for returns and therefore the model risk is considerably reduced.
3. The parameter estimates of the limit distributions depend on the number of extreme observations used. The choice of a threshold \( u \) should be large enough to satisfy the limit conditions that are assumed by the method (\( u \) tends towards infinity). Therefore, there is a trade-off between the number of observations, which need to be sufficient for the parameter estimation and the theoretical limit condition. Different methods of making this choice can be used, but in general estimation risk can be an issue.
4. An important assumption in the theory is that observations are i.i.d. Deviations from the assumption of i.i.d. random variables may result from trends, periodicity, autocorrelation and clustering. By one or more methods one has to try to filter out some residuals to which the estimation methods can be applied. Clustering effects
which may be the result of stochastic or time-dependent volatility may be handled using a two-step approach based on GARCH and EVT as described in this thesis.

These remarks made by Nyström and Skoglund (2002) confirm the relation between EVT and TSE and give further motivation to apply EVT methods to the residuals in order to capture the extreme risk observed in equity data.

In order to confirm the hypothesis that the residuals of our AR/MA-GARCH models are not normally distributed we looked at a quantile-quantile (QQ-plot) plot of each index. The QQ-plot displays a quantile to quantile plot of the sample quantiles of the variable analyzed versus theoretical quantiles from a normal distribution. If the distribution of the variable is normal, the plot will be close to linear. In other case, it will deviate from that form. Figures 13 and 14 (for Cnsmr and Other indices) and for the other indexes in appendix A, show the departure from normality of the residuals in all five indices, particularly in the extremes.

Figure 13: QQ plot of Normal distribution vs. the residuals of the ‘Cnsmr’ index.
This suggests that an assumption of conditional normality is unrealistic, and that the innovation process has fat tails or is leptokurtic. We further explore this hypothesis with a formal test called the Jarque-Bera Normality Test.

The Jarque-Bera test the null hypothesis that the sample in the variable under analysis comes from a normal distribution with unknown mean and variance, against the alternative that it does not come from a normal distribution. It’s a two-sided goodness-of-fit test suitable when a fully-specified null distribution is unknown and its parameters must be estimated. The test statistic is

\[ JB = \frac{n}{6} \left( s^2 + \frac{(k - 3)^2}{4} \right) \]

where \( n \) is the sample size, \( s \) is the sample skewness, and \( k \) is the sample kurtosis. For large sample sizes, the test statistic has a chi-square distribution with two degrees of freedom.

We perform the Jarque-Bera test for the returns of each index and also over the standardized residuals of the AR/MA-GARCH models. As we can see from table 8, the null hypothesis of normality is rejected in all cases at 99% for both the returns and the residuals. This indicates
that even after filtering out the returns they are still far from being ‘normal’. This completes our motivation to use EVT for the modeling of the tails of their distribution.

<table>
<thead>
<tr>
<th>J-B Test</th>
<th>Cnsmr</th>
<th>Manuf</th>
<th>HiTec</th>
<th>Hlth</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>h-residuals</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>p</td>
<td>0,001</td>
<td>0,001</td>
<td>0,001</td>
<td>0,001</td>
<td>0,001</td>
</tr>
<tr>
<td>p-residuals</td>
<td>0,001</td>
<td>0,001</td>
<td>0,001</td>
<td>0,001</td>
<td>0,001</td>
</tr>
<tr>
<td>jbstat</td>
<td>121416,012</td>
<td>275558,236</td>
<td>77267,7346</td>
<td>64700,251</td>
<td>66931,3398</td>
</tr>
<tr>
<td>jbstat-residuals</td>
<td>2541,98262</td>
<td>5006,13106</td>
<td>3112,56971</td>
<td>2155,3668</td>
<td>3450,41484</td>
</tr>
</tbody>
</table>

Table 8: Jarque-Bera test for the returns and residuals

3.4.1 The Peaks over a Threshold method

The EVT method we will implement in this thesis to model the tails of the distributions is called Peaks over a Threshold (POT). According to McNeil (1996) the main result in EVT the Pickands-Balkema-de Haan theorem (Balkema & de Haan 1974, Pickands 1975) essentially says that for a vast majority of the distributions employed in statistics and natural sciences (e.g. normal, lognormal, chi-squared, F, t, gamma, exponential, uniform, beta, etc.), losses which exceed high enough thresholds follow the generalized Pareto distribution (GPD). The GDP can be successfully adjusted for high enough thresholds. Additionally the method requires enough data over the threshold to estimate the parameters of the GDP. According to McNeil and Saladin (1997) if these two conditions are satisfied, the POT method allows us to use the GDP as an important tool for extreme events modeling.

3.4.1.1 Theoretical Background

In this section we will show some of the theoretical background of the POT method. Most of the content is taken from McNeil and Saladin (1997) and McNeil (1999). We refer the reader for a more general background on EVT to Embrechts, Klüppelberg & Mikosch (1997).
Now we suppose that each series of residuals of each index are a sequence of i.i.d. observations \( X_1 \ldots X_n \), from an unknown distribution function \( F \). We are interested in excess losses over a high threshold \( u \). Let \( x_0 \) be the finite or infinite right endpoint of the distribution \( F \). The distribution function of the excesses over a threshold \( u \) is given by

\[
F_u(x) = P\{X - u \leq x \mid X > u\} = \frac{F(x + u) - F(u)}{1 - F(u)},
\]

for \( 0 \leq x < x_0 - u \). Hence, \( F_u(x) \) is the probability that a loss exceeds the threshold \( u \) by an amount no more than \( x \), given that the threshold is exceeded.

Given this last equation and noting that \( X = x + u, y X > u \) we have that

\[
F(X) = [1 - F(u)]F_u(x) + F(u),
\]

this is valid because \( X > u \).

The distribution which comes to the fore in the modeling of excesses is the generalized Pareto distribution (GPD) which is usually expressed as a two parameter distribution with d.f.

\[
G_{\xi, \beta}(x) = \begin{cases} 
1 - (1 + \frac{\xi x}{\beta})^{-\frac{1}{\xi}} & \text{if } \xi \neq 0, \\
1 - \exp(-\frac{x}{\beta}) & \text{if } \xi = 0,
\end{cases}
\]

where \( \beta > 0 \), and the support is \( x \geq 0 \) when \( \xi \geq 0 \) and \( 0 \leq x < -\beta/\xi \) when \( \xi < 0 \). The GDP has implicitly three other distribution under its parametrization. When \( \xi > 0 \) we have a reparametrized version of the usual Pareto distribution; if \( \xi < 0 \) we have a type II Pareto distribution; \( \xi = 0 \) gives the exponential distribution.

This distribitional choice is motivated by a theorem (Balkema & de Haan 1974, Pickands 1975) which states that, for a certain class of distributions, the GDP is the limiting distribution for the distribution of the excesses as the threshold tends to the right point. Formally, we can find a positive function \( \beta(u) \) such that
\[
\lim_{u \to +\infty} \sup_{0 \leq x \leq u} \left| F(x) - G_{\xi, \beta(u)} \right| = 0,
\]

if and only if \( F \) is in the maximum domain of attraction (MDA) of the extreme value distribution \( H_{\xi} \), a condition we write as \( F \in MDA(H_{\xi}) \). This condition is fulfilled by all the common continuous distributions of statistics and actuarial science (normal, lognormal, \( \chi^2 \), t-student, F, gamma, exponential, uniform, beta, etc.). But what does this condition mean?

The generalized extreme value distribution (GEV) has the distribution function

\[
H_{\xi}(x) = \begin{cases} 
\exp(-(1 + \xi x)^{-1/\xi}) & \text{if } \xi \neq 0, \\
\exp(-e^{-x}) & \text{if } \xi = 0,
\end{cases}
\]

where \( x \) is such that \( 1 + \xi x > 0 \) and \( \xi \) is known as the shape parameter. The three extreme value distributions are special cases of the GEV: if \( \xi > 0 \) we have the Fréchet distribution; if \( \xi < 0 \) we have the Weibull distribution; \( \xi = 0 \) gives the Gumbel distribution.

