Analysis of current risk measures and study of model sophistication, via Copulas and Extreme Value Theory, in the estimation of risk within the regulatory framework

Louvain School of Management

Authors: ANTOINE BODART & ADRIEN KINART

Supervisor: PR. FRÉDÉRIC VRINS

Université Catholique de Louvain

The thesis has been submitted in fulfillment of the requirements for the degree of MASTER IN BUSINESS ENGINEERING, option in Financial Engineering, at the *Louvain School of Management*.

Academic year 2018-2019

Abstract

The crisis of 2008 has brought to light deficiencies in the risk estimations performed by financial institutions. Despite warnings, the underlying risk of many portfolios has been underestimated. As such, in the view of the regulator and the bank within the current regulatory environment, one may wonder what is the best risk measure and what are the best estimation techniques with regard to the presumed riskiness of a portfolio. We show that, despite its non-elicitability and lack of robustness, the Expected Shortfall (ES) should continue to gain ground over the Value-at-Risk (VaR) as the risk measure reference. Actually, these two features are not even problematic because the non-elicitability does not prevent backtestability and non robustness to outliers might actually foresee future large losses. We also examine the properties of the risk measure procedure to estimate the VaR/ES, and shade light on the importance of the *economical bias* that arises in the plug-in estimation of the VaR and ES in the Gaussian framework. Although the bias is small for a 250 days based VaR/ES, the approach remains interesting and we recommend its use. We also show that there is no panacea to estimate the VaR and ES. Depending on the market conditions and the portfolio itself, the best estimation techniques vary. Through Copulas and Extreme Value Theory, we show that more sophisticated does not necessarily mean better because it may cause risk overestimation. This is shown via simulations of Gaussian log-return (low-risk) versus Johnson distributed log-returns with unforeseeable loss jumps (high risk), and via the VaR and ES estimations of an ETFs portfolio (presumed low-risky) versus a portfolio of cryptocurrencies (presumed high-risky). The EVT is the most conservative method since it produces the least number of violations in the backtesting phase. However, it tends to overestimate the risk if no outliers are found out of sample. It may receive the favor of regulators, or even of prudential risk managers. The Clayton copula is found to be the copula to advocate and its benefit is greater in context of high risk. They both outperform traditional parametric methods based on the Gaussian assumption, whereas the historical simulation and Monte Carlo jump remain good fit as long as the underlying environment is constant.

Dedication and acknowledgements

We, Antoine and Adrien, wish to express our sincerest gratitude and warm appreciation to Mr. Frédéric Vrins, thesis advisor, who largely contributed by helping us shape and reshape this piece of work. We thank him for his interest in our progress, for answering our questions, and for giving us valuable feedback all along the writing.

We also extend our general appreciation to our faculty, its members and professors who helped us build the knowledge we have today, otherwise it would not have been possible to provide this thesis.

I, Adrien, extend my deepest gratitude to my friend and co-authour, Antoine Bodart, in the realisation of this thesis. I deeply thank him of his support, understanding and dedication throughout this thesis.

I, Antoine Bodart, similarly and sincerely would like to thank my friend and co-author of this thesis, Adrien Kinart. His input, help, and advice undoubtedly helped steer this thesis in the right direction. The complementarity of our work and affinities has made this accomplishment possible. For that, I am grateful.

AUTHORS' DECLARATION

We, Antoine Bodart and Adrien Kinart, declare that, except where indicated by specific reference in the text, this work is our own work. The work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are ours.

Table of Contents

Page

List of Tables xi

List of Figures xiii

Introduction 1

I Theoretical perspective 5

1 Risk, risk measure and properties 7
- 1.1 Introduction . . . 7
- 1.2 The different types of risk . . . 8
- 1.3 Formal framework of risk . . . 9
- 1.4 Definition of a risk measure . . . 10
- 1.5 Desirable properties of a risk measure . . . 11
  - 1.5.1 Coherence . . . 11
  - 1.5.2 Comonotonic additivity . . . 11
  - 1.5.3 Law-invariance . . . 12
  - 1.5.4 Elicitability . . . 12
  - 1.5.5 Tail-risk . . . 13
- 1.6 Some well-known risk indicators . . . 13

2 Risk indicators and their Properties 15
- 2.1 Introduction . . . 15
- 2.2 Variance . . . 15
  - 2.2.1 Definition . . . 15
  - 2.2.2 Properties . . . 16
- 2.3 Value-at-Risk . . . 17
  - 2.3.1 Definition . . . 17
  - 2.3.2 Properties . . . 18

2.4 The Expected Shortfall . . . 20
2.4.1 Definition . . . 20
2.4.2 Properties . . . 21
2.5 Comparison . . . 22
2.5.1 The VaR and ES over the variance . . . 22
2.5.2 Tail-risk . . . 23
2.5.3 Coherence . . . 23
2.5.4 Elicitability and backtestability . . . 23
2.5.5 Robustness and sensitivity . . . 24
2.6 VaR and ES in regulation . . . 24
2.7 Summary . . . 25

3 The desirable properties of the risk estimator 27
3.1 Introduction . . . 27
3.2 Statistical bias . . . 27
3.3 Consistency . . . 28
3.4 Efficiency . . . 29
3.4.1 Relative efficiency . . . 29
3.4.2 Robustness of the estimation procedure . . . 30
3.5 The Economical bias . . . 30
3.5.1 The problem at hand and its solution . . . 30
3.5.2 Economically unbiased VaR estimator in the Gaussian framework . . . 31
3.5.3 Economically unbiased ES estimator in the Gaussian framework . . . 33
3.6 A new property: the economical consistency . . . 35
3.6.1 Case of the value-at-risk . . . 35
3.6.2 Case of the expected shortfall . . . 37
3.7 Conclusion . . . 39

4 Conventional methods of estimation for the VaR and ES 41
4.1 Introduction . . . 41
4.2 Non-parametric approach . . . 41
4.2.1 Historical Simulation . . . 42
4.3 Parametric approach . . . 43
4.3.1 Risk metrics . . . 43
4.3.2 GARCH processes . . . 44
4.3.3 Monte-Carlo methods . . . 45
4.4 Discussion . . . 46

5 The method of copulas 47

5.1 Introduction . . . 47
5.2 Definition . . . 47
5.3 Important properties of the Copula . . . 48
5.4 Dependence structure . . . 49
5.4.1 Kendall's tau and Spearman's rho . . . 49
5.4.2 The tail dependence . . . 50
5.5 Implicit copula . . . 50
5.5.1 The Gaussian Copula . . . 51
5.5.2 The Student t-Copula . . . 51
5.6 Explicit copula . . . 51
5.6.1 Clayton Copula . . . 52
5.6.2 Gumbel Copula . . . 52
5.7 Estimation of the VaR and ES . . . 53

6 The Extreme Value Theory **55**
6.1 Introduction . . . 55
6.2 The Block Maxima model . . . 55
6.2.1 Generalized extreme value distribution . . . 56
6.3 The Threshold Exceedance model . . . 57
6.3.1 Generalized pareto distribution . . . 58
6.3.2 Choice of the threshold . . . 58
6.3.3 VaR and ES in POT . . . 59
6.4 Discussion . . . 60

7 Backtesting framework **61**
7.1 Introduction . . . 61
7.2 The VaR backtesting framework . . . 61
7.3 The ES backtesting possibilities . . . 64
7.4 The Basel framework for backtesting . . . 66

II Practical and empirical analysis **69**

8 Risk assessment of a simulated low and high risk portfolio **71**
8.1 Introduction . . . 71
8.2 Low risk simulations description . . . 72
8.3 High risk simulations description . . . 74
8.3.1 Drivers of the portfolio riskiness . . . 74
8.3.2 Unbounded Johnson distributed log-returns . . . 75
8.4 Copulas and EVT fitting and comparisons for the simulations . . . 77

8.4.1 Fitting Copula model . . . 77
8.4.2 Fitting Extreme Value Theory model . . . 79
8.4.3 Fitting GARCH model . . . 82
8.5 Backtesting results and performance comparisons . . . 83
8.6 Stability of the estimation of the VaR and ES . . . 88
8.7 Conclusions . . . 89

9 Risk assessment of a presumed low and high risk real portfolio 91
9.1 Introduction . . . 91
9.2 Technical sheet to assess the riskiness of a portfolio . . . 92
9.3 Presumed low risky products: the Exchange-Traded Funds . . . 92
9.3.1 Why to choose ETF as low risk profile . . . 92
9.3.2 Exploratory analysis of an equally weighted portfolio of ETF . . . 93
9.4 Presumed high risky products: the Cryptocurrencies . . . 97
9.4.1 Why to choose cryptocurrencies as high risk profile . . . 97
9.4.2 Exploratory analysis of an equally weighted cryptocurrencies portfolio . . . 98
9.5 Backtesting results and observations . . . 101
9.5.1 ETF portfolio . . . 101
9.5.2 Cryptocurrencies portfolio . . . 103
9.6 Conclusions . . . 106

Conclusion and further research 106

A Appendix 111
A.1 Results of the backtesting . . . 111
A.1.1 Normal-distributed portfolio . . . 111
A.1.2 Johnson-distributed portfolio . . . 113
A.1.3 ETF portfolio . . . 115
A.1.4 Cryptocurrencies portfolio . . . 117
A.2 GPD fitting . . . 119
A.2.1 Normal-distributed portfolio . . . 119
A.2.2 Johnson-distributed portfolio . . . 120
A.2.3 ETF portfolio . . . 120
A.2.4 Cryptocurrencies portfolio . . . 121
A.3 Economical unbiasedness: sensitivity analysis for the ETF portfolio . . . 123
A.4 Stability of the VaR and ES: samples mean and discrepancy with the true value . 123

List of Tables

Table **Page**

2.1 Properties of risk measures 25

7.1 Backtesting zones for VaR, computed assuming $\alpha = 0.01$ and N = 250 observations . 67
7.2 Backtesting zones for ES, computed assuming $\alpha = 0.025$ and N = 250 observations . 68

8.1 Number of exceedances of the 99% VaR and 97.5% ES estimates calculated using various estimation methods, and for the series of **normal simulated log-returns** (low-risk). Expected failures=2.5. 84
8.2 Number of exceedances of the 99% VaR and 97.5% ES estimates calculated using various estimation methods, and for the series of **Johnson's S_U log-returns** (high-risk). Expected failures=2.5. 86
8.3 Standard deviation associated with the estimation of the $ES_{1\%}^{\text{plug-in}}$, the $VaR_{1\%}^{\text{plug-in}}$ (a), the $ES_{2.5\%}^{\text{plug-in}}$ (b) over the samples and for different size of time window. The difference between (b) and (a) is also provided for comparison. 88

9.1 Values of the first four empirical moments of the sample for the ETF portfolio log-returns.) 96
9.2 Descriptive statistics for the crypto portfolio's log-returns. 100
9.3 Number of exceedances of the 99% VaR and 97.5% ES estimates calculated using various estimation methods, for the **ETF portfolio's log-returns**. Expected failures=2.5. 101
9.4 Non absorbed loss over a $\$1,000,000$ worth portfolio in ETF, and a maximal loss of $\$51,059$. 103
9.5 Number of exceedances of the 99% VaR and 97.5% ES estimates calculated using various estimation methods, and for the series of **crypto portfolio's log-returns**. Expected failures=2.5. 104

A.1 Difference between the samples mean and the true value of the risk measure for the $ES_{1\%}^{\text{plug-in}}$, the $VaR_{1\%}^{\text{plug-in}}$, and the $ES_{2.5\%}^{\text{plug-in}}$ for different time window size. 124

List of Figures

Figure **Page**

2.1 Illustration of the variance and standard deviation for the standard normal distribution . . . 16
2.2 Illustration of the VaR and CVaR/ES . . . 17
2.3 Levels of VaR (95% confidence level) and ES. . . . 20
2.4 Historical volatility of the S&P 500. *Source: Macrotrends: VIX Volatility Index - Historical Chart.* . . . 22

3.1 Illustration of the asymptotic convergence of the plug-in VaR to the real VaR with regard to the window size, that goes from 20 to 1500 observations by steps of 5. For each of them, 1,000,000 simulations were run. . . . 36
3.2 Illustration of the asymptotic convergence of the plug-in ES to the real ES with regard to the window size, that goes from 45 to 1500 observations by steps of 5. For each of them, 100,000 simulations were run. . . . 36
3.3 Average exceedance rate (over 350,000 simulations) with regard to an increasing time window size on which the plug-in VaR, the ec. unbiased VaR and the real VaR are computed ($\alpha = 1\%$). . . . 37
3.4 Approximated a_n derived from 1 million simulations for each window size (in red) and in blue the extrapolated a_n derived from a trend line (model 'power3P' in R) . 38
3.5 Average exceedance rate (over 350,000 simulations) with regard to an increasing time window size on which the plug-in ES, the ec. unbiased ES and the real ES are computed ($\alpha = 2.5\%$). . . . 39

6.1 Densities for the Fréchet, Weibull and Gumbel functions. . . . 56
6.2 Example of a Mean excess plot. . . . 59

8.1 Illustration of the simulated correlated stock prices ($\rho = 0.711$). . . . 74
8.2 Illustration of the simulated low risk log-returns (left) and high risk log-returns (right). The high risk portfolio is populated with more and lower negative log-returns. 76