The GEV is the natural limit distribution for normalized maxima. Let\( s\) consider the maximum of a block of \( n \) observations to be \( M_n = \max(X_1,\ldots,X_n) \). Suppose it is possible to find sequences of numbers \( a_n > 0 \) and \( b_n \) such that the distribution of \( (M_n - b_n)/a_n \) converges to some limiting distribution \( H \) as the block size increases. If this occurs \( F \) is said to be in the maximum domain of attraction of \( H \). This is fulfilled under the condition that the underlying distribution is “reasonable behaved” (non-degenerate). It turns out that the non-degenerate possibilities for this distribution \( H \) are limited. It was shown by Fisher & Tippett (1928) that the extreme value distributions (possibly with changes of scale and location) are the only non-degenerate limit distributions for appropriately normalized sample maxima.

If this condition is fulfilled, then in this case, by the theorem of Pickands, Balkema and de Haan, it can be inferred that the distribution function of the excesses above a high threshold
lean to a Generalized Pareto with shape parameter $\xi$, as the threshold tends to the right limit of $F$.

### 3.4.1.2 The Threshold choice and Parameter estimation

According to McNeil and Saladin (1997) these theoretical considerations suggest that when we have data from unknown underlying distributions we can successfully approximate the distribution of excess amounts over sufficiently high thresholds using a generalized Pareto distribution, $G_{\xi,\beta}(x)$ for some values of $\xi$ and $\beta$. This is a modeling approach which can be found in several papers in the statistical literature (Smith 1989, Davison & Smith 1990).

As mentioned before, within this method we need to find an appropriate high threshold above which the approximation of Pickands, Balkema and de Haan’s theorem is good and above which we still have enough data to find accurate estimates of the GDP parameters. One tool commonly use to choose suitable thresholds is the sample mean excess plot (ME)

$$\{(u,e_n(u)), X_{n,n} < u < X_{1:n}\},$$

where $X_{1:n}$ and $X_{n:n}$ are first and $nth$ order statistics of the data sample $e_n(u)$ is the sample mean excess function, is defined as

$$e_n(u) = \frac{\sum_{i=1}^{n} (X_i - u)^+}{\sum_{i=1}^{n} 1_{\{X_i > u\}}},$$

i.e. the sum of the excesses over the threshold $u$ divided by the number of data points over that threshold. The sample mean excess function is an empirical estimate of the mean excess function which is defined as $e(u) = E[X - u|X > u]$. According to McNeil and Saladin (1997) the mean excess function describes the expected overshoot of a threshold given that the exceedance occurs. In particular, if the empirical plot seems to follow a reasonably straight line with positive gradient above a certain value of $u$, then this is an indication that the
excesses over this threshold follow a GPD with positive shape parameter. This is clear since for the GPD

\[ e(u) = (\sigma + \xi \lambda)/(1 - \xi), \]

where \( \sigma + \lambda \xi > 0 \).

Examples of the mean excess functions for the negative returns (left tail) of the Manuf and Hlth indices are shown in figures 15 and 16 and for the positive returns of the Other index in figure 17 (in Appendix A are the ME plots for the negative and positive returns for the rest of the indices). We can see from the graphs that the ME of the loses (left tail) after a certain
threshold that the function looks indeed as a straight line with positive slope. However, for the positive returns (right tail) this situation doesn’t seem to hold.

Figure 17: Mean excess function for the losses of the ‘Hlth’ index.

For reasons that will be explained in section 4.4.3 of Backtesting, and to be consistent trough the whole process we will choose the threshold as the 95\textsuperscript{th} quantiles in the data (for both, looses and gains). These values are around the threshold suggested by the ME graphs. The thresholds equivalent to the 95\textsuperscript{th} and 5\% quantiles for the right and left tail correspondingly, are shown in table 9.

<table>
<thead>
<tr>
<th>Thresholds</th>
<th>Csmr</th>
<th>Manuf</th>
<th>HiTec</th>
<th>Hlth</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Tail</td>
<td>-1.6456</td>
<td>-1.6592</td>
<td>-1.62</td>
<td>-1.6611</td>
<td>-1.6755</td>
</tr>
<tr>
<td>Right Tail</td>
<td>1.5717</td>
<td>1.5619</td>
<td>1.6122</td>
<td>1.5746</td>
<td>1.5459</td>
</tr>
</tbody>
</table>

Table 9: Threshold of left and right tails using the 95\% and 5\% criterion

A similar Threshold choice procedure for Backtesting purposes is done in McNeil (1999).

We estimate the GPD parameters using the Maximum likelihood method, explained in Embrechts, Kl"uppelberg, & Mikosch (1997); the results are in table 10. The values of the shape parameter confirm the hypothesis that came from the ME plots: the looses seem to follow a GPD with positive shape parameter $\xi$, which means that the underling distribution
should be a heavy tailed one; the positive returns do not present this situation very clear, meaning that their underlying distribution might have a lighter tail. This illustrated the asymmetry on the returns distribution.

<table>
<thead>
<tr>
<th></th>
<th>Csmr</th>
<th>Manuf</th>
<th>HiTec</th>
<th>Hlth</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Tail</td>
<td>ξ</td>
<td>0.1178</td>
<td>0.1371</td>
<td>0.1474</td>
<td>0.1311</td>
</tr>
<tr>
<td></td>
<td>β</td>
<td>0.5153</td>
<td>0.5201</td>
<td>0.4494</td>
<td>0.519</td>
</tr>
<tr>
<td>Right Tail</td>
<td>ξ</td>
<td>-0.015</td>
<td>-0.058</td>
<td>0.0063</td>
<td>-0.0726</td>
</tr>
<tr>
<td></td>
<td>β</td>
<td>0.5201</td>
<td>0.5151</td>
<td>0.5357</td>
<td>0.5395</td>
</tr>
</tbody>
</table>

*Table 10: GPD parameters Maximum Likelihood estimates*

### 3.4.2 EVT VaR and ES

Now, we will use the GDP to estimate the VaR and ES of each asset. This method has the advantage to have a parametric calculation for them. This is of interest in particular to test the robustness of our risk-measure estimates. If we re-write equation (1) we may have for \( X > u \)

\[
1 - F(X) = [1 - F(u)](1 - F_u(x - u))
\]

According to McNeil and Frey (2000) if we estimate the first term on the right hand side of the equation using the random proportion of the data in the tail \( N/n \), and if we estimate the second term by approximating the excess distribution with a GDP fitted by maximum likelihood, we get the tail estimator.

\[
\hat{F}(X) = 1 - \frac{N}{n} \left(1 + \frac{\hat{\xi}}{\hat{\beta}_N} \frac{X - u}{\hat{\beta}_N}\right)^{-\frac{1}{\hat{\xi}_N}}
\]

As mentioned in the precedent section, in order to be able to Back-test our method, the amount of data considered in the tail will be fixed to be \( N = k \) where \( k << n \). This effectively gives us a random threshold at the \((k + 1)th\) order statistic. Given \( z_{(1)} \geq z_{(2)} \geq \cdots \geq z_{(n)} \) are the ordered residuals, we will fit the GDP with parameters \( \hat{\xi} \) and \( \hat{\beta} \) to the data.
\((z_{(1)} - z_{(k+1)}, \ldots, z_{(k)} - z_{(k+1)})\), the excess quantities for all residuals exceeding the threshold. The solution for the tail estimator of \(\hat{F}_z(z)\) is then

\[
\hat{F}_z(z) = 1 - \frac{k}{n} \left( 1 + \hat{\xi}_k \frac{z - z_{(k+1)}}{\hat{\beta}_k} \right)^{-1/\hat{\xi}_k}
\]

For \(q > 1 - k/n\) we can invert this expression of the tail to get the quintile formula

\[
\hat{z}_q = z_{(k+1)} + \frac{\hat{\beta}_k}{\hat{\xi}_k} \left( \frac{1-q}{k/n} \right)^{-1/\hat{\xi}_k} - 1
\]

Given the conditional quantile of the distribution (VaR) we can calculate the conditional ES. McNeil and Frey illustrates that as follows:

The conditional (one-step) Expected Shortfall is given by

\[
S_q = \mu_{r+1} + \sigma_{r+1} E[Z | Z > z_q].
\]

We need an estimate of the ES for the innovation distribution \(E[Z | Z > z_q]\) to estimate this risk measure. For a random variable \(W\) with an exact GPD with parameters \(\xi < 1\) and \(\beta\) it can be shown that

\[
E[W | W > w] = \frac{w + \beta}{1 - \xi},
\]

where \(\beta + w\xi > 0\). Supposing that the exceedances over a threshold \(u\) have exactly this distribution, by noting that \(z_q > u\) we can write

\[
Z - z_q | Z > z_q = (Z - u) - (z_q - u) | (Z - u) > (z_q - u),
\]

it can be shown that

\[
Z - z_q | Z > z_q \sim GPD(\xi, \beta + \xi(z_q - u)),
\]
So that the excesses over the higher threshold \( z_q \) also have a GPD with the same shape parameter \( \xi \) but with different scaling parameter. We can use equation (3) to get

\[
E[Z | Z > z_q] = z_q \left( \frac{1}{1-\xi} + \frac{\beta - \xi}{(1-\xi)z_q} \right).
\]

This is obtained using the estimator of the quintile showed in equation (2) and replacing \( \xi \) and \( \beta \) by their maximum likelihood estimators and \( u \) by \( z_{k+1} \). This results in the conditional estimator for the \( ES \)

\[
S'_{q} = \mu_{t+1} + \sigma_{t+1} \left( z_q \left( \frac{1}{1-\xi} + \frac{\beta - \xi}{(1-\xi)z_q} \right) \right).
\]

(4)

3.4.3 Backtesting of VaR

In order to make a robustness check of our estimate of the VaR, we will use a Backtesting method. This consists in performing a test of the method using historical data. The idea is to simulate the procedure we would do on a daily basis (in the future) but as if we have done it every day in the past. This implies taking the data available until one day in the past and then makes the forecast for the next day using equation (2). We repeat this procedure and look at the number of times that the actual returns violate our VaR estimate.

In order to be able to make any conclusion about the robustness of our risk-metric, we need to use a significant number of days to test the methodology. However, the procedure explained above presents the “complication” of the threshold choice, which requires certain analysis of the data each time the procedure is implemented. In the backtesting exercise, given the high number of times we need to repeat the procedure the analysis is not feasible. What we do instead is the procedure suggested in McNeil and Frey (2000), where we fix the number of data considered to be in the tail. We considered that 5% lowest residuals were on the tail of
the distribution. The corresponding threshold, given this assumption corresponds to values around -1.6 for every index. The choice of the threshold is not very far from the one we would choose based on the ME-criterion when looking at the whole sample. Furthermore, the threshold that we might have choose given a limited amount of data in the past may change and is more likely to be a lower value, since less extreme cases would be considered when we look at sub samples as opposed to the whole sample.

For comparison purposes we did the Backtesting exercise over the VaR estimate based on three methods: Our Conditional EVT methodology explained, a conditional mean and variance assuming normality of the residuals (same models for the conditional mean and variance) and an unconditional EVT model (not TSA modeling, but only EVT directly over the returns).

We estimated the 99% quantile during 2000 days using 5000 days of data to estimate the parameters of the models. A plot of the three VaR estimates corresponding to the ‘Cnsmr’ index can be seen in Figure 18 and in appendix A for the rest of the indexes. We also present in figure 19 plots of the two conditional VaR estimates (EVT and normal) together with the negative returns of the ‘Cnsmr’ index and the same graphs for the rest of the indices in appendix A. We can see from those figures the reaction of the Conditional VaR to the changes in volatility of the variables, particularly visible around the market crash of 1987. We should notice also that the unconditional VaR does not react considerably not even around that date.
We looked at the number of failures of the three VaR estimates and compare it to the expected number of failures. In this case this last quantity is simply the 1% of the number of days that the strategy was implemented. The results are presented in table 11.
Using the violation ratio of Danielsson we find that the Conditional EVT methodology is more robust than the other two methodologies for most of the indices (with the exception of the series ‘Other’ where the unconditional EVT outperforms and the C-EVT overestimates the VaR). The normal assumption is not robust in any case and it seriously underestimates the risk.

McNeil and Frey (2000) also test a conditional model with t distributed innovations. They find that the t distribution, on the other hand, underestimates the losses and overestimates the gains. This illustrates the drawbacks of using a symmetric distribution with data which are asymmetric in the tails. They point out that with more symmetric data the conditional t distribution often works quite well and that in fact, can be viewed as a special case of “our” method (their method which is the same we are using at this stage of the process). The reason for this is, as already mentioned, that the t distribution is an example of a heavy-tailed distribution, i.e. a distribution whose limiting excess distribution is a GPD with $\xi > 0$.

### 3.4.4 Fitting the semi-parametric distribution

Once we have confirmed the robustness of our TSA-EVT model to estimate extreme risks by using the VaR Backtesting methodology, we can proceed to use this method “looking forward”. We will fit the GP distribution for both, left and right tail separately using the thresholds mentioned in table of section 4.4.1.2 and from this model we will obtain the innovation process. This methodology allows us to take into account the fat-tails and the
asymmetry of the return’s distributions. Subsequently we will reintroduce the stochastic volatility feature with the GARCH forecasts; all this integrated in one model!

In figures 20 and 21 (and in appendix A for the rest of the indexes), we can now visually analyze the goodness-of-fit of the EVT model to the empirical data for both tails of the distribution. We also compare it to the fit of a normal and a t distribution.

![Figure 20: Left tail fit of using EVT, Normal and t distributions. ‘Other’ index.](image)

![Figure 21: Right tail fit of using EVT, Normal and t distributions. ‘Other’ index.](image)
As we can see from the zoom on figures 20 and 21, that the EVT model presents in general the best fit of the three methods, especially when we look into the higher quantiles. In particular we can see that the normal distribution and the student t, both underestimate de looses and overestimate the positive returns on the extreme cases. This result is in line with the finding of McNeil and Frey (2000).

We will need to estimate the complete marginal distributions (tails plus body) in order to proceed with the next stage of the general process: simulation. Since we want to be consistent in the model selection process, we will stay away from the pure normal distribution assumption even for the body of the distribution. In this sense we will use a non-parametric smoothed interior obtained with a Kernel approximation and the parametric GPD for the upper and lower tails to construct a composite semi-parametric CDF for each index.

The semi-parametric distribution allows for interpolation within the interior of the CDF and extrapolation (function evaluation) in each tail. Extrapolation is very desirable from a risk management perspective, since it allows estimation of quantiles outside the historical record.

The kernel-smoothing method computes a probability density estimate of the sample in the vector of residuals $X$ of size n. The Kernel estimators smooth out the contribution of each observed data point over a local neighborhood of that data point. The estimate is based on a normal kernel function, using a window parameter ('width') that is a function of the number of points in $X$. The density is evaluated at 100 equally spaced points that cover the range of the data in $X$. The Kernel density estimator is given by

$$
\hat{f}(x, h) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h} \right)
$$
with kernel \( K = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} u^2\right) \) and bandwidth \( h \). Under mild conditions (\( h \) must decrease with increasing \( n \)). The kernel estimate converges in probability to the true density.\(^{10}\)

Figure 22 illustrates the fit of the semi-parametric distribution fitted to the empirical cdf. In red and blue are drawn the left and right tails correspondingly, fitted with the GPD and in black the kernel-smooth distribution. As we can see they are almost indistinguishable from the empirical distribution, which means the fit of the distribution is good.

Having the marginal distributions fitted for each of the assets on the portfolio, there is only one step more before going to the simulation stage. This is the dependency among them and its modeling.

### 3.5 Dependency

As we pointed out before, a great deal of the “raison d’être” of portfolio theory relies on the dependence assessment between risky asset returns and therefore risk management in general.

\(^{10}\) See help file of Matlab for ‘kernel’
This is the reason that motivated us to look into mode complete dependence models beyond linear correlation, which only works in a normal word with linear dependence. This two last assumptions are demonstrated to be unrealistic in most financial applications.

3.5.1 Asymptotic Dependence

According to Poon, Rockinger and Tawn (2003) to understand extremal dependence, one must first appreciate that the form and degree of such dependence determine the chance of obtaining large values of both variables. For instance if we consider two variables S and T with a common scale, events of the form \( \{ S > s \} \) and \( \{ T > s \} \), for large values of \( s \), represent extreme events for each variable. As all such probabilities will tend to zero as \( s \to \infty \) it is natural to consider conditional probabilities of one variable given that the other is extreme. Specifically, consider \( \Pr(T > s|S > s) \) for a large \( s \). If (S, T) are perfectly dependent then

\[
\Pr(T > s|S > s) = 1.
\]

In contrast, if (S, T) are exactly independent then

\[
\Pr(T > s|S > s) = \Pr(T > s),
\]

which tends to 0 as \( s \to \infty \).

They give two measures defining asymptotic dependence, the first one is:

\[
\chi = \lim_{s \to \infty} \Pr(T > s|S > s), \quad 0 \leq \chi \leq 1
\]

we say that variables are asymptotically dependent if \( \chi > 0 \) and asymptotically independent if \( \chi = 0 \). Clearly \( \chi \) measures the degree of dependence that persists in the limit.