8.3 Comparison between the Johnson (J) and Normal (N) simulated densities. $\gamma_J = -1.27$; $Kurt_J = 5.9$, and $\gamma_N = -0.31$; $Kurt_N = 3.3$. μ and σ are the same in both simulations. 76
8.4 10,000 simulated points from two normally assumed log-returns with dependence structure from Copula: (a) Gaussian, (b) t, (c) Clayton, and (d) Gumbel. $\tau = 0.515$. . 77
8.5 3D plot of the probability density function under the Clayton copula. The left tail is particularly inflated. 78
8.6 Levels of value-at-risk for the Copula at various levels alpha ($\alpha \in [0.01; 0.05]$). 78
8.7 Levels of expected shortfall for the Copula at $\alpha \in [0.01; 0.05]$. 78
8.8 10,000 simulated points from two Non-normal (Johnson) marginals Y1 and Y2 for the log-returns with dependence structure from respectively Copula: (a) Gaussian, (b) t, (c) Clayton, and (d) Gumbel. $\tau = 0.505$. 79
8.9 Method of blockmaxima and threshold exceedances over the low risk simulated data. The block window is defined to 20 days and the threshold to 0.01 to best fit the data. 80
8.10 Mean excess plot for (a) Normal, and (b) Johnson simulated log-returns. 81
8.11 Diagnostic plots of fit for the GPD with Gaussian simulated log-returns (250 first observations). The threshold $u = -0.005$. The figure at the top represents the comparisons of quantiles and the figure at bottom shows the fitted GPD against the empirical distribution. 82
8.12 Daily losses for the backtesting period with risk measure estimate $VaR_{1\%}$, and exceedances for the methods in the legend. All backtesting results for VaR and ES to be found in Appendix A.1.1. *Portfolio: Normal*. 85
8.13 Daily losses for the backtesting period with risk measure estimate $VaR_{1\%}$, and exceedances for the methods in the legend. All backtesting results for VaR and ES to be found in Appendix A.1.2. *Portfolio: Johnson*. 87

9.1 Daily prices of the ETFs constituting the studied ETF portfolio (502 observations). . 93
9.2 Historical daily log-returns of the ETFs and the equally weighted portfolio (bottom right). 94
9.3 ACF and PCF of the log-returns of equally weighted portfolio of ETF. 95
9.4 Gaussian fitting on to the log-returns of the equally weighted portfolio of ETFs (left) and its respective boxplot (right). 96
9.5 Scatter plots of the ETF margins against each other. 97
9.6 Price index of the Bitcoin against the dollar over the period from 05-2017 to 06-2019. The *Bitcoin bubble* is visible just before 2018. 98
9.7 Time series of log-returns of the (a) Bitcoin, (b) Ethereum, (c) Litecoin, and (d) XRP for the period from 05-2017 to 06-2019. The mean appears in green. 98

9.8 Time series of log-returns of the crypto equally weighted portfolio. The mean appears in green. 99

9.9 ACF and PCF of the log-returns of the crypto portfolio. 100

9.10 Backtesting of the **ETF** portfolio: Daily losses for the backtesting period with risk measure estimate $VaR_{1\%}$, and exceedances for the methods in the legend. All backtesting results for VaR and ES to be found in Appendix A.1.3. 102

9.11 Daily losses for the backtesting period with risk measure estimate $VaR_{1\%}$, and exceedances for the methods in the legend. All backtesting results for VaR and ES to be found in Appendix A.1.4. *Portfolio: Cryptocurrencies*. 104

9.12 A comparison between the $VaR_{1\%}^{\text{Plug-in}}$ and the Economically unbiased $VaR_{1\%}$ in backtesting for the **crypto. portfolio**. 106

A.1 Backtesting results of the $VaR_{1\%}$ for all the considered risk estimation methods. Only the negative log-returns (loss side) are plotted. While the EVT-based methodologies tend to overestimate the risk, the difference in term of the number of failures between the other methods is quite narrow. We can also point out the irregularity of the GEV, which can be assumed to be a noticeable pitfall of the method. 111

A.2 Backtesting results of the $ES_{2.5\%}$ for all the considered risk estimation methods. Only the negative log-returns (loss side) are plotted. The results obtained for the expected shortfall at the 2.5% significance level are very close to the VaR 1%. We may therefore argue that the higher severity of the ES (α can still be increased) is desirable, as well as its relevant properties. 112

A.3 Backtesting results of the $VaR_{1\%}$ for all the considered risk estimation methods. Only the negative log-returns (loss side) are plotted. We remark that the different estimation methods are more scattered than for the normal log-returns. The riskiness is a key determinant for the relevance of certain methods. The copula are good performer, with the Clayton providing a superior estimation, followed by the t, Gaussian, and Gumbel copulas in order. Parametric methods on the behalf of the RM, GARCH and MC produce poor results. The EVT is the most conservative, prudential method. 113

A.4 Backtesting results of the $ES_{2.5\%}$ for all the considered risk estimation methods. Only the negative log-returns (loss side) are plotted. The results are similar to those of the VaR 1%. The estimations may, however, sometimes be slightly greater. This is inherent to the behavior of the ES that looks further in the tail. Tail where we find more low returns in this portfolio. 114

A.5 Backtesting results of the $VaR_{1\%}$ for all the considered risk estimation methods. Only the negative log-returns (loss side) are plotted. The variability out of sample symbolized by the few peaks around the 100^{th} and 200^{th} observations leads to a powerful EVT (GPD) that outperforms all other estimation methods. It is also interesting to note the superior performance of the Clayton copula compared to the other copula. The difference is more pronounced here. The GARCH and RM are found very reactive to market changes but produce many hits since there is no clustering of volatility. Parametric methods are again bad performers, while the HS(-GARCH) and MC with jump are satisfying. 115

A.6 Backtesting results of the $ES_{2.5\%}$ for all the considered risk estimation methods. Only the negative log-returns (loss side) are plotted. The backtesting results are very close to those of the VaR 1%. 116

A.7 Backtesting results of the $VaR_{1\%}$ for all the considered risk estimation methods. Only the negative log-returns (loss side) are plotted. The taking into account of the *Bitcoin bubble* produces high estimations for, more or less, all the methods. The GPD remains the methods associated with the lowest number of failures, however slightly overestimating the risk in this case (the adjustment at the end is good point). The superiority of the Clayton copula against other copula is even more established here. The Gumbel copula is pretty low. It is also interesting to highlight how the copula estimates are narrowing as far as the outliers disappear from the estimation sample. 117

A.8 Backtesting results of the $ES_{2.5\%}$ for all the considered risk estimation methods. Only the negative log-returns (loss side) are plotted. The backtesting results are very close to those of the VaR 1%. 118

A.9 Diagnostic plots of fit for the GPD with Normal log-returns. The threshold $u = -0.005$. 119

A.10 Diagnostic plots of fit for the GPD with Johnson log-returns. The threshold $u = -0.01$. 120

A.11 Mean excess plot for the ETF's log-returns. 120

A.12 Diagnostic plots of fit for the GPD with the ETF's log-returns. The threshold $u = 0$. . 121

A.13 Mean excess plot for the log-returns of the crypto portfolio. 121

A.14 Diagnostic plots of fit for the GPD with the crypto portfolio's log-returns. The threshold $u = 0$. 122

A.15 A comparison between the $VaR_{1\%}^{\text{Plug-in}}$ and the Economically unbiased $VaR_{1\%}$ in backtesting for the **ETF portfolio**. Over an increasing time window size, from 10 to 400 days, we recorded the number of exceedances of the plug-in and econ. unbiased $VaR_{1\%}$, respectively. The results obtained are similar as for the crypto-currencies portfolio, whose analysis is provided 9.5.2. We observe that the plug-in VaR returns more violations as far as the time window considered for the backtesting is small. After what it converges to the economical form (around a window size of 200). Over a 250 days backtesting window, as per the Basel regulation, no real difference is to be noted. 123

INTRODUCTION

The today's financial environment cannot be dissociated from the notion of risk. In particular, since the financial crisis of 2008, concerns about risk evaluation are pointed out. This has led to numerous criticisms about the existing risk measures and their estimation methods, and motivated the search for more appropriate methodologies able to cope with rare events that have heavy consequences. The crisis had indeed dramatic consequences on the economy and compelled the financial sector and its core institutions to question their processes.

Especially, although it cannot be taken responsible on its own, the risk estimation methods used by financial institutions at that time showed weaknesses that precipitated the collapse of the whole system. Among all different financial risk measures, the Value-at-Risk (VaR) is one of the most known and widely used in the financial industry to calculate risk exposures. Generally speaking, based on a loss distribution, the VaR is the maximum loss that can be expected at a given confidence level α, over a given period of time, and on a given portfolio of financial assets. The VaR is thus a quantile-based risk measure that directly depends on the shape of the probability density of losses. In addition, knowing a key tenant of financial theories assumed that market returns are normally distributed, the probability that a crisis occur is almost null and is certainly not adequately captured by the model. Therefore, the fatter the tails of the asset return distribution under consideration, the larger the discrepancy with the normal distribution, and the larger the errors made in VaR estimation. It may lead to risk underestimation triggering crises. Already at the end of the 20th century, it was admitted that crisis events were in the 'tails' of distributions and that traditional estimation methods of VaR were less performing for measuring and monitoring market risk in that context (Yamai and Yoshiba, 2001).

Even, if the use of VaR remains widespread across the financial sector, it is commonly argued that it needs to be improved, aimed to avoid more crisis. In order to address that issue, the Basel Committee has reinforced its policies and notably requires now that banks hold an increasing amount of capital to respond the risk over their assets. That has also fostered a shift in the predominant risk measure from VaR $\alpha = 99\%$ to the benefit of Expected shortfall (ES) $\alpha = 97.5\%$ in 2019 (BCBS, 2019b). As it will be discussed, the ES has more desirable properties that may tackle the concerns raised by the VaR.

Concretely, to ensure a competent regulatory framework with respect to Value-at-Risk, and

Expected shortfall as it is gaining in importance, for establishing a bank's capital adequacy requirements, the approaches for estimating risk need to incorporate the fat tails apparent in the return distributions of financial assets (Hiusman et al., 1998). In line with that, the Copula and Extreme Value Theory (EVT) showed up as serious candidates to substantially improve the risk estimation. Whereas the Copula produce a joint distribution by capturing specific dependence aspects in the marginals, the EVT is dedicated to inflate the distribution's tails. Overall, improving the risk measurement appears as a necessity for the health and stability of financial markets, yet of our economy. It is even an ethical duty toward the society as a whole.

The purpose of this thesis is, for the most relevant risk measures, to review their conventional methods of estimation, and examine whether sophisticated methods based on Copula and EVT may improve the capital determination. We show that the EVT is the most conservative method since it produces the least number of exceedances in backtesting. However it tends to overestimate the risk if no outliers are to be found in the out of sample. It may receive the favor of regulators and appears to be the only model able to help banks withstand periods of market stress (e.g. crisis). The Clayton copula is found to be the copula to preconize in risk management and its benefit is greater in context of high risk. With this thesis, we hope to provide input that may help enhance the risk assessments performed by financial institutions.

The paper is organized as follow:

- Chapter 1 introduces the notion of risk and risk measure. A formal framework for the risk and the definition of a risk measure are firstly given, before we deeply examine the desirable properties it should have. This chapter ends with the enumeration of some common risk indicators.

- Chapter 2 contains a comparative analysis about the ability of the variance, VaR, and ES to fulfill the desirable properties of a risk measure. The place of VaR and ES in regulation is also presented.

- Chapter 3 focuses on the risk estimator. It reviews the traditional properties that are expected from a good estimator, and most importantly, investigates the notion of economic unbiasedness, firstly introduced by Pitera and Schmidt (2018), particularly valuable in the context of economic capital estimates.

- Chapter 4 reviews the conventional estimation methods for the VaR and ES, whereas Chapter 5 and 6 are dedicated to the presentation of the methods of Copula and EVT, respectively.

- Chapter 7 introduces the backtesting procedure for the VaR and ES, firstly in a theoretical approach, and secondly considered within the Basel framework.

- Chapter 8 and 9 contain an empirical analysis split in two parts: (i) an analysis over low risk and high simulated log-returns, (ii) an analysis over two real portfolios made out

of Exchange Traded Funds and Cryptocurrencies, respectively, aimed to reflect two risk profiles. The data are extracted for the two past years. We lastly conclude the document by summarizing our key results and providing some suggestions for future research.

Part I

Theoretical perspective

CHAPTER 1

RISK, RISK MEASURE AND PROPERTIES

1.1 Introduction

Risk Management is one of the most crucial skills of financial practitioners. It is a particularly general discipline consisting in identifying, assessing and prioritizing risks that can be defined as the *effect of uncertainty on objectives*. It is the discipline of living with the possibility that future events may cause adverse effects (McNeil et al., 2005). Therefore, in order to avoid any deflection of a firm's endeavors from its business objectives due to certain risks and/or uncertainties, companies have to put in place various risk management processes whose primary objective is to mitigate either the impact or the likelihood of the identified risks.

Risks are traditionally assessed through the impact/likelihood framework which aims at knowing which risks need to be prioritized over others, in order to minimize the resources used to mitigate the risks. Obviously, risks characterized with a high likelihood of occurrence and important severity will need to be dealt with first.

Accurately assessing the risk exposure is of crucial importance in any company/sector, since it can be defined as the likelihood that a given risk is realized times its impact, i.e. its expected losses (or gains). As said above, managers often base their risk management processes and prioritization on this risk exposure. Therefore, being able to properly assess it is particularly important, since it'll be the backbone of the whole risk management policy of a given company.

Given that from the crisis increasing regulatory requirements have emerged, risk management is a field that has gained in importance in the financial industry. It indeed plays a major role in the computation of economic capital so that banks need to accurately quantify their risk, outlining fair values that must be derived through reliable and very precise methods. Therefore, in this work, the risk evaluation will be rather tackled in the form of the determination of

economic capital hold as a buffer against unexpected losses.