In general, when \( \chi = 0 \), the two random variables are not necessarily exactly independent. For example, if the dependence structure is that of a bivariate normal random variable with any value for the correlation coefficient less than one, then \( \chi = 0 \).
Poon et al argue that when exact independence is rejected, “traditional” multivariate extreme value methods assume \( \Pr(T > s | S > s) = \chi > 0 \) for all large \( s \). Furthermore, if the true distribution of the variables is asymptotically independent, the use of the traditional multivariate extreme value methods will over-estimate \( \Pr(S > s, T > s) \) and all other probabilities of joint extreme events since \( \Pr(T > s | S > s) \to 0 \) as \( s \to \infty \). The degree of bias will be depend on the difference between \( \chi \) and \( \Pr(T > s | S > s) \), which is determined by the value of \( s \) and the rate at which \( \Pr(T > s | S > s) \to 0 \) as \( s \to \infty \).

The authors present an alternative measure of dependence when \( \chi = 0 \). Although the random variables are asymptotically independent in that case, different degrees of dependence are reached for finite level of \( s \). The measure was suggested first in Coles, Heffernan and Tawn (1999):

\[
\bar{\chi} = \lim_{s \to \infty} \frac{2 \log \Pr(S > s)}{\log \Pr(S > s, T > s)} - 1, 
\]

\(-1 \leq \bar{\chi} \leq 1\), is an appropriate measure of asymptotic independence as it gives the rate that \( \Pr(T > s | S > s) \to 0 \). Values of \( \bar{\chi} > 0 \), \( \bar{\chi} = 0 \) and \( \bar{\chi} < 0 \) “loosely” mean positive association in the extremes, exactly independence and negative association between \( S \) and \( T \). For the bivariate normal dependence structure \( \bar{\chi} \) is the correlation coefficient.

More importantly Poon et al say that the couple \((\chi, \bar{\chi})\) together provide all the necessary information to characterize the form and degree of extremal dependence. For asymptotically dependent variables \( \bar{\chi} = 1 \) with the degree of dependence given by \( \chi > 0 \). For asymptotically independent variables \( \chi = 0 \) with the degree of dependence given by \( \bar{\chi} \). They remark that it
is important to test if $\chi = 1$ first before drawing conclusions about asymptotic dependence based on estimates of $\chi$.

### 3.5.1.1 Asymptotic Dependence Estimation

Poon, Rockinger and Tawn (2003) list the weak assumptions required for estimating $\chi$ and $\chi$. The first three of them are essentially the same as those required when estimating the univariate EVT model for the tails: (i) the joint distribution $(S, T)$ has a joint tail behaviour that is bivariate regularly varying, satisfying the conditions of Ledford and Tawn (1998)–counter examples to this form are given by Schlather (2001); (ii) both $\chi$ and $\chi$ are limit properties, so it is necessary to assume that the sample characteristics of the empirical joint distribution, above some selected threshold, reflect the limiting behaviour; (iii) the series has sufficient independence over time for the sample characteristics to converge to the population characteristics of $\chi$ and $\chi$; (iv) the marginal variables can be transformed to identically distributed Fréchet variables.

To present the extreme dependence measures estimation, Poon et al. introduce a special example of threshold modelling linked to the generalized Pareto distribution for the case where $\xi > 0$, i.e., Fréchet tail. In this case the tail of the variable $Z$ above a high threshold $u$ can be approximated as

$$
\Pr(Z > z) \sim \frac{L(z)}{z^{\frac{1}{\xi}}} \text{ for } z > u,
$$

Where $L(z)$ is a slowly varying function of $z$. Treating the slowly varying function as a constant for all $z > u$, say $L(z) = c$, under the assumption of independent observations the maximum likelihood estimators for $\xi$, known as the Hill’s estimator and $c$ are

$$
\hat{\xi} = \frac{1}{n_u} \sum_{j=1}^{n_u} \log \left( \frac{z_{(j)}}{u} \right),
$$
\[
\hat{\xi} = \frac{n_u}{n} u^{\hat{\xi}},
\]
(5)
where \(z_{(1)}, \ldots, z_{(n_u)}\) are the \(n_u\) observations of the variable \(Z\) that exceeded \(u\).

Poon, Rockinger and Tawn (2003) use the results in Ledford and Tawn (1996, 1997, 1998) to estimate \(\xi\) and \(\chi\) as

\[
\hat{\chi} = \frac{2}{n_u} \left( \sum_{j=1}^{n_u} \log \left( \frac{z_{(j)}}{u} \right) \right) - 1,
\]
\[
Var(\hat{\chi}) = \frac{(\hat{\chi} + 1)^2}{n_u}
\]
The asymptotically normality of this estimator is ensured by results in Smith (1987). If \(\hat{\chi}\) is significantly less than 1 (i.e., if \(\hat{\chi} + 1.96 \sqrt{Var(\hat{\chi})} < 1\)) then we infer the variables are asymptotically independent and take \(\chi = 0\). Only if there is no significant evidence to reject \(\chi = 1\) we do estimate \(\chi\) (since we do it under the assumption that \(\chi = 1\)). Using the maximum likelihood estimator by (5) and under the constrain \(\chi = 1\), the estimator for \(\chi\) is

\[
\hat{\chi} = \frac{u n_u}{n},
\]
\[
Var(\hat{\chi}) = \frac{u^2 n_u (n - n_u)}{n^3}.
\]
Following the procedure proposed by Poon et. al. to characterize the extreme dependence between two variables, we estimated \(\chi\) for every pair among the five industry indices, then we tested for the hypothesis that \(\chi = 1\) and then calculated \(\chi\). The results are in table 12. They show that four out of ten pairs are asymptotically dependent. The results for \(\chi\) where the \(\chi = 1\) hypothesis is rejected are shown in red, indicating they should not be interpreted since they where calculated under the assumption that the hypothesis was not rejected.
Although some of the pairs present asymptotic dependence, this condition is not fulfilled by most of the pairs (more than half), confirming the results of Poon, Rockinger and Tawn: not all financial series present asymptotic dependence, and hence we can not take this condition as given, by we should test it first and take into account the class of extreme dependence we are looking at.
Both forms of extremal dependence (asymptotic dependence and asymptotic independence) permit dependence between rather large values of each variable. However, only when the variables exhibit asymptotic dependence the very largest values from each variable can occur together. To illustrate this, Figure 23 presents scatter-plots of daily return of the pairs HiTec vs. Cnmsr and Manuf vs. Csmr. We also show in figure 24 the graphs presented on Poon, Rockinger and Tawn (2003) of stock market returns pairs (US vs. Japan and Germany vs. France equity indices). The graph on the left, for both figures, show an asymptotically dependent pair and on the right graph an asymptotically independent pair (however not exactly independent).

Poon, Rockinger and Tawn (2003) also document the effect of heteroskedasticity on the tail index and tail dependence among international stock market returns. They find that tail index and tail dependence can be partially captured by models for heteroskedasticity. In particular their results show that all $\tau$ estimates are greater than zero but the tail dependence of heteroskedasticity filtered returns is substantially reduced.

In this sense, we also estimated $\tau$ of the return after filtering out the heteroskedasticity effect (what we called residuals in the time series section). The results are presented in table 13.
After testing the hypothesis of asymptotic dependence $\mathcal{Z} = 1$, we found that all filtered returns are asymptotically independent. This is also the finding of Poon et. al with respect to the filtered returns for all the financial paired series they tested.

The results suggest that there is significant dependence between large values of the paired series but that the very largest values do not occur concurrently. According to Poon, Rockinger and Tawn (2003), when most of the pairs are asymptotically independent, multivariate extreme value models that assume asymptotic dependency among stock returns are likely to have overestimated portfolio joint risk.

In this sense we will use a method to model dependencies among assets that allow them to have extreme dependence, but not assuming asymptotic dependence necessarily. As we will see, a particular method that fulfills these conditions is one based on the t-copula.

### 3.5.2 Dependence with Copulas

As seen on the last section, it is very common among multivariate studies, to remove the influence of marginal aspects first by transforming the original variables to a common marginal distribution. After such a transformation, differences in distributions are solely due to dependence features. This idea of separating the multivariate distribution into dependence structure and marginal behavior, leads to the well known concept of a copula. Hence, in this approach the dependence measure are no longer influenced by the form of the marginal distributions.
According to Schmidt (2003) Copulas are of interest in finance because of two reasons: First, as a way of studying the dependence structure of an asset portfolio irrespective of its marginal asset-return distributions; and second, as a starting point for constructing multidimensional distributions for asset portfolios, in a simulation context. The latter focus is the one we take in this thesis, since our aim is to construct a portfolio (not analyze a given one).

Copulas can bee seen either as functions that join or ”couple” multivariate distributions to their corresponding marginal distributions or as a multivariate distribution function with uniform margins on the interval $[0, 1]$. We take now the latter definition

Copula Definition: Let $C : [0,1]^n \rightarrow [0,1]$ be an $n$-dimensional distribution function on $[0,1]^n$. Then $C$ is called a copula if it has uniformly distributed one-dimensional margins on the interval $[0,1]$.

The following theorem is the most important result behind the application of Copulas since it characterize is as a dependence structure of a multivariate distribution

Sklar’s Theorem: let $F$ be an $n$-dimensional distribution function with margins $F_1 \ldots F_n$. Then there exists a copula $C$, such that for all $x \in \mathbb{R}^n$

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)).$$  \hfill (6)

If $F_1 \ldots F_n$ are all continuous, then $C$ is unique; otherwise $C$ is uniquely determined on $\text{Ran} F_1 \times \cdots \times \text{Ran} F_n$, where $\text{Ran}$ stands for the corresponding range. Conversely, if $C$ is a copula and $F_1 \ldots F_n$ are one-dimensional distribution functions, then the function $F$ given (6)
is an n-dimensional distribution function with margins $F_1 \ldots F_n$. For a proof we refer the reader to Sklar (1996) and Nelsen (1999).

The following Corollary is derived from Sklar’s theorem and it shows the way to construct a copula of a multivariate distribution.

Corollary: Let $F$ be an n-dimensional continuous distribution function with margins $F_1 \ldots F_n$. Then the corresponding copula $C$ has representation

$$C(u_1, \ldots, u_n) = F(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)), \quad 0 \leq u_1, \ldots, u_n \leq 1,$$

(7)

Where $F_1^{-1}, \ldots, F_n^{-1}$ denote the generalized inverse (distribution) functions of $F_1 \ldots F_n$.

One of the most cited properties of Copulas is that the invariance under increasing transformations (see Yan 2006) that emphasizes the fact that the copula captures those properties of the joint distribution which are independent from the marginal. In other words it represents the dependence structure of a multivariate random vector (Schmidt, 2003).

For more details regarding the theory and financial applications of copula we refer the reader to Cherubini, Luciano and Vecchiato (2004).

### 3.5.2.1 T-student Copula

The t-copula comes from of a more general class of functions called the elliptical copulas often used in finance applications given its desirable tractability properties. They allow to model multivariate extreme events and forms of non-linear dependencies. Elliptical copulas are simply the copulas of elliptical distributions.
One copula function often chosen based on the tractability criteria is the Gaussian Copula. However, according to Martellini and Meyfredi (2003) one problem with the choice of a Gaussian copula function is that it does not allow extreme co-movements to be taken into account since it does not exhibit tail dependence.

According to Rank (2007) the Gaussian copula is asymptotically independent in both upper and lower tails; which implies no tail dependence in general. In our case, as we saw from the results of section 3.5.1, extreme dependence is present among equity indices (not always) and therefore it might be preferable to choose a function capable to generate tail dependence.

An alternative to the Gaussian copula, which belongs to the elliptical class of dependence functions, is the Student t copula. This copula has an “advantage” over the Gaussian one because it presents tail dependence, which is more appropriate in case we observe joint extremes in the data. We say this is an advantage, because in case we are in a situation where the extreme dependence is not present, the coefficient of tail dependence, in a t copula, tends to zero as the number of degrees of freedom tends to infinity for (for the non-perfect correlation case).

Having said this, we can argue then that the t copula has desirable properties as a dependence model since: (1) it has a desirable tractability property inherited from the elliptical distribution class, and (2) it is flexible enough to allow for tail dependence. The latter property implies that using empirical data to adjust the parameters of the t copula, instead of imposing a constraint over the tail dependence (as the Gaussian copula does) we can instead “let the data talk” and adjust the model in this sense trough the degree of freedom. Given this reasons, we decided to employ a t-copula to model the dependence between the assets in the portfolio.
The copula that corresponds to the multivariate t-Student distribution is the t copula, and Embrechts, Lindskog and McNeil (2001) present it as follows:

Given a random n-variate vector \( X \) with a t-Student distribution with \( \nu \) degrees of freedom, mean vector \( \mu \) (for \( \nu > 1 \)) and variance-covariance matrix \( \frac{\nu}{\nu-2} \Sigma \) (for \( \nu > 2 \))\(^{11}\). Then \( X \) can be represented as

\[
X^d = \mu + \frac{\sqrt{\nu}}{\sqrt{S}} Z, \quad (8)
\]

where \( \mu \in \mathbb{R}^n, S \sim \chi^2_\nu \) and \( Z \sim N_n(0, \Sigma) \) are independent, then \( X \) has an n-variate \( \nu \)-distribution with mean \( \mu \) and covariance matrix \( \frac{\nu}{\nu-2} \Sigma \). If \( \nu \leq 2 \) then \( \text{Cov}(X) \) is not defined.

In this case we just interpret \( \Sigma \) as the shape parameter of the distribution of \( X \).

The copula of \( X \) is the t-Student copula with \( \nu \) degrees of freedom. From equation (7) this is:

\[
C_{\nu,R}^t(u) = t_{\nu,R}^t(t^{-1}_\nu(u_1),\ldots,t^{-1}_\nu(u_n)),
\]

where \( R_{ij} = \Sigma_{ij} / \sqrt{\Sigma_{ii} \Sigma_{jj}} \) for \( i, j \in \{1, \ldots, n\} \) and where \( t_{\nu,R}^n \) denotes the distribution function of \( \sqrt{\nu} Y / \sqrt{S} \) and \( S \sim \chi^2_\nu \) and \( Y \sim N_n(0, R) \) are independent. The \( t_\nu \) in this case represent the equal marginal distributions of \( t_{\nu,R}^n \) and \( R \) is the linear correlation matrix of the corresponding n-variate \( t_\nu \)-distribution if \( \nu > 2 \).

In order to simulate dependent random variables using a t-copula, we need to estimate the parameters of the copula. We use historical data and maximum likelihood method to do that (explained in the next section). Once we have the parameters of the copula, we first simulate dependent t-student random variables and transform them into uniform variables using the cumulative distribution function (CDF) of the t-student. Then we transform the uniform

\(^{11}\) If \( \nu \leq 2 \) the covariance matrix is not defined
dependent variables to standardized residuals using the inverse CDF of the semi-parametric marginal distributions estimated before for each of the 5 indexes. The standardized residuals are later on used as the innovation process for the AR/MA-GARCH process that will produce the final simulated returns of the five indexes.

Embrechts, Lindskog and McNeil (2001) propose the following algorithm to simulate random variables from the t-copula $C_{v,R}^t$:

- Find the Cholesky decomposition $A$ of $R$.
- Simulate $n$ independent random variables $z_1, \ldots, z_n$ from $N(0,1)$.
- Simulate a random variable $s$ from a $\chi^2$ d.f. independent from $z_1, \ldots, z_n$.
- Set $y = Az$.
- Set $x = \sqrt{v/y}$.
- Set $u_i = t_s(x_i), i = 1, \ldots, n$.
- $(u_1, \ldots, u_n) \sim C_{v,R}^t$.

3.5.3 Parameter Estimation

The parameter estimation method we used is called the Canonical Maximum Likelihood method (CML). This method is desirable because it does not make assumptions about the parametric form of the marginal distributions. The estimation process in general is performed in two steps:

1. Transforming the dataset $(x'_1, \ldots, x'_n), t = 1, \ldots, T$, into uniform variables $(a'_1, \ldots, a'_n)$, using the empirical marginal distributions.
2. Estimating the copula parameters maximizing the log-likelihood function with respect to the set of parameters $\alpha$ i.e. $\alpha = \arg \max \sum \ln c(a'_1, \ldots, a'_n; \alpha)$.
However, the maximum likelihood function for the t student copula with parameters R and ν tends to have convergence difficulties in high dimensions. Given this difficulty, we used the estimation function for the t student copula parameters included in Matlab (version 2007) GARCH-demos called ‘tCopulaFit’. Detailed information about this demonstration is on the Mathworks web-page\textsuperscript{12}. This function performs the maximum likelihood estimation in a two step nested process – in programming terms could be understood as an inner loop within an outer loop or process. The maximum likelihood estimation (MLE) is accomplished in a two-step process. The inner step maximizes the log-likelihood with respect to the linear correlation matrix, given a fixed value for the degrees of freedom. Then, that conditional maximization is placed within a 1 to d\textsuperscript{13} maximization with respect to the degrees of freedom, thus maximizing the log-likelihood over all parameters. The function being maximized in this outer step is known as the profile log-likelihood for the degrees of freedom.