Banks' performance. For the bank, the amount of capital to be added is also relevant from the business performance perspective. Indeed, the capital pledged by the bank is immobilized in its balance sheet, hence unavailable for investment or other actions that may increase the bank's profitability while it still needs to remunerate its shareholders. Then, increasing regulatory constraints may prevent banks from efficiently allocating resources. It is a shortfall. Overall, regulatory compliance can be seen costly, in line with the wide-spread belief that the cost of equity is greater than the cost of debt because of tax distortions. This is in spite of the *Modigliani–Miller Theorem* that states that the financial structure of a firm is irrelevant for determining the firm's value (Modigliani and Miller, 1958). The essence of the MM theorem is as follow: a higher leverage makes equity more risky, so if leverage is brought down the required return on equity financing is likely to fall. It is also argued that the MM theorem does not hold to the full extend in practice such that the weighted average cost of capital of the bank goes up slightly when it has to increase its capital (David Miles and Marcheggiano, 2012). In this work, we will therefore assume that an increased capital cushion is likely to be less favorable for the bank profitability. However, it clearly preserves financial stability, thereby ensuring the bank solvency which is key for the going concern of the institution, and for shareholders' interests, or the society as a whole.

Hence, for all levels and parties, the risk estimations are crucial and need to be conducted in accordance.

1.2 The different types of risk

Many types of risk exist, which a given financial institution will have to deal with throughout its lifetime. Here are introduced the major risk categories.

1. **Market risk**: Market risk is defined as the risk of losses arising from movements in market prices (BCBS, 2019a). It directly relates to the uncertainty inherent to financial markets. That risk cannot be removed through diversification, but can only be hedged using financial instruments.

2. **Credit risk**: risk that comes from borrowers in case they default, i.e. if they fail to pay back promised repayments. It typically comes from bonds (firms) and loans (individuals).

3. **Operational risk**: risk resulting from bad internal processes, people, systems or external events. It summarizes the risks a company undertakes when it operates within a given industry. Operational risk is not inherent in financial, systematic or market-wide risk.

4. **Liquidity risk**: risk associates with the time it can require to sell a financial product, i.e. its marketability. The *funding-liquidity risk* refers to the inability to meet payment obligations which may force early liquidation.

An additional category of risk is the *model risk*. The former is the risk that arises from using a misspecified or inappropriate model for measuring risk.

In this thesis, we will focus on the ***market risk***. We are mainly interested in movements in the value of a financial investment through time that may affect the future level of cash flows that will be earned. These are indeed at the origin of potential future losses incurred by a firm. In addition, forced to recognize that the other types of risk would be more difficult to assess using the VaR or ES measure.

1.3 Formal framework of risk

In the financial industry, risk is mainly related to uncertainty and, hence, to the notion of randomness. It turns out that risk management is principally a statistical issue. Based on historical observation and a model, or risk measurement technique, a statistical estimate of the distribution of the change in value of a position is calculated (McNeil et al., 2005). From a statistical view point, we can refer to the Kolmogorov theory in which a probabilistic model is described by the probability space $(\Omega, \mathcal{F}, P)$. The probability space is the domain of the random variables we introduce hereafter, and the uncertainty about future states of the world is represented through it. P is the probability measure attached to historical outcomes. In a portfolio, one can represent the risky positions a time t (assumed to be the log-prices of the assets in this work) by a random vector $\boldsymbol{Z_t}$, a function of the previously defined probability space. Risk factors are usually assumed to be observable such that Z_t is known at time t. Considering a given portfolio, the loss over the latter can be expressed in term of the portfolio's value across a specific period. Typically, the loss of the portfolio with denoted value, observable at time s, V_s, a random variable, over time horizon Δ, is given by

$$L_{s,s+\Delta} = -(V_{s+\Delta} - V_s)$$

where $L_{s,s+\Delta}$ is the loss from day s to day $s+\Delta$. $L_{s,s+\Delta}$, while observable at time $s+\Delta$, is random from the perspective of s. The distribution of $L_{s,s+\Delta}$ is the so called *loss distribution*. It is termed the loss distribution because in risk management we are mainly interested in the probability of large amounts occurring in the upper tail of the distribution: the loss side. Note that the loss value returns a positive number. The value of the portfolio can be also represented as a function of time and a set of risk factors Z_t, leading to the representation

$$V_t = f(t, Z_t)$$

As per McNeil et al. (2005), this mapping of risk can be conveniently defined according to the series of *risk factors changes* $X_t = Z_t - Z_{t-1}$, the so-called log-returns which are "the objects of interest in most financial studies of financial time series". **Typically, since common practices**

in finance carry analysis on logarithmic prices as risk factors, *risk factor changes* usually corresponds to log-returns of the position. Hence, the portfolio loss can be rewritten

$$\begin{aligned} L_{t+1} &= -\left(f(t+1, Z_{t+1}) - f(t, Z_t)\right) \\ &= -\left(f(t+1, Z_t + X_{t+1}) - f(t, Z_t)\right) \end{aligned}$$

Since Z_t is known at time t, the loss distribution is determined by the distribution of the risk factor change X_{t+1}. Thus, we obviously have

$$L_{t+1} = l_{(t)}(X_{t+1}) \tag{1.1}$$

with $l_{(t)}$ the loss operator.

Note that the loss distribution can be assumed to be conditional to the publicly available information gathered up to time t, i.e. current time, and conveyed by the sigma-field $\mathbb{F}_t$ (McNeil et al., 2005). Typically, $\mathbb{F}_t = \mathscr{F}(X_s : s \leqslant t)$, the sigma-field generated by present and past risk factor changes until, and including, time t. Thus we denote $F_{X_{t+1}|\mathbb{F}_t}$. It represents the conditional distribution of X_{t+1} given current information $\mathbb{F}_t$. Linked to the loss distribution in Eq 1.1, the conditional loss distribution function can be written as follow

$$F_{L_{t+1}|\mathbb{F}_t}(l) = \Pr\left(l_{(t)}(X_{t+1}) \leqslant l \mid \mathbb{F}_t\right) = \Pr\left(L_{t+1} \leqslant l \mid \mathbb{F}_t\right)$$

i.e. the probability that the loss in $t+1$ will be lower than a certain amount, given the information conveyed up to and including time t. Hence the modelling of the risk associated to a risky position is mainly related to its distribution function $F_L(l) = \Pr(L \leq l)$. Most modern measures of the risk are indeed statistical quantities describing the loss distribution of the portfolio over a certain horizon. Conditional distribution are particularly relevant in market risk management (McNeil et al., 2005).

Those are the building blocks of a mathematical description of risk.

1.4 Definition of a risk measure

The definition of risk is taken within the formal framework of risk section 1.3. We denote $\mathscr{L} := \mathscr{L}(\Omega, \mathscr{F}, P)$ the set of random variables related to market risk on $(\Omega, \mathscr{F}, P)$. It represents the portfolio losses over Δ. We assume to have a sample $X_1, ..., X_t$ of risk factors changes observations at hand up to time t, and estimate the risk of a future position X. Formally:

Definition 1.4.1. *A risk measure ρ is a mapping function from $\mathscr{L}$ to $\mathrm{R} \cup +\infty$.*

Based on a RV and its assumed distribution, the risk measure returns a real number (i.e. capital amount) as the estimation of the risk; the risk is *mapped*. A measure thought of as a map from spaces of probability random variables into real numbers is referred to as a *functional*.

From a financial point of view, we define the value $\rho(X)$, assigned by the measure ρ to the risk factor X, as the amount of money one has to add to the position X such that X becomes acceptable (McNeil et al., 2005). In addition, some underlying properties are important for the risk measure definition.

1.5 Desirable properties of a risk measure

The following section help determine the desirable features that a risk measure should have. This is believed to constitute a good theoretical measure of risk.

1.5.1 Coherence

The notion of coherence is a fundamental concept related to the acceptability of a risk measure. It was introduced by Artzner et al. (1999), who proposed four axioms that a risk measure ρ should meet, for a future loss L:

1. **Translation invariance:** $\rho(L+l) = \rho(L) + l$ where l is a deterministic quantity ($\forall L \in \mathscr{L}$). It simply states that adding (or subtracting) a deterministic quantity will alter the capital requirement by exactly this amount.

2. **Subadditivity:** $\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L2)$, $\forall L_1, L_2 \in \mathscr{L}$. It reflects the idea that diversification decreases the overall risk. Hence, subadditivity ensures that the diversification principle holds, and is desirable in that sense.

3. **Positive homogeneity:** $\rho(\lambda L) = \lambda\rho(L)$, $\forall L \in \mathscr{L}$, where $\lambda > 0$. This reflects the proportionality of the risk measure on a position to the size of the risk exposure of that position.

4. **Monotonicity:** $\forall L_1, L_2 \in \mathscr{L} : L_1 \leq L_2,\ \rho(L_1) \leq \rho(L_2)$. It means that the positions that lead to higher losses require more risk capital.

1.5.2 Comonotonic additivity

Complementary to the sub-additive axiom, it is sometimes added that when variables are assessed to be comonotonic, the associated risk measures should be additive (S. Wang et al., 1997).

Definition 1.5.1. *The random variables (RVs) L_1 and L_2 are comonotonic if any change in the value of $L_1(L_2)$ results in a change of value of $L_2(L_1)$ in the same direction. Formally, they are comonotonic if*

$$\Big(L_1(\omega_1) - L_1(\omega_2)\Big)\Big(L_2(\omega_1) - L_2(\omega_2)\Big) \geq 0 \tag{1.2}$$

holds almost surely for any $\omega_1, \omega_2 \in \Omega$. Overall, comonotonicity is considered as the strongest possible dependence of RVs (Emmer et al., 2015).

That explains the term comonotonic, which means *commonly monotone*. Thus if L_1 and L_2 are commonotic, the outcome of L_1 and L_2 always move in the same direction, such that there is no hedge or diversification when pooling the two risks. The strong dependence between both random variables induces by comonotonicity suggests the absence of diversification benefit such that the aggregated risk is exactly additive. This translates to the following axiom:

Comonotonic additivity: If L_1 and L_2 are comonotonic, then a good measure of risk should respect

$$\rho(L_1 + L_2) = \rho(L_1) + \rho(L_2)$$

1.5.3 Law-invariance

Risk measures that depend solely on the distribution of the losses are of special interest because their value can be estimated from loss observations only. This leads to the following definition inspired from the work of Kusuoka (2001) :

Definition 1.5.2. *A risk measure ρ is law-invariant if*

$$\Pr(L_1 \leq l) = \Pr(L_2 \leq l),\ l \in \mathbb{R} \Rightarrow \rho(L_1) = \rho(L_2) \tag{1.3}$$

that is, two random variables L_1 and L_2 with the same distribution are assigned the same value of ρ. It then means that the risk measure can be identified with a function of the loss distribution F_L (see section 1.3) exclusively. Especially, the law-invariant property allows one to compute $\rho(L)$ only by knowing the probability distribution of L.

1.5.4 Elicitability

As pointed out before, the risk measure is a mapping function that informs about an amount of economic capital needed to secure a position. An estimation of the risk measure is required in order to deliver such a result, which forces one to make choices regarding the model, methods, parameters. Therefore one ends up with competing estimation methods that would need to be confronted and validated. The elicitability is precisely a mathematical property that helps for models, and estimation, selection and ranking for a given risk measure. Hence, risk measures for which such validation and estimation performance comparison are possible, are called *elicitable* (Roccioletti, 2016). Note that the concept of elicitability is as a result more determinant with respect to model comparison than model testing (Kratz, 2017).

The idea is based on prediction assessment so that the estimations are mapped via a scoring (or error) function S that measures the deviation of the forecast y from the realization l. For a given risk measure, one consequently wishes to get estimates that minimize the score function. It is formally defined as

Definition 1.5.3. *A scoring function* $\mathbb{R} \times \mathbb{R} \to [0, \infty)$ *is a function satisfying:*

1. $S(y,l) \geq 0 \quad iif\ y = l$
2. $S(y,l)\begin{cases} increasing\ for\ y > l \\ decreasing\ for\ y < l \end{cases}$
3. $S(y,l)$ *is continuous in* y

for any $y, l \in \mathbb{R}$

Accordingly, an **elicitable statistical functional** (risk measure) is a functional for which it exists a scoring function that it can minimize. Formally, it states that:

Definition 1.5.4. *A functional* ρ*, that is defined on a space of d.f.* χ*, is* ***elicitable*** *on* $\chi_\rho \subset \chi$ *if there exist a scoring function* S *such that, for every* $F \in \chi_\rho$*:*

1. $\int_{\mathbb{R}} S(y,l)\, dF(l) < \infty,\ \forall y \in \mathbb{R}$
2. $\rho(F) = \arg\min_{y \in \mathbb{R}} \int_{\mathbb{R}} S(y,l) dF(l)$

We say that, in such case, the scoring function S is strictly consistent for ρ. Applied to risk measures, ρ is elicitable if $y \in \rho(F)$ is an optimal point forecast (estimate) under S, that is if it minimises the expected score $\mathbb{E}\Big(S(y,l)\Big) = \int_{\mathbb{R}} S(y,l)\, dF(l)$ for a RV L that represents the losses with its respective CDF. Undoubtedly, we can affirm that elicitability provides a natural methodology to perform backtesting (Roccioletti, 2016). However, we already point out that backtestability does not resume to the notion of elicitability.