For a complete reference on copula parameters estimation and selection we refer the reader to Durrleman, Nikeghbali and Roncalli (2000). The results of the parameter estimation procedure explained above are in table 14.

<table>
<thead>
<tr>
<th>R</th>
<th>1.0000</th>
<th>0.8444</th>
<th>0.7902</th>
<th>0.7841</th>
<th>0.8679</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8444</td>
<td>1.0000</td>
<td>0.7708</td>
<td>0.7477</td>
<td>0.8538</td>
<td></td>
</tr>
<tr>
<td>0.7902</td>
<td>0.7708</td>
<td>1.0000</td>
<td>0.7085</td>
<td>0.7911</td>
<td></td>
</tr>
<tr>
<td>0.7841</td>
<td>0.7477</td>
<td>0.7065</td>
<td>1.0000</td>
<td>0.7647</td>
<td></td>
</tr>
<tr>
<td>0.8679</td>
<td>0.8538</td>
<td>0.7911</td>
<td>0.7647</td>
<td>1.0000</td>
<td></td>
</tr>
</tbody>
</table>

| DoF | 5.7235 |

\textit{Table 14: Copula Maximum Likelihood parameters estimates}

\textsuperscript{12} \url{http://www.mathworks.com/products/demos/shipping/garch/garchcopulaevtdemo.html?product=GA}

\textsuperscript{13} d being the number of assets; in this case equity indices.
In order to see the dependence structure of the data we looked at the empirical copulas of all residual’s pairs. In figures 25, and 26 below and on Appendix A for the rest of the pairs.

![Figure 25: Empirical Copula for Hlth and Hitec indices.](image1)

![Figure 26: Empirical Copula for Cnsmr and Other indices.](image2)

From these graphs we can see that joint extreme events are possible, and that a t copula might be necessary to generate this kind of dependence. However, strong tail dependence is not present in every pair.

### 3.5.4 Monte Carlo Simulation

Having fitted the dependence structure to the data we are ready jump to the next stage of the procedure: simulation. We now have a complete probabilistic model for the returns of all
assets (or risk factors) that will compose our portfolio. This allows us to simulate many scenarios in order to assess the risk that we might take when choosing a given set of weights. As explained before the integration of the models on the simulation framework is as follows:

1. Use the t copula and the semi-parametric marginal distributions to simulate dependent (among them, not in time) series.

2. This series are used as the innovation process for the AR/MA-GARCH models which reincorporate the serial correlation and heteroskedasticity to the series, generating the simulated returns.

This simple procedure is taking into account for the stylized facts of financial returns and incorporating an appropriate dependence framework to it. We simulated 10000 paths for a period of 22 days (number of trading days in one month) for each index. An illustration of the dependence between the innovations amongst three indices is showed in figure 27.

![Figure 27: Dependence of simulated residuals corresponding to the Cnsmr, Manuf and HiTec indices.](image)
In a real practitioner exercise, we would re-estimate the parameters of the models every certain period in order to include the new information available at that time into the model’s parameters. This might be important in particular for the GARCH models, since we want to account for changes in the volatility level when they occur; as we could see from the Backtesting section, this should make a difference on the risk estimation.

This adjustment on the risk estimation may lead some times to re-adjustment on the portfolio choice. As we will explain in the next section, we use the one month estimation (simulation) using all the information available and then perform a portfolio optimization exercise using the output of the simulation model.

3.6 Portfolio Optimization

The portfolio choice process consists in balancing the amounts invested among the different assets of choice. If we look at these amounts as a proportion of the total capital available to invest, the choice of the portfolio can be seen as a vector of weights $w$ where each element of the vector $w_i$ correspond to the percentage of the capital invested in asset $i$. This implies that the sum of all weights in the portfolio is equal to the unity (or 100%). The returns of the portfolio will be the sum of the asset returns in the portfolio multiplied by its corresponding weights.

Markowitz’s optimal portfolio selection model minimizes the variance of the portfolio given a desired level of return (the dual problem would be to maximize the return given a level of volatility). We can obtain the vector of weights using that criterion by solving the following optimization program
\[
\begin{align*}
\min_w & \quad w^T \Sigma w \\
\text{s.t.} & \quad \sum_i w_i = 1 \quad (9) \\
& \quad E(R_p) = \bar{r} \quad (10)
\end{align*}
\]

where \( \Sigma \) is the variance-covariance matrix\(^{14} \) of all assets returns in the portfolio, \( E(R_p) \) is the expected return of the portfolio and \( \bar{r} \) is the desired total return. The set of solutions of this program for all possible values of \( \bar{r} \) is what Markowitz calls the efficient frontier. More precisely, the efficient frontier excludes the portfolios that are dominated in a risk adjusted basis by other solution-portfolios of the program).

We can often find additional constrains such as short-selling constrains, but we don’t consider them in this thesis. We refer the reader for a more complete explanation to Markowitz (1952) and Fabozzi, Kolm, Pachamanova, and Focardi (2007).

### 3.6.1 ES minimization

Di Clemente and Romano (2003) argue that the minimization of the portfolio’s variance produces misleading results in terms of capital allocation given the unrealistic Gaussian assumptions which is based on. They suggest that in order to minimize portfolio risk effectively and to obtain a correct capital allocation, to use the ES as a measure tail risk of portfolio return distribution. In this sense they apply the methodology to minimize ES proposed by Rockafellar and Uryasev (1999):

Suppose to have a portfolio with \( d \) assets. Let \( f(w, r) \) be the portfolio loss, function of both

\(^{14}\) Each off-diagonal element of a variance-covariance matrix is the covariance between asset \( i \) and asset \( j \), given by \( \rho_{ij} \sigma_i \sigma_j \) and the diagonal elements are \( \sigma_i^2 \), where \( \rho_{ij} \) is the linear correlation between asset \( i \) and \( j \), and \( \sigma_i \) is the standard deviation of asset \( i \).
the vector $w$ of the $d$ positions and the random vector $r \in \mathbb{R}^m$. For each $w$, the loss $f(w, r)$ is a univariate random variable, whose distribution depends from the one of $r$. Let $p(r)$ be the multivariate density function of the random vector $r$. Given this, it is not necessary to give an analytical form to $p(r)$, but it is sufficient to have a procedure for generating historical or Monte Carlo scenarios from $p(r)$. From last section, we now have the Monte Carlo simulation tool set up.

The probability that $f(w, r)$ does not exceed a certain threshold, $\alpha$, is defined as:

$$\psi(w, \alpha) = \int_{f(w, r) \leq \alpha} p(r)dr$$

For fixed $w$, $\psi(w, \alpha)$ (as a function of $\alpha$) is the cumulative distribution function of the loss associated to the portfolio composition $w$. Generally, $\psi(w, \alpha)$ is a non-decreasing function with respect to $\alpha$ and continuous. However, it is assumed that this function does not present jumps and it is continuous with respect to $\alpha$. Without this restriction, the $ES$ optimization technique results mathematically very difficult.

The VaR of the portfolio $w$ with probability level $\beta \in (0,1)$ over the specified holding period is $\alpha_{\beta}(w) = \min\{\alpha \in \mathbb{R} : \psi(w, r) \geq \beta\}$.

The $ES$ with respect to $w$, at the probability level $\beta$ over the holding same period is

$$\phi_{\beta}(w) = \frac{1}{1 - \beta} \int_{f(w, r) \leq \alpha_{\beta}(w)} f(w, r) \cdot p(r) \cdot d(r)$$

Equation (11) represents the conditional expected loss beyond VaR. Actually, the probability of the event $f(w, r) \geq \alpha_{\beta}$ is equal to $1 - \beta$. This equation can also be written as

$$\phi_{\beta}(w) = \alpha_{\beta}(w) + \frac{1}{1 - \beta} \int_{w \in \mathbb{R}^d} [f(w, r) - \alpha_{\beta}(w)]^+ d(r)$$

The optimization is based on the following linear approximation of equation (12)

$^{15}$ where $[a]^+ = a$ when $a > 0$ and $[a]^+ = 0$ otherwise.
In equation (13) $\alpha$ is not the VaR but a simple numeric parameter.