1.5.5 Tail-risk

Generally speaking, while addressing the notion of risk measure, one wishes to define a statistic, based on the loss distribution $F_L(l)$, that measures the severity of the risk of holding a position over a predetermined time period (McNeil et al., 2005). It is indeed acknowledged that portfolio returns might move very far in the tails. As a results, we find it desirable for a risk measure ρ to take into account the tail-risk of the distribution, and to be able to deal with the skewness it may depict.

1.6 Some well-known risk indicators

First of all, one could be tempted to come up with the **variance** $\boldsymbol{\sigma^2}$ as risk measure. It measures the average dispersion of data from its expected value. Historically, ever since Markowitz (1952), it has become common to use σ^2 to measure the risk of a position. Another important risk indicator, probably the most widely used risk measure in finance up to today after it

was introduced by RiskMetrics (J.P. Morgan) in 1994, is the so-called **Value-at-Risk** (VaR). It assesses the level of risk related to a portfolio via an estimation of the quantile of its distribution. It gained much importance since the rules on capital requirements from the Basel II that enforced financial institutions to meet capital requirements based on VaR estimates. However, it has recently lost some ground with the advent of the so-called **Expected Shortfall** (ES). It assesses the *tail risk* of an investment by assessing what happens beyond a given threshold and confidence interval. It accounts for the average losses exceeding the VaR threshold and depends on the full distribution in the tails. It took full effect from the 1^{th} of January 2019 in the capital requirements of the banks. Finally, it seems that a new risk measure estimation is growing in importance: the **expectiles**. They are similar to quantile regressions (QR) except that they are defined by tail expectations rather than tail probabilities (Newey and Powell, 1987). QR is itself similar to linear regression except that rather than estimating an expected conditional value, it aims at estimating, conditionally to independent variables, a certain quantile (e.g. the median) of a dependent variable.

In the next chapter, we emphasis the difference between those standard risk measures within the framework of desirable properties, except for the *expectiles* since they are beyond the scope of this work and their interpretation in the context of banking regulations is less immediate.

CHAPTER 2

RISK INDICATORS AND THEIR PROPERTIES

2.1 Introduction

This chapter's purpose is to identify the risk measure with the most desirable properties. This will be achieved by investigating whether the variance, Value-at-Risk, and Expected shortfall fulfill the set of desirable properties as framed in chapter 1. This includes the *tail-risk*, *coherence*, *elicitability*, *comonotonic additivity*, and *law-invariance*.

2.2 Variance

2.2.1 Definition

The variance became the dominating risk measure in finance after Markowitz contributions in portfolio optimization.

Definition 2.2.1. *The variance of the RV L is its expected value of the squared deviation from the mean:*

$$Var(L) = \sigma_L^2 = \mathbb{E}\left[\left(L - \mu_L\right)^2\right]$$

where $\mu_L = \mathbb{E}(L)$

The graphical illustration of the variance and standard deviation in given by the following figure:

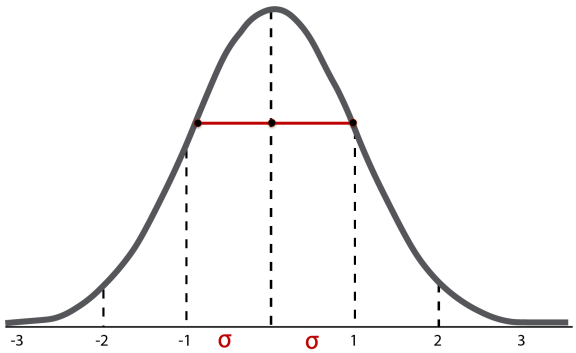

Figure 2.1: Illustration of the variance and standard deviation for the standard normal distribution

2.2.2 Properties

By definition, the notion of risk is related to the possible losses upon the value of a portfolio. To this regard, a first obvious drawback of the variance is its inherent consideration of the returns regardless their direction. It consequently assumes that the distribution is symmetric. However, it has been shown than many assets depict negative skewness so that the variance invariably underestimates the risk. When this is the case, the variance as a risk measure becomes invalid. Secondly, the variance does not account for the kurtosis, which translates to the phenomenon of "fat tails" in loss distributions. Indeed, the variance requires that the risks are random variables with a finite variance, while the variance tends to infinity in fat tails. In these aspects, the variance fails to take into account the tail-risk (see subsection 1.5.5), which is considered key in this thesis.

As framed by Artzner et al. (1999), the variance is **not coherent**. This crucial property can be proven with regard to the four fundamental axioms stated in section 1.5.1.

First, the variance is *not positive homogeneous*. Indeed, for a RV L_1 multiplied by a scalar λ , the property of the variance tells us that $\sigma^2(\lambda L_1) = \lambda^2 \sigma^2(L_1)$. It is straightforward that $\lambda^2 \sigma^2(L_1) \neq \lambda \sigma^2(L_1)$, which make it not positively homogeneous.

Moreover, we know that the variance of sum of two RV is

$$\sigma^2(L_1 + L_2) = \sigma^2(L_1) + \sigma^2(L_2) + 2\rho_{L_1,L_2}\,\sigma(L_1)\sigma(L_2) \tag{2.1}$$

where ρ_{L_1,L_2} is the correlation between L_1 and L_2.
Therefore, the *subadditivity* property of the variance depends on the sign of the correlation, which makes it not subbadditive.

Moreover, it can be seen from Equation 2.1 that the variance does not respect the property of *comonotonic additivity*. Indeed, when L_1 and L_2 are comonotonic, their correlation is 1. This makes the aggregate variance *super-additive*[1].

[1] i.e. $\sigma^2(L_1 + L_2) = \sigma^2(L_1) + \sigma^2(L_2) + 2\,\sigma(L_1)\sigma(L_2)$

Regarding the *monotonicity* of the variance, taking L_1 and L_2 such that $L_1 \geq L_2$, we cannot ensure that $Var(L_1) \geq Var(L_2)$ since the variance simply measures the dispersion around the mean. It is consequently not monotonic.

Finally, the variance is neither *translation invariant* since the variance of a constant is null. Formally, for a constant a: $\sigma^2(L_1 + a) = \sigma^2(L_1)$.

Additionally to the non coherence of the variance, S. Lambert et al. (2008) showed that the variance are **not elicitable**. Nonetheless, like every widely used risk measures, the variance is known to be **law-invariant**.

Overall, although being very common and widely used, one can see that the variance lacks many important properties to be a desirable measure of risk. This surely explains the rise of new risk measures.

2.3 Value-at-Risk

2.3.1 Definition

Value-at-Risk, or VaR, is a very important metric and is widely used in quantitative finance. It is computed with regard to three components: a certain time-frame (e.g. N days), a confidence interval $1 - \alpha$ and a given loss amount. It is very simple to interpret as it is the quantile of a N-days returns at which the area under the curve is equal to $1 - \alpha$ (Bommareddy et al., 2014). This corresponds to the predicted maximum loss with a specified probability level over a certain period of time (Cornuéjols and Tutuncu, 2007):

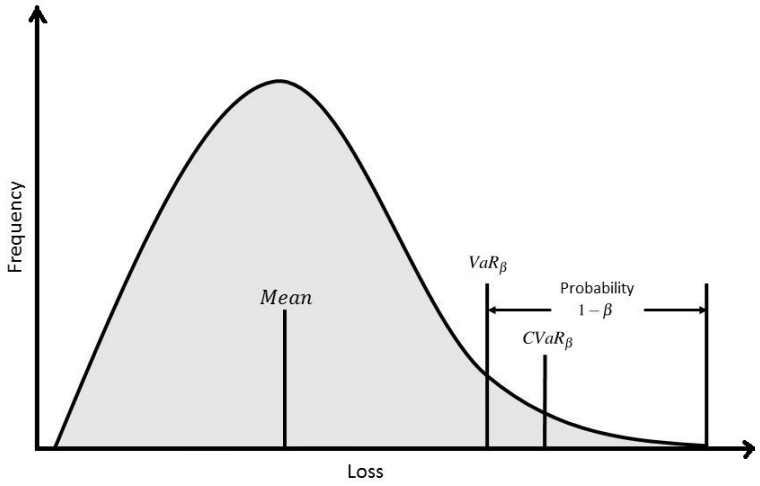

Figure 2.2: Illustration of the VaR and CVaR/ES

The expected shortfall here illustrated will be deeply explained subsequently.

Portfolios are generally regarded as multivariate random variables (RV). Therefore, they can be represented by a cumulative distribution function (CDF) $F_X : R \rightarrow [0,1]$:

$$F_X(x) := Pr(X \leqslant x) = p \tag{2.2}$$

If the function F is continuous and strictly monotonically increasing, the associated quantile function Q (proper to the VaR) returns the value x such that:

$$\begin{aligned} & Q(p) = x \quad \text{iff } p = F(x) \\ \Leftrightarrow\ & Q = F^{-1} \end{aligned}$$

Otherwise, if a manager wishes to incur the smallest loss l that does not exceed the loss L with a confidence level α, the associated VaR, denoted $VaR_\alpha(L)$, is the smallest number l such that the probability that the loss L exceeds l is not larger than $(1-\alpha)$[2] (McNeil et al., 2005, p.38):

$$\begin{aligned} VaR_\alpha(L) &= \inf\left\{l \in \mathbb{R} : \Pr\left(L > l\right) \leq 1-\alpha\right\} \\ &= \inf\left\{l \in \mathbb{R} : F_L(l) \geq \alpha\right\} \\ &= F_L^{-1}(\alpha) \end{aligned} \tag{2.3}$$

Whereas the definition of the VaR is simple and straightforward, one of the biggest challenges associated to it is how to represent the CDF F_L of the loss distribution. It is indeed expected to vary with regard to the market environment (bearish or bullish), the nature of the portfolio, etc. However the conceptual clarity and intuitive interpretability of the VaR have contributed to its great popularity.

2.3.2 Properties

It has been pointed out that the elicitability property is contingent on the existence of a scoring function for a given risk measure. Since it is possible to find a scoring function that is strictly consistent for the VaR (see definition 1.5.4), the VaR is **elicitable**. Indeed, for any $0 < \alpha < 1$, the statistical functional $\rho(F_L) := VaR_\alpha(L)$ is elicitable on the set of strictly increasing d.f. with finite mean. For example, the following scoring function is consistent for the VaR:

$$S_\alpha(x,y) = \left|\mathbb{1}_{\{y \geq l\}} - \alpha\right| \left|l - y\right|$$

and can be minimized by taking the derivative of the expectation of this expression (see McNeil et al. (2005), p357). This makes the VaR elicitable and desirable if estimation models wished to be compared using a scoring function.

Looking at the **coherence** property, as developed in section 1.5.1, the VaR appears to share only three fundamental axioms of coherence:

[2] In risk management, the loss distribution is mostly used following a convention that losses are positive numbers, what thus provide an immediate interpretation. We accordingly focus on the upper right tail of the **loss** distribution. For that reason, α is assumed to be the confidence level, such that $\alpha = 0.95$ or 0.99.

1. It is *positively homogeneous*: $VaR(\lambda L) = \lambda VaR(L)$. If the size of a portfolio increases, so the VaR does, in exactly the same multiplier.

2. It is *translation invariant*: $VaR(L + d) = VaR(L) + d$. Add a constant to a quantile add this quantile with exactly the constant.

3. The VaR is *monotone*: for $L_1 \leqslant L_2 \Rightarrow VaR(L_1) \leqslant VaR(L_2)$. The VaR is indeed related to the range the RVs take.

Therefore the VaR is not a coherent risk measure. Indeed it seems that the VaR is only *subadditive* in the idealized situation where all portfolios can be represented as linear combinations of the same set of underlying elliptically distributed risk factors (e.g. under the assumption of normality) (McNeil et al., 2005, p. 242). As soon as the portfolio shows skewness in its losses or when there are strong dependence between the products that constitute the portfolios, the axiom does not stand anymore. In other words, there exists a portfolio of financial instruments (X,Y) such that $VaR_\alpha(X + Y) > VaR_\alpha(X) + VaR_\alpha(Y)$. As a consequence, a decentralization of risk management would be compromised since one cannot be sure that the aggregated level of risk using VaR for different units will provide a bound for the overall risk (Emmer et al., 2015). A financial institution is likely to mis-estimate its risk exposure, and eventually hold an amount of economic capital that differ from the one desired (Danielsson and al., 2005).

On the other hand, an analysis conducted by Danielsson and al. (2005) has shown that the VaR is subadditive for most practical applications. They especially demonstrated that for fat-tailed distributions (tail index[3] < 1) (but not super fat tailed), the VaR is still subadditive in the tails, what denies the *absolute* non-coherence of the VaR in the tails. This suggests that an exclusion of the VaR from the set of relevant risk measures solely based on that criteria would be inappropriate.

Considering the **comonotonic additivity** of the VaR, McNeil et al. (2005) found that quantiles of the sum of comonotonic variables are perfectly additive. From definition 2.3 of the VaR, it can be shown that it is additive, i.e. $VaR(L_1 + L_2) = VaR(L_1) + VaR(L_2)$.

Moreover, the VaR is **law-invariant** since it is a strictly distribution-based risk measure. The risk measure estimation of the loss $\rho(L)$ is derived from the loss distribution F_L. It means that the sole knowledge about the probabilities distribution allows the estimation of the VaR.

Finally, while the VaR is a tail risk measure that focuses in the loss section of the distribution, it is rather **tail-risk** insensitive. It tells us the maximum loss L that would be incurred at the α confidence level, but it does not give any clue about the size of the loss within the remaining $(1 - \alpha)$. This may lead to an underestimation of the tail-risk in the view of subsection 1.5.5. The VaR is therefore unable to distinguish between portfolios embodying different levels of risk. For an intuitive visualization, see Figure 2.3:

[3]The tail index is the shape parameter of these heavy tailed distributions. The most popular estimator for the tail index of heavy tailed distributions is the Hill (1975) estimator.