Rockafellar and Uryasev (1999) demonstrated that function (13) is convex with respect to positions $w$ held in a portfolio and, hence, particularly useful to solve optimization problems. They also showed that the ES can be computed from equation (13) without calculating the VaR before. Since it is easy to optimize over differentiable and convex functions using numerical methods, the problems simplifies over the function on (13).

Given an $d$-variate distribution function of the asset’s returns $p(r)$ the portfolio’s random returns are given by the sum of each asset’s return $r_i$ weighted by its corresponding weight $w_i$. Hence, portfolios loss is defined as the negative return of the portfolio

$$f(w, r) = - [w_1 r_1 + \ldots + w_d r_d] = -w^T r$$

Then by the result of Rockafellar and Uryasev (1999) and replacing $-w^T r$ in equation (13), the objective function is to minimize

$$F_\beta(w, \alpha) = \alpha + \frac{1}{1-\beta} \int_{\mathbb{R}^d} \left[ w r - \alpha \right]^+ p(r) dr$$

(14).

Which is convex and often differentiable with respect to $w, \alpha$.

Now let $\mu(w) = w_1 r_1 + \ldots + w_d r_d = w^T r$ be the portfolio expected return, where $r = (r_1 \ldots r_d)^T$ is the vector of the expected returns of the $d$ assets (this might be obtained using historical data). In order to construct an efficient set of portfolios (or efficient frontier as Markowitz called it) we introduce an additional constrain; the portfolio expected return must be equal to a given objective return $\pi$:

$$\mu(w) = \pi$$

(15)
Now if we consider the set of feasible portfolios after taking into account the linear constrains given in (15) and the wealth constrain- given by equation 9 – we are looking at convex optimization problem.

Furthermore, equation (14) may be simplified by generating q historical or Monte Carlo scenarios, \( r_1 \ldots r_q \) of the random vector \( r \), sampled from the probability distribution \( p(r) \):

\[
\bar{F}_\beta(w, \alpha) = \alpha + \frac{1}{q(1-\beta)} \sum_{i=1}^{q} \left[ -wr_k - \alpha \right]^+
\]  \hspace{1cm} (16)

Minimizing Eq. (16) with respect to \((w, \alpha)\) with constraints (9) and (15) corresponds to the minimization of the portfolio’s ES. After applying this procedure we obtain the optimal portfolio, given by the decisional vector \( w \) with refer to a chosen portfolio’s expected return, \( \pi \). We can also calculate the VaR of the optimal portfolio.

We repeat this same process for different values of the objective return \( \pi \), to obtain the “efficient” set of portfolios.

### 3.6.2 Optimization Results

Using the 10,000 simulated paths, we implement the portfolio optimization procedure explained above, where we minimize the ES given a set of different objective returns \( \pi \). The set of portfolios corresponding to each \( \pi \) compose the efficient frontier (or efficient set of portfolios). For the sake of comparison, we also performed an optimization minimizing the portfolios volatility, using the optimization program expressed in equations (9) and (10).

Table 15 shows the set of weights assigned to each industry portfolio corresponding to each \( \pi \) for the ES and Vol minimization. The bold numbers are the corresponding to the minimum risk portfolios; where risk refers to ES and volatility respectively.
We can see that the allocation to the different industry indices using the average and extreme risk minimization leads to different allocations. Figures 28 show the efficient set of portfolios using that minimize ES. Figure 29 plots the expected return vs. the VaR of the same portfolios; when we look at the two figures together we can see that the portfolios with lowest

![Efficient frontier: ES vs. Expected Returns of min-ES portfolios](image)

Table 15: Efficient frontiers using minimum ES and minimum Volatility

<table>
<thead>
<tr>
<th>E return</th>
<th>0.800%</th>
<th>0.700%</th>
<th>0.800%</th>
<th>0.900%</th>
<th>1.000%</th>
<th>1.100%</th>
<th>1.200%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weights</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum ES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES</td>
<td>-1.570%</td>
<td>-1.340%</td>
<td>-1.230%</td>
<td>-1.200%</td>
<td>-1.270%</td>
<td>-1.420%</td>
<td>-1.630%</td>
</tr>
<tr>
<td>Vol</td>
<td>1.036%</td>
<td>0.886%</td>
<td>0.815%</td>
<td>0.809%</td>
<td>0.861%</td>
<td>0.945%</td>
<td>1.069%</td>
</tr>
</tbody>
</table>

| Weights  |        |        |        |        |        |        |        |
| Minimum Vol |      |        |        |        |        |        |        |
| ES       | -1.530%| -1.350%| -1.240%| -1.210%| -1.280%| -1.430%| -1.660%|
| Vol      | 0.935% | 0.853% | 0.805% | 0.797% | 0.829% | 0.897% | 0.994% |

Table 15: Efficient frontiers using minimum ES and minimum Volatility
ES are also the ones with the lowest VaRs. This is an expected result, given the close relation between these two extreme risk metrics.

Figure 29: VaR vs. Expected Returns of min-ES portfolios

Figure 30: Efficient Frontier: Volatility vs. Expected Returns of min-Vol portfolios

Figure 30 illustrates the efficient frontier under the Markowitz Criteria. It can also be notice that the Vol criteria also seem to have a relation with ES and VaR. We can see that by looking at the level of expected return in both frontiers; the group of portfolios with the lowest ES and
VaR have around the same level of expected return as the group of portfolios with minimum variance.

### 3.6.3 Out-of-sample Performance

In order to assess the impact or added value (if any) that the proposed methodology would have in a real investment scenario, we decided to Backtest the model and estimate its out-of-sample risk and performance.

In this exercise we used a rolling window of 6500 daily data points to estimate the parameters of the models every month (22 days), then simulated 5000 paths for the next 22 days and performed the optimization procedure explained in section 4.6.2. We repeat the same procedure every month for ten years.

For comparison purposes, we performed a similar exercise based on the classic assumptions of Markowitz model. In this case we generated a Multivariate Normal distribution and estimated its parameters (Expected returns and Sample Covariance matrix) using the same data each time (the last 6500 days of data); then simulated 5000 paths and calculated the Global Minimum Variance (GMV) and the Maximum Sharpe Ratio Portfolios (assuming zero interest rate). The Backtest was performed during ten years, due to the time that each iteration required to be completed. The starting date of the Backtest is 08/05/1989 (data point number 6500) and the end date is 03/04/2000; the latter corresponds to 120 months after the starting date if we count each month as 22 days (the 22 days is our approximation of the average number of trading days in a month).
Table 16 shows the out-of-sample risk and performance measures for the three portfolios. As we can see the minimum ES presents not only the lowest ES (out-of-sample) but also the highest Sharpe ratio! This result confirms two facts: (1) The strong relationship between and return and (2) the sample average return as an estimate of the expected return (as proposed by classic MPT) is not a good estimate for the expected return. The former is particularly evidenced by the fact that the Sharpe ratio is not actually maximized out-of-sample (is the worse among the three portfolios). Furthermore the fact that the volatility is a slightly lower for the min-ES portfolio that the one of the GMV shows that the Normality assumption does not hold and it has an impact on the portfolio optimization risk.
4 Conclusions, limits and further research

Portfolio choice is one of the main subjects of study in Market Finance and as we argue here, it is not a trivial problem to solve and improvements from its classic approach should be done.

In a portfolio construction process the definition of the objectives is crucial and they are typically related to the risk and return that an investor wants to have in the portfolio she/he chooses. In addition, we argue that the accuracy in the measures of the risk and return is also a fundamental part of the process and hence needs special attention.

In particular, we believe that using inaccurate estimates for risk and return can only give us misleading information about them, and in consequence our choice would be based on weak grounds and end up in disappointing results. In this sense, we decided to not include estimates of the Expected Return into the analysis, because the assumptions and estimation of its classic measure are not sound enough. On the other hand, we think that some of the models and measures currently available to estimate the risks of some assets are more trustable and accurate enough to use them in the portfolio selection process.

Given the limitations on the accuracy of the estimation of future expected returns, we employ a “risk management” approach to the problem. However, our results suggest that the return is not taken out of the picture given its strong tie to risk.