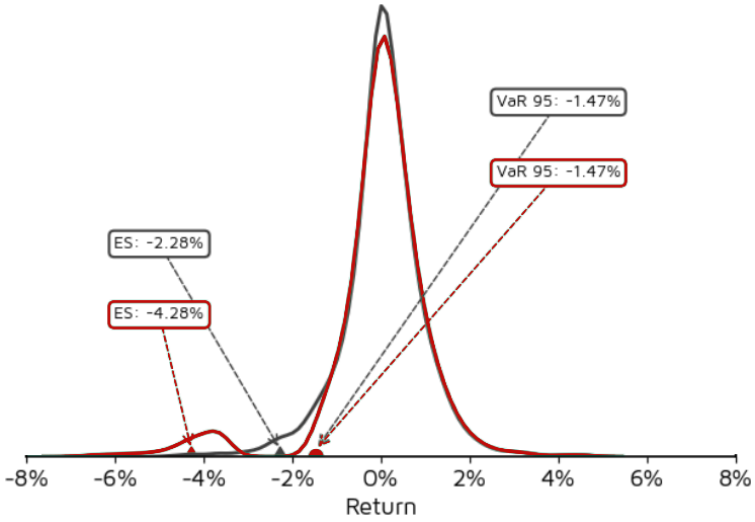

Figure 2.3: Levels of VaR (95% confidence level) and ES.

Overall, the commonly agreed **non-coherence of the VaR** and its inconsideration for the amplitude of the tail-risk have led to increasing interests towards the Expected Shortfall. It has been often mentioned as a great alternative to the VaR and has recently been enforced by regulators (for details, see section 2.6). The next section gives further insights about it.

2.4 The Expected Shortfall

2.4.1 Definition

The ES came to live after the VaR with the aim to construct a risk measure that would be aligned with the four axioms of coherence, although being closely related to the VaR. It looks at all the losses L such that they are bigger than a confidence interval given by the VaR, and takes the average of the amount that might be lost beyond that VaR (BCBS, 2019b). It results that an accurate estimation of the tail of the distribution is especially important for the effectiveness of ES. For a graphical representation of the ES, please refer to Figure 2.3.1. As the ES is an expectation over a threshold (which is the VaR at a level α), the ES can be defined as the conditional expected loss in case the VaR is exceeded (provided that F_L is integrable) :

$$\begin{aligned} ES_\alpha &= \frac{1}{1-\alpha}\mathbb{E}\Big(L; L \geq q_\alpha(L)\Big) \\ &= \mathbb{E}\Big(L | L \geq VaR_\alpha\Big) \end{aligned} \tag{2.4}$$

Hence, because of its definition, we always have $ES_\alpha \geq VaR_\alpha$. Moreover, as an expectation is given by an integral, the ES can be defined as follow:

Definition 2.4.1. *For a loss L (with finite expectation) and a CDF F_L, the expected shortfall at confidence level $\alpha \in (0,1)$ is defined as:*

$$\begin{aligned} ES_\alpha &= \frac{1}{1-\alpha}\int_\alpha^1 q_u(F_L)du \\ &= \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(L)du \end{aligned} \tag{2.5}$$

where $q_u(F_L)$ is the quantile function of F_L

2.4.2 Properties

First of all, the ES is maximized when the losses are **comonotonic**. This means that the worst risk for the aggregation of assets is when those are comonotonic.

Secondly, the ES is **law-invariant** for the same reasons as the VaR. It is a loss distribution-based risk measure and can be computed as a sole function applied on $F_L(l)$.

Thirdly, the ES takes the full **tail-risk** into consideration since it looks at the returns/losses further in the tail, beyond the VaR threshold. This allows the ES to take into account the full risk embedded in the loss distribution of the portfolio, which increases its risk sensitivity and its aptitude to account for extreme adverse events that may lead to significant losses. This characteristic of the expected shortfall makes it more conservative, what is desirable from the regulatory view point but might tighten the performance of banks by forcing them to rise the cushion of capital to pledge.

Then, the ES is **non-elicitable** (proven in Gneiting (2011, theorem 3.5)) since no scoring function from the ES can be drawn. This has challenged the ES functional as a predictive measure of risk. It has been even accordingly thought that the ES is not backtestable. Recently, however, Acerbi and Szekely (2014) claimed that elicitability has nothing to do with a traditional backtesting procedure (e.g. exceedance rate assessment discussed in chapter 7). Hence they argue that the belief under which the ES cannot be backtestable is erroneous, and show that the mathematical property of elicitability concerns model selection rather than model testing. Therefore, the ES can be backtested, although the backtesting procedure may be expected to be more arduous in comparison with the VaR backtesting. Besides, if the non-elicitability of the ES is not contestable, it has been proven that the ES is conditionally elicitable. In particular, thinking of the ES as comparable quantiles of the loss distribution, i.e. $E\Big(L|L \geq q(L)\Big)$, one can notice that it is elicitable conditionally on VaR. It results that focusing on one tail of the distribution only, Expected shortfall is assumed elicitable. The latter is an exception, but from interest as far as we are concerned.

At last, the ES fulfills all the four axioms of coherence in any circumstances such that it is a **coherent** risk measure (Danielsson and al., 2005). In particular the expected shortfall is subadditive, i.e.

$$ES_\alpha\Big(\sum_{i=1}^n L_i\Big) \leq \sum_{i=1}^n ES_\alpha(L_i)$$

Several proofs of the subadditivity of the ES have been drawn. We strongly refer to the paper of Embrechts and Wang (2015).

2.5 Comparison

2.5.1 The VaR and ES over the variance

The variance lacks many desirable properties, such as *coherence*, *comonotonic additivity*, and *elicitability*. The VaR and ES contain the considerable advantage of being interested in extreme events, and not only in the simple deviation around the mean, since they correspond to quantiles in the upper tail of the distribution. In the context of intense globalization, markets have never been so complex. Globalization has indeed increased the sensitivity of a market to the flow of information such that in the past few years, the (inter)connectivity of risks soared [4]. This is notably likely to induce clustering of volatility in the returns of financial series. This stylized fact can be observed in the historical series of the S&P 500.

Figure 2.4: Historical volatility of the S&P 500. *Source: Macrotrends: VIX Volatility Index - Historical Chart.*

One can notice that around the big crisis of 1987, 2000 and 2008 highlighted in grey on the chart, higher levels of volatility ensued. They lead to fatter end-tails in the distribution. Facing this, it is clear that models of risk based on the variance only would produce bias estimations and very poor insights for a reliable management of risk. In that respect, the VaR and ES were shown to be more suitable to deal with tail-risk.

Finally, the information conveyed by the variance (or standard deviation) is already taken into account by the VaR, hence ES. Its use would thus be redundant. Consequently, as opposed to the variance, the originality of the VaR and ES is entailed in their enlarged scope of measure-

[4] Dr. Peter Schenk, MEAG's Head of Investment Controlling

ment, not limited to a risk measure in terms of a deviation or volatility, but aimed to return an absolute measure of the potential loss, within a specified likelihood and time horizon. For this reason, we won't further address the variance in this comparative overview.

2.5.2 Tail-risk

The VaR may attribute the same risk level to two portfolios loss distribution though one depicts a greater riskiness, whereas the ES will return representative different values for the risk estimation. This is easy explained: the VaR is a quantile which cannot discern different risk behaviors of assets having two different tails behind the quantile. This is exposed in Figure 2.3. Moreover, for capital reserve determination, it seems to make sense to look at multiple quantiles instead of single levels (Kerkhof and Melenberg, 2003). By tackling the tail risk, the ES provides an answer to the *what if* question, crucial to determine the potential insolvency of a bank due to adverse market conditions. Answering that natural and legitimate question on the risk makes the ES a simple concept, at least as straightforward as the VaR (Acerbi and Tasche, 2002).

2.5.3 Coherence

The VaR is not coherent risk measure. Especially, because the VaR is a quantile on the loss distribution and not an expectation, so that the shape of the tail before and after the VaR probability does not have any bearing on the actual VaR number, it does not fulfill the sub-additivity property according to Artzner et al. (1999)'s definition (Danielsson and al., 2005). As opposed, the ES, yet computed as the average of those, is subadditive, i.e. coherent, which seems necessary for a good capital adequacy requirement in banking supervision (Acerbi and Tasche, 2002).

2.5.4 Elicitability and backtestability

The backtesting procedure is considered as a key tenant of the capital regulation for banks. It is indeed the process whereby the quality of the risk measure estimates can be assessed.

The backtesting of the ES is not as straightforward as for the VaR, notable characteristic of its lack of elicitability. The VaR backtesting simply counts the frequency of losses above the estimated VaR, while on the other hand, the ES backtesting needs to compare the average of realized losses beyond the VaR level with the predicted ES (Yamai and Yoshiba, 2002b). However, although it was commonly believed that the non-elicitability of the ES prevented its backtestability, we showed that the traditional backtesting, i.e. via an exceedance process for example, is also feasible for the expected shortfall. We especially pointed out that it is conditionally elicitable on the VaR, thereby easing a backtesting procedure for ES. In spite of that, the assessment of the goodness of the ES remains contingent on the finding of effective

backtesting methodologies. Notably, the difficulties encountered in the backtesting of the ES most likely slowed down its move into regulation.

2.5.5 Robustness and sensitivity

The ES estimates are known to be less robust than the VaR estimates (Yamai and Yoshiba, 2002a). Cont et al. (2010) conclude that the ES estimates appear to be largely affected with infrequent and large losses in the sample whereas the VaR estimates remain more insensitive to them. Therefore, as the loss distribution becomes fat-tailed, the ES estimates become more sensitive to infrequent large loss or outliers so that their estimation error increases larger than the estimation error of the VaR.

In rapidly changing market conditions, a risk measure characterized by a greater risk sensitivity in the tails, although it creates larger variations in the estimates, may be desirable for an financial institutions that seeks to prevent itself from harmful unexpected extreme events. Since the ES is more sensitive to new data, it might be more desirable than the VaR in such circumstances. That is actually part of the reasons that drove an introduction of the ES, as a remedy to the VaR (Emmer et al., 2015).

To become more robust, the ES needs a larger sample size than it is required for the VaR for the same level of significance (Yamai and Yoshiba, 2002a).

2.6 VaR and ES in regulation

It is common to use the VaR in the Banking sector through the Basel II pillar II that defined the economic capital as the 99%-VaR of the core equity tier I capital (CET1). However, following recent events, the VaR has been criticized about its, sometimes, weakness for the calculation of regulatory capital in banking. Notably, (1) the VaR does not say something about the *what if* question, unlike the ES, (2) it suffers from statistical and model uncertainty at high confidence level, and, as we said, (3) it may add up the wrong way (subadditivity) (Embrechts and al., 2014). As those "critics" have become more and more relevant, a slight move from VaR to ES is considered by the Basel committee. Indeed, although ES also suffers from (2), it help correct (1) and, assuredly, (3).

Then, the BCBS (2019b) requires a shift from 99%-VaR to 97.5%-ES as the reference risk measure in capital calculations. It specifies that the "revised internal models approach will eventually replaces VaR and stressed VaR with a single ES metric that is calibrated to a period of significant market stress", stressing the fact that it may address deficiencies spotted in the Basel 2.5 framework (BCBS, 2019b). This decision is clearly supported by our comparative analysis conducted in section 2.7. Hence, the importance of ES has grown in regulations.

2.7 Summary

Table 2.1 provides an overview of the considered risk measures and their properties, with respect to the framework of desirable properties that has been described in section 1.5, plus some additional relevant features.

Risk measure	Variance	Value-at-Risk	Expected shortfall
High tail-risk consideration			x
Coherent			x
Elicitability		x	
Backtestability		x	x
Interpretability [a]		x	x
Robustness to new data points		x	

Table 2.1: Properties of risk measures

[a] To be understood as the interpretability of the risk measure in the context of banking regulation to determine an amount of economic capital.

The latter table clearly states a strong dominance of the VaR and Expected shortfall against the variance. This shows that VaR and ES are more adequate measures to assess risk in theory. This particularly justifies the focus of this work, and the reason why VaR and ES have received the favor of the regulators and risk specialists in financial institutions so far.

Both VaR and ES have attractive properties. However, as shown by the coherence and tail-risk principle, the ES depicts better theoretical properties than the VaR, especially since its backtestability was demonstrated. Overall, this certainly explains why the ES is increasingly preferred as "risk functions", notably in regulation (for details, see section 2.6). The interesting properties of the VaR should not be disregarded either, such that one may be tempted to adopt a framework that relies on both risk measures as to benefit from their respective advantages (even though in practice a one model risk measurement procedure is generally preferred). Sharing this vision, Embrechts and al. (2014) pointed out "warnings concerning the near-impossibility of accurate estimation of single risk measures like VaR and ES at high confidence levels", as the "one number can suffice" still holds for the ES at extreme events.

Because of their properties and current uses, we decided to conduct our estimations on both VaR and ES. This makes even more sense since the VaR is still priorly needed to estimate the ES, and both are still mentioned in the new regulations. That last fact also explains why we do not treat the *expectiles*. However, finding a suitable risk measure that meets the desirable properties that a risk measure should have is extremely crucial, but not sufficient. Cont et al. (2010) highlight the need to consider the entire "risk measurement procedure" in order to find a good theoretical measure of risk that also embeds a good estimation procedure. This naturally leads to the question of the performance of the risk estimation based on these measures.

CHAPTER 3

THE DESIRABLE PROPERTIES OF THE RISK ESTIMATOR

3.1 Introduction

Once a risk measure ρ has been chosen, in practice one has to estimate the loss distribution of L from available data, then compute the risk measure using this distribution, thereby delivering the risk estimation. This leads to the notion of risk measurement procedure (RMP) (Cont et al., 2010) that integrates both steps. If M is a mapping function from the data set to the loss portfolio and ρ a mapping function from the loss portfolio to the risk,

Definition 3.1.1. *A RMP is a couple* (M, ρ)*, where* $\rho : \mathscr{L} \to \mathbb{R}$ *is a risk measure and* $M : L \to \mathscr{L}$ *is an estimator for the loss distribution. Therefore, the estimate* $\hat{\rho}$ *is given by* $\rho \circ M$*.*

The mapping function M, known as the estimator is defined as follow (Wackerly et al., 2008, p. 391):

Definition 3.1.2. *An estimator is a rule, often expressed as a formula, that tells how to calculate the value of an estimate based on the measurements contained in a sample.*

In order to have the best estimator as possible, one should try to follow procedures with regard to the different aspects that are presented in the next sections. These aspects consequently turn to be desirable properties of the risk estimator.

3.2 Statistical bias

Statistically speaking, having the right estimate on average is desirable. This means having an unbiased estimate of risk.

Definition 3.2.1. *$\hat{\rho_n}$ is statistically unbiased if:*

$$\mathbb{E}_\theta\left[\hat{\rho}_n\right] - \rho_\theta(L) = 0 \quad \forall\theta \in \Theta,$$

where $\mathbb{E}_\theta$ denotes the expectation operator under the probability measure $\Pr_\theta$, with $\theta \in \Theta$ the possible scenarios.

Let's take the Gaussian framework where the VaR is given by $\mu_L + \sigma_L\,\Phi(\alpha)$. The respective plug-in estimator is $\widehat{VaR}_\alpha^{\text{Gauss. plug-in}} := \hat{\mu}_L + \hat{\sigma}_L\,\Phi(\alpha)$ of the VaR in the Gaussian framework (call it $VaR^{\text{Gauss.}}$). From definition 3.2.1, it can be shown that this estimator of the VaR is biased. For $l_i \sim^{\text{iid}} (\mu_i, \sigma_i^2)$ with $L := (l_1, l_2, .., l_n)$.

$$\begin{aligned}
\mathbb{E}\left[\widehat{VaR}_\alpha^{\text{Gauss. plug-in}}\right] &= \mathbb{E}\left[\hat{\mu} + \hat{\sigma}\,\Phi(\alpha)\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} l_i\right] + \mathbb{E}\left[\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(l_i - \bar{l})^2}\;\Phi(\alpha)\right] \\
&= \mathbb{E}[l_i] + \Phi(\alpha)\mathbb{E}\left[\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(l_i - \bar{l})^2}\right]
\end{aligned}$$

where $\Phi(\alpha)$ is a constant, $\mathbb{E}[l_i] = \mu_L$, and $\mathbb{E}\left[\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(l_i - \bar{l})^2}\right] < \sigma_L$ since $\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(l_i - \bar{l})^2}$ is a biased estimator for the standard deviation. The square root is a concave function such that, from the Jensen's inequality, the square root of the sample variance is an underestimate. Therefore we have that

$$\mathbb{E}\left[\widehat{VaR}_\alpha^{\text{Gauss. plug-in}}\right] \leqslant VaR_\alpha^{\text{Gauss}}$$

This means that, provided the Gaussian framework, a sample that would estimate the true mean and variance would still underestimate the true VaR and ES on average. This will be discussed and solved in section 3.5.

3.3 Consistency

In addition to the unbiasedness, it is desirable to manipulate consistent risk measure estimators $\hat{\rho}$ of ρ. It is said to be consistent if it converges in probability to the true value of the parameter:

$$\lim_{n\to\infty} \Pr\left[|\hat{\rho}_n - \rho| > \epsilon\right] = 0 \tag{3.1}$$

for any $\epsilon > 0$.

For unbiased estimators, the consistency can be spotted by looking at their variance when n tends to infinity. If it is null, then the unbiased estimator is said consistent. The VaR and ES estimators that have been presented are all consistent. More particularly for the Gaussian framework, it will be shown in section 3.5 that, although being biased, the Gaussian plug-in VaR/ES are consistent to the true VaR/ES.

Note that it is utterly possible to design unbiased but inconsistent estimators, and conversely, biased but consistent estimators.
On the other hand, the concept of convergence in probability can be generalized by the concept of *convergence in measure*.

Definition 3.3.1. *Let $f, f_n \in \mathbb{N} : L \to \mathbb{R}$ be measurable functions on a measure space $(L, \mathscr{L}, \rho)$ (Anosov, 1994). The sequence f_n is said to converge (globally) in measure to f if, for every $\epsilon > 0$*

$$\lim_{n\to\infty} \rho\left(\left\{l \in L : |f(l) - f_n(l)| > \epsilon\right\}\right) = 0$$

In addition to the consistency per se, it is also important to consider the **rate of convergence** of the (consistent) estimator. Therefore, *ceteris paribus*, one should prefer estimators that quickly converge. It can be measured (Fu, 1975; Liu and Yang, 2012), and can be easily illustrated in simulations (see section 3.6).

3.4 Efficiency

We refer efficiency as the quality of an estimator. Generally speaking, an efficient estimator is one that estimates the quantity of interest in the best possible manner, notably by maximizing the use of materials, time and energy, e.g. the computation intensiveness or sample size, to reach a given level of performance. We here break down the efficiency term into two aspects: the *relative efficiency* and the *robustness of the estimation procedure*.

3.4.1 Relative efficiency

This property arises when one faces multiple unbiased estimates as candidates. When this occurs, one wishes to have an unbiased estimate that is *relatively more efficient* than the other ones. Formally, it is defined in terms of the variance of the estimate (Wackerly et al., 2008, p. 445):

Definition 3.4.1. *Given two unbiased estimators of risks $\hat{\rho}_1$ and $\hat{\rho}_2$ with variance $V(\hat{\rho}_1)$ and $V(\hat{\rho}_2)$, respectively, then the efficiency of $\hat{\rho}_1$ relative to $\hat{\rho}_2$, denoted by $eff(\hat{\rho}_1, \hat{\rho}_2)$ is defined to be the ratio*

$$eff(\hat{\rho}_1, \hat{\rho}_2) = \frac{V(\hat{\rho}_2)}{V(\hat{\rho}_1)} \tag{3.2}$$

If the estimators are all biased, we would choose the mean squared error (MSE) instead of the variance[1]. Therefore, $\hat{\rho}_1$ is relatively more efficient than $\hat{\rho}_2$ when the previous ratio is greater than 1.

[1]Note that the MSE can be defined in terms of variance and bias (Wackerly et al., 2008, p. 393). Indeed $MSE(\hat{\rho}) = V(\hat{\rho}) + \text{Bias}^2(\hat{\rho})$

3.4.2 Robustness of the estimation procedure

The efficiency property is closely related to the notion of robustness. Indeed, the efficiency also measures the degree to which the estimation is stable from sample to sample. That is, the less subject to sampling fluctuation a (risk) estimation is, the more efficient it will be (less variant).

Thus, the estimator for the portfolio's risk is said to be robust if small variations in the loss distribution result in small variations in the estimator (Cont et al., 2010). This is studied through sensitivity analysis.

3.5 The Economical bias

3.5.1 The problem at hand and its solution

Given the underestimation of risk in the Gaussian framework for a very simple case as the plug-in estimator, biasedness is crucial to be considered. It has however been usually disregarded by financial institutions or regulators (Pitera and Schmidt, 2018). Traditionally, one estimates the parameters of the statistic (typically via maximum of likelihood (MLE)), and then plug them in the estimator (an example is given in section 3.2). While the classical (statistical) definition of bias might be desirable from a theoretical point of view, it might be economically biased, which has not been prioritized by financial institutions or regulators (Pitera and Schmidt, 2018). In order to consider the importance of the economical sense of the risk measures, Pitera and Schmidt (2018) propose the notion of *economical bias*:

Definition 3.5.1. *The estimator $\hat{\rho_n}$ is called unbiased for $\rho(X)$, if*

$$\rho_\theta(X + \hat{\rho}_n) = 0 \quad \forall \theta \in \Theta$$

where X is a RV.

This can be interpreted the following way: for a risky position X, $\hat{\rho}_n$ is the estimated amount of capital to be added to X to make the *secured position* $X + \hat{\rho}_n$ acceptable under all possible scenarios $\theta \in \Theta$ with the risk measure ρ. Taking the regulator and risk manager views, we seek the value of $\hat{\rho}_n$ such that the position is *just* secured. This means that we wish to find to minimal amount of capital to be added to secure the risky position. Mathematically, any $\hat{\rho}$: $\hat{\rho} > \hat{\rho}_n$ or $\hat{\rho} < \hat{\rho}_n$ is not desirable.

Francioni and Herzog (2012) highlight the fact that the interplay between the model estimation and the model prediction for the VaR has never been investigated. The problem with the plug-in method is rooted in the the manner the distribution parameters (e.g. mean and standard deviation) are treated. They are commonly treated as deterministic values while $\hat{\theta}$ is rather a random variable. This introduces a bias which leads to a systematic underestimation of risk, at least in the Gaussian framework. This issue is tackled by Pitera and Schmidt (2018) for

the VaR and the ES in the Gaussian framework, and it is the matter of the section. It shows how the plug-in estimator of the VaR and ES are economically biased, which leads to systematic underestimation of risk. It also provides the rate of convergence of the plug-in estimators, as they are consistent to the true measure.

3.5.2 Economically unbiased VaR estimator in the Gaussian framework

As said earlier, the VaR is a quantile function (Q_α) that can be written as the inverse of the CDF of the P&L represented by X (this is just the opposite of L) for a level of confidence $(1-\alpha)$. Generally speaking, if $\theta := (\theta_1, \theta_2, ..)$ are the parameters of F, then

$$Q_\alpha = F_\theta^{-1}(\alpha) \Leftrightarrow F_\theta(Q_\alpha) = \alpha$$

The *plug-in method* estimates $\hat{Q}_\alpha := F_{\hat{\theta}}^{-1}(\alpha)$ where the parameters of the distribution are generally derived by MLE. Under the normality assumption of a portfolio distribution X with a mean μ and variance σ^2, the corresponding VaR at level α is given by:

$$VaR_\alpha^{\text{Gauss}}(X) = -\mu - \sigma\Phi^{-1}(\alpha) \tag{3.3}$$

By assuming an iid sample $\mathbf{X}:= X_1, ..., X_n$ at hand, the associated estimators for μ and σ are:

$$\begin{aligned} \hat{\mu}_X &:= \frac{1}{n}\sum_{i=1}^{n} X_i \\ \hat{\sigma}_X &:= \sqrt{\frac{1}{n-1}\sum_{i=1}(X_i - \hat{\mu}_X)^2} \end{aligned} \tag{3.4}$$

Then, by plugging the estimates in the VaR function, one obtains the following estimate:

$$\widehat{VaR}_\alpha^{\text{Gauss. plug-in}} := \hat{\mu}_X - \hat{\sigma}_X\Phi^{-1}(\alpha) \tag{3.5}$$

The goal pursued by Francioni and Herzog (2012) is to consider Q_α over F_θ $(\forall\theta)$ such that it also has this *threshold* property. In this regard, they came up with the notion of *probability unbiasedness*:

Definition 3.5.2. *The estimate of the quantile function $\hat{Q}_\alpha$ is probability unbiased for $X_i \overset{iid}{\sim} F_\theta$ if*

$$\mathbb{E}_\theta\left[F_\theta(\hat{Q}_\alpha)\right] = \alpha \tag{3.6}$$

For the VaR in particular, we want its estimate to be exceeded by the risk factor X with a probability α. The VaR is consequently considered as a specific statistical measure that needs to be treated as any statistical measure estimated from a finite amount of data.