Having said this, we motivate the main purpose of this thesis: to look at possible improvements in the portfolio construction process with respect to the classic approach and implement them to assess their gains. We approach the problem by using risk-measurement
methods with more sound assumptions, which makes them more reliable than their classic counterparts.

Moreover, we focus on extreme risk estimation and minimization; because we believe this type of risk (and its corresponding measures and assumptions) are better representations of investors’ concerns and reality than the classic risk-metric and model assumptions, such as standard deviation and the so-called normality assumption. We also focus on extreme risk rather than volatility because for banks this type of risk is of particular interest for regulatory purposes.

We asses the differences and gains of using the proposed risk models with respect to the classic approach in three ways: First we estimate the accuracy of the (extreme) risk measures using a Backtesting method. Secondly, we compare the allocation suggested by the two methods looking forward and finally we perform an out-of-sample exercise using historical data by means of a rolling window of information available at each point in time (just as if would have performed the portfolio construction process in the past).

Using some of the new technologies available to model assets’ returns, we integrate them into a probabilistic model that allows us to perform simulations of possible future scenarios. We used the output of these models in order to estimate extreme risk measures and then to construct portfolios minimizing this risk. These “new technologies” step away from unrealistic assumptions of classic portfolio theory and take into account the so-called stylized facts of financial time series. These methods are able to model returns’ characteristics such as fat-tails, asymmetry, stochastic volatility and extreme dependence.
Using a Backtesting method we find that taking into account for these stylized facts, improves the accuracy of the risk measures. In order to determine the robustness of the proposed estimation method, we compare it to other two methods: the first assumes conditional “normality” and the other one an unconditional version (e.g. no stochastic volatility) of the EVT method. We find that the improvement is significant in both cases for most of the series.

We also compare the allocation produced following the classic assumptions and minimizing volatility to the extreme risk minimization proposed. We find that the portfolio allocation can differ significantly; although we see that there is a relation between extreme risk and volatility.

On an out-of-sample exercise, using historical data, we also find that during a ten years period, the method proposed outperforms the classic approach in terms of risk and risk-adjusted performance measures, such as Expected Shortfall and Sharpe ratio, respectively.

In general, we can conclude that the assumptions underlying the models used for risk assessment and management do matter and can have an important impact on investment decisions and their corresponding real performance. Therefore, it is important for portfolio managers to make special attention on the choice of their risk models and their calibration, because their decisions may be highly influenced by this risk’s estimation.

Further robustness checks of the optimization process could be done such as: increasing the length of the testing window, varying the rebalancing period, modifying the amount of data to calibrate the models, changing the rolling window approach by a growing window one and
using other confidence levels for the extreme risk measures (alpha). These tests are not performed in the present document due to time constraints.

One interesting extension of the work done in this thesis would be to look at the allocation and portfolio performance using different asset classes and international markets, and to explore the marginal contribution of these new asset classes and markets in terms of diversification.

Another interesting extension of this work could be to include alternative measure of the Expected return’s explicitly into the portfolio construction process. Some of the options in this field are related to factor models and fundamental approaches. However, they may also have model risks and estimation challenges that should be assessed and taken into account at the model selection stage.
5 References


Cool T. (1999). Proper definitions for Risk and Uncertainty (July update with (a) better notation, (b) relative risk (c) an application to insurance). General Economics and Teaching 9902002, EconWPA.


6 Appendix A

For completeness purposes, in this appendix we present the different graphs that do not appear in the other sections of the thesis. There are no different types of graphs here but only the same kind of graphs that appear on the different sections of the thesis; although they correspond to the rest of the indexes.

Figure 31: ACF of returns, ‘Cnsmr’ index

Figure 32: ACF of returns, ‘Manuf’ index
Figure 33: ACF of returns, ‘HiTec’ index

Figure 34: ACF of returns, ‘Hlth’ index

Figure 35: PACF of returns, ‘Manuf’ index
Figure 39: ACF of squared returns, ‘Manuf’ index

Figure 40: ACF of squared returns, ‘Hlth’ index

Figure 41: ACF of squared returns, ‘Other’ index
Figure 42: residuals and standard deviation GARCH forecasts, ‘Cnsmr’ index

Figure 43: residuals and standard deviation GARCH forecasts, ‘Hlth’ index
Figure 44: residuals and standard deviation GARCH forecasts, ‘Other’ index

Figure 45: ACF of Standardized Residuals, ‘Cnsmr’ index
Figure 46: ACF of Squared Standardized Residuals, ‘Cnsmr’ index

Figure 47: ACF of Squared Standardized Residuals, ‘Manuf’ index

Figure 48: ACF of Standardized Residuals, ‘HiTec’ index
Figure 49: ACF of Standardized Residuals, ‘Hlth’ index

Figure 50: ACF of Squared Standardized Residuals, ‘Hlth’ index

Figure 51: ACF of Standardized Residuals, ‘Other’ index
Figure 52: ACF of Squared Standardized Residuals, ‘Other’ index

Figure 53: Standardized Residuals, ‘Manuf’ index

Figure 54: Standardized Residuals, ‘HiTec’ index
Figure 55: Standardized Residuals, ‘Other’ index

Figure 56: QQ plot of Standardized Residuals vs. the normal dist., ‘Manuf’ index

Figure 57: QQ plot of Standardized Residuals vs. the normal dist., ‘HiTec’ index
Figure 58: QQ plot of Standardized Residuals vs. the normal dist., ‘Hlth’ index

Figure 59: Mean Excess function of Negative Standardized Residuals, ‘Cnsmr’ index

Figure 60: Mean Excess function of Negative Standardized Residuals, ‘HiTec’ index
Figure 61: Mean Excess function of Negative Standardized Residuals, ‘Other’ index

Figure 62: Mean Excess function of Positive Standardized Residuals, ‘Cnsmr’ index

Figure 63: Mean Excess function of Positive Standardized Residuals, ‘Manuf’ index
Figure 64: Mean Excess function of Positive Standardized Residuals, ‘HiTec’ index

Figure 65: Mean Excess function of Positive Standardized Residuals, ‘Hlth’ index
Figure 66 to 73 show the GPD fit to the whole data and compare it to the Normal and t-student distributions. On the lower tail the Normal and t-student distributions underestimate the risk. They also overestimate the upper tail with respect to the EVT model.

Figure 66: Pareto, Normal, t and empirical lower tail, ‘Cnsmr’ index

Figure 67: Pareto, Normal, t and empirical upper tail, ‘Cnsmr’ index
Figure 68: Pareto, Normal, t and empirical lower tail, ‘Manuf’ index

Figure 69: Pareto, Normal, t and empirical upper tail, ‘Manuf’ index
Figure 70: Pareto, Normal, t and empirical lower tail, ‘HiTec’ index

Figure 71: Pareto, Normal, t and empirical upper tail, ‘HTec’ index
Figure 72: Pareto, Normal, t and empirical lower tail, ‘Hlth’ index

Figure 73: Pareto, Normal, t and empirical upper tail, ‘Hlth’ index
Figures 74 to 81 show the Backtest for the VaR for the Conditional Normal, Conditional EVT and Unconditional EVT models. As we can see the Unconditional methodology does not react fast enough to volatility changes.

![Figure 74: VaR estimation for the conditional normal, conditional EVT and unconditional EVT methods for the ‘Manuf’ index.](image1)

![Figure 75: VaR estimation for the conditional normal and conditional EVT with negative returns of the ‘Manuf’ index.](image2)
Figure 76: VaR estimation for the conditional normal, conditional EVT and unconditional EVT methods for the ‘HiTec’ index.

Figure 77: VaR estimation for the conditional normal and conditional EVT with negative returns of the ‘HiTec’ index.
Figure 78: VaR estimation for the conditional normal, conditional EVT and unconditional EVT methods for the ‘Hlth’ index.

Figure 79: VaR estimation for the conditional normal and conditional EVT with negative returns of the ‘Hlth’ index.
Figure 80: VaR estimation for the conditional normal, conditional EVT and unconditional EVT methods for the ‘Other’ index.

Figure 81: VaR estimation for the conditional normal and conditional EVT with negative returns of the ‘Other’ index.
Figure 82: Return Series ‘Cnsmr’ index.

Figure 83: Return Series ‘Manuf’ index.
Figure 84: Return Series ‘Hlth’ index.

Figure 85: Return Series ‘Other’ index.