Statistical unbiasedness is always desirable from a theoretical point of view. However, the average value of estimated capital reserve (i.e. $\mathbb{E}_\theta[\hat{\rho}_n]$) fails to determine the risk of the so-called *secured position* $(X + \hat{\rho}_n)$. One wants an estimator that behaves well within the backtesting phases given by the regulator. In the case of the VaR, the so called *exceedance rate* is considered. It will be deeply explained in Chapter 7 but note that this function is **non linear** while the average value of an estimated capital reserve is, by definition, linear. From this assessment, Pitera and Schmidt (2018) propose an estimator $\hat{\rho}_n$ for $\rho(X)$ that is *unbiased* in its economical sense, as referred to definition 3.5.1.
Using the definition of the VaR (as in equation 2.3), and just taking the returns instead of the losses, we have:

$$\rho_\theta(X) := VaR_{\alpha,\theta} = \inf\left\{x \in \mathrm{R} : \Pr_\theta(X > x) \leq \alpha\right\}$$

for any $\theta \in \Theta = \mathbb{R} \times \mathbb{R}_{>0}$ and a fixed $\alpha \in (0,1)$.
Accordingly, using definition 3.5.1, the VaR_α is economically unbiased if:

$$\rho_\theta(X + \hat{\rho}) = \inf\left\{x \in \mathbb{R} : \Pr_\theta(X + \hat{\rho} \leq x) \leq \alpha\right\} = 0 \quad \forall \theta \in \Theta$$

Because this is equal to 0, i.e. $x := 0$, the 'inf' operator can be removed, and the inequality term becomes equality. Therefore, the above expression is equivalent as saying:

$$\Pr_\theta(X + \hat{\rho} \leq 0) = \alpha \quad \forall \theta \in \Theta \tag{3.7}$$

In the Gaussian setting, $(\theta_1, \theta 2) := (\mu, \sigma^2)$, so $X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$. Accordingly, a possible unbiased estimator (in the economical sense) is

$$\hat{\rho} := -\hat{\mu}_X - \hat{\sigma}_X \sqrt{\frac{n+1}{n}}\, t^{-1}_{n-1}(\alpha) \tag{3.8}$$

where t^{-1}_{n-1} is the inverse function of a t-Student distribution with $n-1$ degrees of freedom. Let's prove this affirmation. The *unbiasedness* equation (3.7) becomes:

$$\begin{aligned} &\Pr\left(X - \hat{\mu}_X - \hat{\sigma}_X \sqrt{\frac{n+1}{n}}\, t^{-1}_{n-1}(\alpha) \leq 0\right) = \alpha \\ \Leftrightarrow &\Pr\left(\sqrt{\frac{n}{n+1}}\, \frac{X - \hat{\mu}_X}{\hat{\sigma}_X} \leq t^{-1}_{n-1}(\alpha)\right) = \alpha \end{aligned} \tag{3.9}$$

So, to be effectively unbiased, the left hand side term of the inequality in the Pr() (call it T) should be t-distributed with $n-1$ df.

$$\begin{aligned} T &:= \sqrt{\frac{n}{n+1}}\, \frac{X - \hat{\mu}_X}{\hat{\sigma}_X} \\ &= \frac{X - \hat{\mu}_X}{\sqrt{\frac{n+1}{n}}\, \sigma_X} \left(\sqrt{\frac{\frac{1}{\sigma^2_X} \sum_{i=1}^{n} (X_i - \hat{\mu}_X)^2}{n-1}} \right)^{-1} \end{aligned}$$

Yet, the Student distribution can be broken down wrt a normal and Chi-square distribution (Wackerly et al., 2008, p. 360):

Definition 3.5.3. $T := \frac{Y}{\sqrt{\frac{W}{v}}}$ *is t-distributed with v df (i.e.* $T \sim t_v$ *) if* $Y \sim \mathcal{N}(0,1)$ *and* $W \sim \chi^2_v$.

By definition, $X \sim \mathcal{N}(\mu, \sigma^2)$, and $\hat{\mu}_X \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, and it can be proven that X, $\hat{\mu}_X$ and $\hat{\sigma}_X$ are independent (Basu, 1955). The addition (and substraction) of normally distributed variables is normally distributed such that:

$$\begin{aligned} X - \hat{\mu}_X &\sim \mathcal{N}\Big(\mathrm{E}(X - \hat{\mu}_X), Var(X - \hat{\mu}_X)\Big) \\ &\sim \mathcal{N}\Big(0, \sigma_X^2 + \frac{\sigma_X^2}{n} - 0\Big) \\ \Leftrightarrow \frac{X - \hat{\mu}_X}{\sqrt{\sigma_X^2 + \frac{\sigma_X^2}{n}}} &\sim \mathcal{N}\Big(0,1\Big) \\ \Leftrightarrow \frac{X - \hat{\mu}_X}{\sqrt{\frac{n+1}{n}}\sigma_X} &\sim \mathcal{N}\Big(0,1\Big) \end{aligned}$$

In addition, by definition, $\frac{1}{\sigma_X^2}\sum_{i=1}^{n}(X_i - \hat{\mu}_X)^2$ follows a Chi-square distribution with $n-1$ df. According to definition 3.5.3, T effectively follows at t-Student distribution with $n-1$ df. It can consequently be used as a pivotal quantity, and equations 3.7 and 3.9 stand together:

$$\Pr_\theta\Big(X + \hat{\rho} < 0\Big) = \Pr_\theta\Big(T < t_{n-1}^{-1}(\alpha)\Big) = \alpha$$

This shows that this estimator for the VaR is economically unbiased when one assumes that X are iid normally distributed.

3.5.3 Economically unbiased ES estimator in the Gaussian framework

Pitera and Schmidt (2018) show the same type of flaw regarding the plug-in method for the ES. To do so, the same setting is used, i.e. $X \sim \mathcal{N}(\mu, \sigma^2)$, $(\theta := (\mu, \sigma))$. As given earlier, the general definition (see equation 2.4) of the ES (once again taking the P&L instead of the losses) is:

$$ES_\alpha = \mathbb{E}_\theta\big[-X | X < q_X(\theta, \alpha)\big]$$

while in the Gaussian framework, the ES is defined as (McNeil et al., 2005, p. 45) :

$$ES_\alpha^{\text{Gauss.}} = -\mu - \sigma\frac{\phi\big(\Phi^{-1}(\alpha)\big)}{\alpha} \tag{3.10}$$

The respective plug-in estimator of the ES is given by:

$$\widehat{ES}_\alpha^{\text{Gauss. plug-in}} = -\hat{\mu} - \hat{\sigma}\frac{\phi\big(\Phi^{-1}(\alpha)\big)}{\alpha} \tag{3.11}$$

Accordingly, Pitera and Schmidt (2018) propose an estimator based $\hat{\rho} := -\hat{\mu}_X - \hat{\sigma}_X\, a_n$ where a_n is found using the economical unbiasedness property as a constraint, that is :

$$\rho_\theta\left(X - \hat{\mu}_X - \hat{\sigma}_X\, a_n\right) = 0$$

This is done as follow: provided that the ES is positively homogeneous (see subsection 2.4.2), one can insert $\sigma\sqrt{\frac{n+1}{n}}$ to the denominator and numerator of the expression, and move the numerator expression outside of the risk measure. As such, the condition of unbiasedness can be rewritten to become

$$\sigma\sqrt{\frac{n+1}{n}}\,\rho_\theta\left(\frac{X-\hat{\mu}_X}{\sigma\sqrt{\frac{n+1}{n}}} - \frac{\hat{\sigma}_X\, a_n}{\sigma\sqrt{\frac{n+1}{n}}}\right) = 0$$

One can immediately see that $Z := \frac{X-\hat{\mu}_X}{\sigma\sqrt{\frac{n+1}{n}}} \sim \mathcal{N}(0,1)$ (as previously shown). On the other hand, the other term is closely related to a χ distribution. Indeed, $V_n := \sqrt{\frac{(n-1)\hat{\sigma}_X^2}{\sigma^2}} \sim \chi_{n-1}$. We consequently get

$$\sigma\sqrt{\frac{n+1}{n}}\,\rho_\theta\left(Z + V_n b_n\right) = 0$$

where $b_n := -a_n\sqrt{\frac{n}{(n+1)(n-1)}}$, and where Z and V_n are pivotal quantities. Then because $\sigma\sqrt{\frac{n+1}{n}}$ is deterministic, it can be removed from the condition, which ultimately leads to

$$\rho_\theta\left(Z + V_n b_n\right) = 0 \tag{3.12}$$

One can prove that b_n and then a_n exist using the *intermediate value theorem* and more particularly the *Bolzano's theorem* that states that if a function is continuous, has one value below and one value above zero, then the function necessarily passes by the x-axis.
Here, ρ_θ is assumed to be continuous. Moreover, when $b_n = 0$, then equation 3.12 becomes $\rho_\theta(Z)$, which is necessarily larger than 0 as the ES of a standard normally distributed RV is necessarily positive (see equation 3.3). Then, when $b_n \to \infty$, and by factorising b_n (this is still feasible because of the homogeneity of the ES measures), the condition from 3.12 is rewritten:

$$\lim_{b_n\to\infty}\; b_n \rho_\theta\left(\frac{Z}{b_n} + V_n\right)$$

Given the Lebesgue property of the ES, the above expression can be rewritten as

$$\lim_{b_n\to\infty} b_n\;\; \rho_\theta\left(\lim_{b_n\to\infty}\frac{Z}{b_n} + V_n\right) \tag{3.13}$$

And specifically

$$\rho_\theta\left(\lim_{b_n\to\infty}\frac{Z}{b_n} + V_n\right) \to \rho_\theta(0 + V_n)$$

By definition, $V_n \sim \chi_{n-1}$, which implies that the distribution is only defined over $\mathbb{R}^+$. This means that the amount of capital to add to make the position acceptable is actually negative, in other words that $\rho_\theta(V_n) < 0$. Therefore, for b_n large, a large positive value (b_n) is multiplied by a value closed to $\rho_\theta(V_n)$, which is negative. The ES is consequently negative when b_n is very large. This proves that b_n, and consequently a_n, exists such that the risk measure is economically unbiased.

3.6 A new property: the economical consistency

3.6.1 Case of the value-at-risk

It can be easily shown that the plug-in estimator is asymptotically unbiased in its economical sense. Indeed, for the VaR, one can see that the plug-in estimator (from equation 3.5) tends to the *economically unbiased* estimator (from equation 3.8) since a student distribution always converges to a normally distribution when its df (linked to n) tends to the infinity. In particular

$$\begin{aligned} & \sqrt{\frac{n+1}{n}}\, t_{n-1}^{-1}(\alpha) \xrightarrow{n \to +\infty} \Phi^{-1}(\alpha) \\ \Rightarrow\ & \hat{\mu}_X - \hat{\sigma}_X \sqrt{\frac{n+1}{n}}\, t_{n-1}^{-1}(\alpha) \xrightarrow{n \to +\infty} \hat{\mu}_X - \hat{\sigma}_X\, \Phi^{-1}(\alpha) \\ \Rightarrow\ & VaR^{\text{econ. unb.}} \xrightarrow{n \to +\infty} VaR^{\text{plug-in}} \end{aligned} \tag{3.14}$$

Hence we obtain that the larger the sample size, the closer the estimators are to each other, such that the plugged estimator matches its unbiased form (the bias of the plug-in estimator decreases), which is the definition of the *economical asymptotic unbiasedness*. Analogously to the statistical consistency, this could be called the *economical consistency* property. This can be written as:

Definition 3.6.1. *the risk estimator $\hat{\rho}_n$ is said to be asymptotically economically unbiased, or economically consistent iif*

$$\rho_\theta\big(X + \hat{\rho}_n\big) \xrightarrow{n \to +\infty} \rho_\theta\big(X + \rho_\theta(X)\big) = 0$$

The now called economical consistency of the Gaussian plug-in VaR to the true value of the VaR can be shown empirically. This is done with simulated data. Indeed, in Figure 3.1 and 3.2, one can first notice the average underestimation of the true $VaR_{1\%}^{\text{Gauss}}$ and $ES_{2.5\%}^{\text{Gauss}}$ and that the calibration errors ($\hat{\rho}_n - \rho$) converge suitably to 0 when the time period tends to infinity (i.e. when $n \to \infty$):

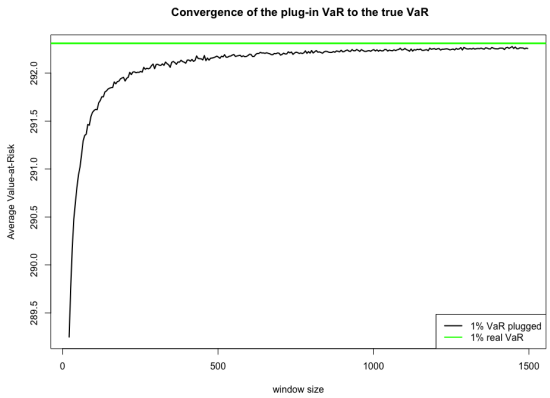

Figure 3.1: Illustration of the asymptotic convergence of the plug-in VaR to the real VaR with regard to the window size, that goes from 20 to 1500 observations by steps of 5. For each of them, 1,000,000 simulations were run.

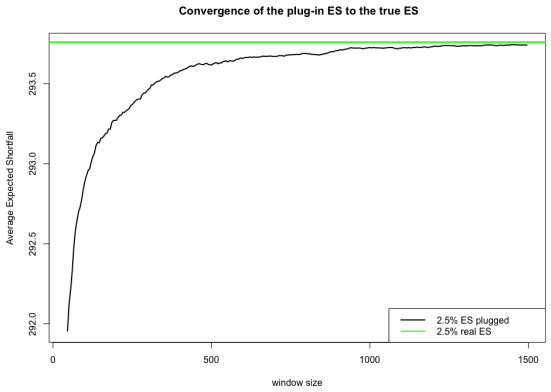

Figure 3.2: Illustration of the asymptotic convergence of the plug-in ES to the real ES with regard to the window size, that goes from 45 to 1500 observations by steps of 5. For each of them, 100,000 simulations were run.

In practice, the main goal of the risk manager is to define an estimator that behaves well in a backtesting or stress-calibrating procedure (Pitera and Schmidt, 2018). As said earlier in subsection 3.5.2, the exceedance rate procedure is considered and assesses the number of exceedances[2] between the prediction and the realized loss. Therefore, it can be used to measure the economical bias. If $\hat{\rho}_n$ is economically unbiased, one may expect that it will be violated as many times as the true value-at-risk. Formally formatted, one wants the probability of each being exceeded to be equal, i.e.

$$\Pr\left(\hat{\rho}_{n,t} < L_t\right) = \Pr\left(\rho_t < L_t\right) = \alpha \ \ \forall n$$

where n represents the time window over which the failure procedure is conducted, and α the significance level. Therefore, unbiasedness consists in a convergence of the exceedance rate towards the level α. In terms of the secured position achieved through capital adequacy, it means that the probability that the capital is not enough to secure the position X should tend to α (Pitera and Schmidt, 2018). This is represented by Equation 3.7.

In order to capt this behaviour, a simulation of this backtesting procedure has been conducted for each window size on which the VaRs are computed. We generated Gaussian observations from daily $\mu = 0.00019$ and $\sigma = 0.0126$. At each step, 350,000 scenarios are generated. Figure 3.3 shows the average exceedance rate with regard to the time window on which the VaRs are computed:

[2] In this thesis, we talk about exceedances, but other words are synonyms in this context, such as violations, breaches, hits or exceptions.

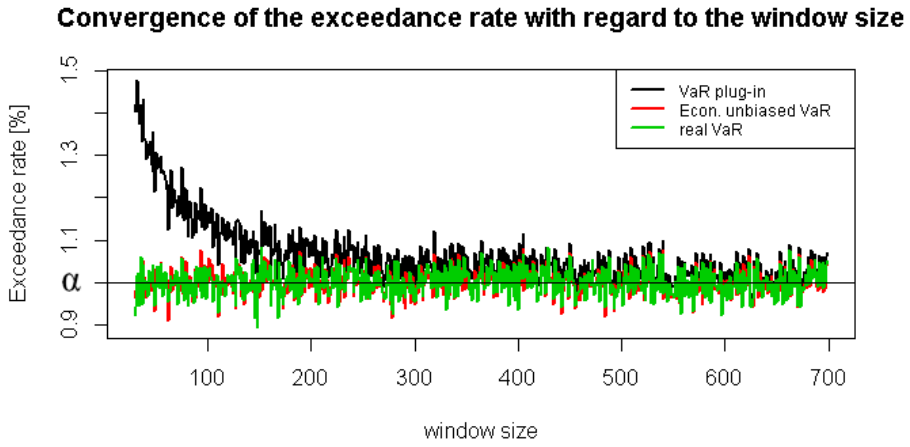

Figure 3.3: Average exceedance rate (over 350,000 simulations) with regard to an increasing time window size on which the plug-in VaR, the ec. unbiased VaR and the real VaR are computed ($\alpha = 1\%$).

This study shows that, while the economically unbiased estimator of the VaR provides efficient estimations regardless the window size, it takes quite a large time window size for the plug-in estimator to approach the true VaR performance (more or less 400 days). This supports the claim that the plug-in VaR produces reliable estimations only asymptotically. **Furthermore, the plug-in VaR underestimates the risk because the number of exceedances of the plug-in estimator is superior to the number of exceedances of the true VaR for typical value of N**. This should call to mind of the regulator because in the Basel context, the time window is 250 days, and the economical bias of the plug-in VaR remains substantial at this level.

3.6.2 Case of the expected shortfall

The ES was backtested using the exceedance rate procedure. The calculation of the unbiased ES according to $\hat{\rho} := \hat{\mu}_X - \hat{\sigma}_X \, a_n$ requires the prior determination of a_n. This can be done using numerical approximations of equation 3.12. Using the definition of ES (see definition 2.4), the economically unbiasedness condition can be written as:

$$\mathbb{E}\Big(Z + b_n V_n \big| Z + b_n V_n < Q_{Z+b_n V_n}(\alpha)\Big) = 0$$

where $Z \sim N(0,1)$ and $V_n \sim \chi_{n-1}$. Any expectation can be easily approximated by simulating a large number of times the respective distribution and taking the empirical expectation form. In this case, one needs to find b_n such that this conditional expectation is equal to zero. Then a_n is found. Moreover, we know that b_n must be positive such that the whole equation is null. As such, one can use the simple approach which test the resulting expectation for a given range of b_n. We did this with regard to the window size and obtained their respective a_n. A good

way to obtain good results is by extrapolating the points by drawing a trend line from the estimated points (as shown in blue in Figure 3.4):

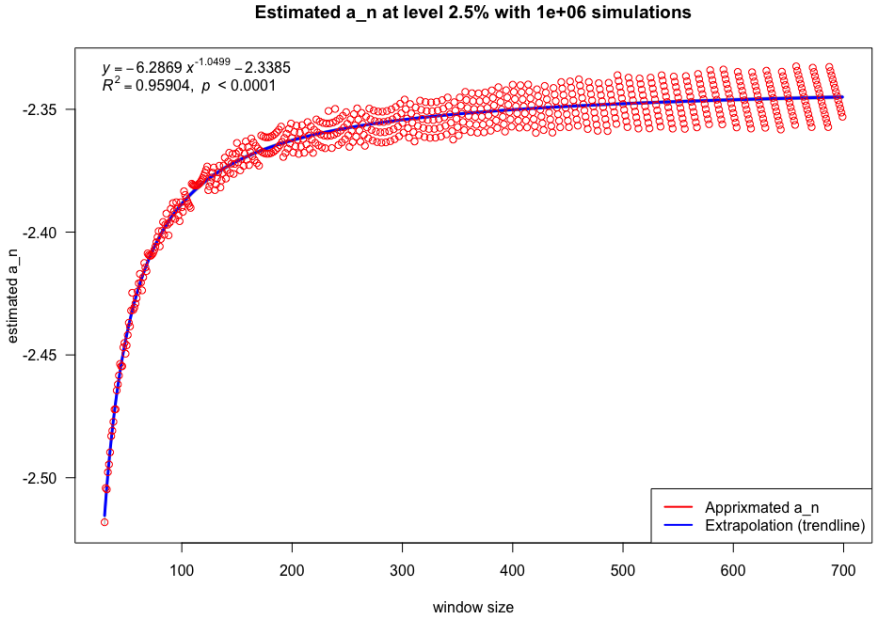

Figure 3.4: Approximated a_n derived from 1 million simulations for each window size (in red) and in blue the extrapolated a_n derived from a trend line (model 'power3P' in R)

Once the a_n coefficients have been derived, one can also study the frequency of exceedances of the ES and see if a similar behavior as in the VaR case is observed.

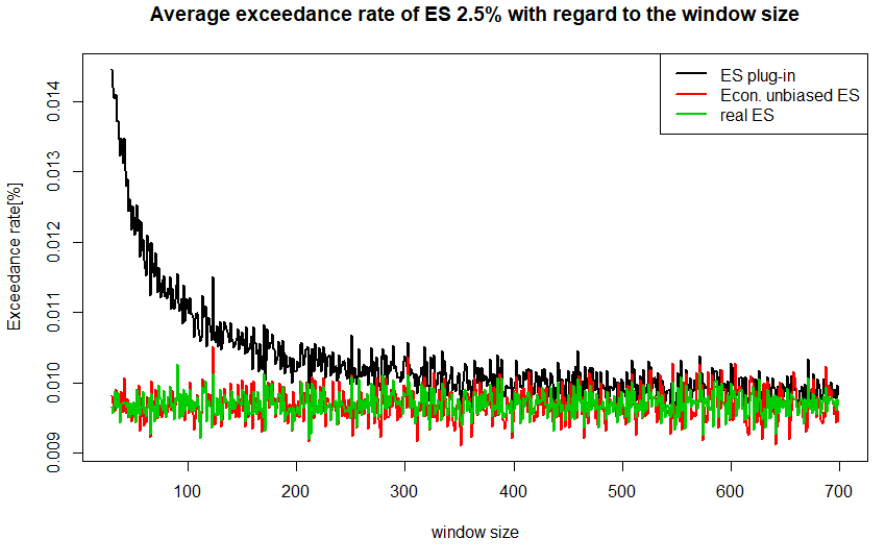

Figure 3.5: Average exceedance rate (over 350,000 simulations) with regard to an increasing time window size on which the plug-in ES, the ec. unbiased ES and the real ES are computed ($\alpha = 2.5\%$).

The same conclusion derived for the VaR can be transferred to the Expected shortfall. Indeed, the Gaussian plug-in ES estimator (red line) (see equation 3.11 for formula) can be also shown to be economically consistent, which is illustrated in Figure 3.5. It also depicts an underestimation of the risk at typical N level, and a low rate of converge. As opposed, the economically unbiased form of the ES estimator (blue line) outlines a rate of exceedances that directly aligns with the one of the true ES. It turns out that the plug-in ES requires a larger time window to achieve similar results. It should be also noted that the exceedance rate for the ES estimated at the level $\alpha = 2.5\%$ is close to the 1% significance level (for the true and economically unbiased ES) or tend to it (for the plug-in ES estimates). This supports the regulatory claim in favor of a 97.5^{th} percentile ES to replace the 99^{th} VaR.

3.7 Conclusion

In any condition, the statistical properties of the estimators (e.g. statistical bias, statistical consistency, etc) are desirable. Even, in the context of risk assessment, and especially in the view of the banking regulations, what is called *economical unbiasedness* should be prioritized. However, we have seen that (economical) bias may arise during the computation of the risk estimate. Based on our theoretical analysis, we highlight the fact that, since the Basel framework deals with a window of 250 historical data, the economical unbiasedness of the VaR and ES estimators proposed by Pitera and Schmidt (2018) should promote the consideration of the *economical unbiased property* as a very desirable characteristic of a risk estimator. Although the

study has been conducted in the Gaussian framework, this property should be (or at least should try to be) taken into account in the regulatory framework. We argue that the desirable economical properties should outweigh the desirable statistical properties in the context of capital adequacy determination.

CHAPTER 4

CONVENTIONAL METHODS OF ESTIMATION FOR THE VAR AND ES

4.1 Introduction

In this chapter, we explore the conventional methods that were previously introduced to estimate ρ using the VaR and ES. Although they may be considered too simplistic nowadays, they constitute the backbone of risk quantification. Three main approaches exist to compute them: non-parametric, semi-parametric, and parametric methods. Each of them presents various advantages and disadvantages (that will be addressed). Whereas non-parametric approaches do not set assumptions of a distribution for random variables, parametric approaches supposed that the returns follow of well-defined distribution. Semi-parametric approaches share features from both categories.

Additionally, we draw the attention of the reader on the fact that the level α will be now assumed to be the significance level (e.g. $\alpha = 0.05$ or 0.01)[1].

4.2 Non-parametric approach

The distribution of the P&L is not imposed a priori. Many approaches exist, such the Historical Simulation (HS), Bootstrapped Historical Simulation (BHS), Weighted Historical Simulation (WHS), and Filtered Historical Simulation (FHS). See details in P. Richardson et al. (1997)). However, we will only present the classical HS in order to grasp the intuition as well as the pro's and con's of the methods.

[1] and not the confidence level. This discrepancy comes from the fact that we now consider the P&L and not the loss distribution.

4.2.1 Historical Simulation

By only assuming that the returns are iid over the time window, it is possible to construct the empirical distribution function. Thereby, the VaR is estimated simply by taking the empirical quantile of the past returns:

$$\widehat{VaR_t}^{HS}(\alpha) = -\text{quantile}\left(\{x_i\}_{i=1}^N, \alpha\right) \tag{4.1}$$

This is a convergent estimator of the VaR for a coverage rate $\alpha\%$ since that by using the iid assumption, we have that

$$\widehat{VaR_t}^{HS}(\alpha) \underset{N\to+\infty}{\longrightarrow} VaR(\alpha)$$

This estimation is unconditional, which leads to the fact that the corresponding estimator will be relatively invariant to changes in the portfolio value. Hence, it is a robust method of estimation, what can be a big flaw, especially in the case of high volatility periods.

Regarding now the ES relative to the HS, it is the average of the losses corresponding to the quantiles beyond the VaR at the α-quantile. Formally, from definition 2.4, we have for the P&L the ES

$$\begin{aligned} ES_\alpha &= -\mathbb{E}(x_i : x_i \leq -VaR_\alpha) \\ &= -\frac{1}{\alpha}\mathbb{E}\left(x_i : x_i \leq q_\alpha\left(\{x_i\}_{i=1}^N\right)\right) \end{aligned} \tag{4.2}$$

where x_i is the i^{th} return from the empirical P&L distribution. In practice, the estimation is obtained by taking the average that are over the $-\widehat{VaR_t}^{HS}(\alpha)$ threshold:

$$\widehat{ES_t}^{HS}(\alpha) = \frac{1}{\sum_{i=1}^N I_t} \sum_{i=1}^N x_i I_t \tag{4.3}$$

where $I_t := \mathbb{1}_{\left\{x_t < -\widehat{VaR_t}^{HS}(\alpha)\right\}}$

This method presents the advantage to be very easy to use and it does not require any assumption regarding the returns distribution. This means that possible fat tails are accounted for (Berkowitz and M. O'Brien, 2001) and that it can be adapted to any type of portfolio (Jorion, 2000). However, the HS method can be easily altered by the sample size (Oanea and Gabriela, 2015) . In addition, it is desirable to have a number a extreme scenarios in the historical records to provide an informative estimates of the tail. According to McNeil et al. (2005), it is argued that the conditional approach is more relevant for a day-to-day market risk management. Moreover, HS calculation appears to be ineffective if the portfolio contains complex securities (Hammersley, 1964). Overall, it is rather adapted to small and simple portfolios.

4.3 Parametric approach

Unlike the non-parametric approach, there here is an assumption on the form of the P&L distribution.

4.3.1 Risk metrics

Developed by J.P. Morgan in the early 90's, it has been a very popular metric for popularizing the concept of VaR. It is based on the so called *Variance-Covariance method* or the *delta normal method*. Risk Metrics (RM) generally assumes that the changes in the market prices (e.g. returns) are normally distributed, i.e. $X_t \sim \mathcal{N}(\mu, \Sigma)$, where μ, Σ are the mean vector and covariance matrix of the portfolio returns, respectively. Their estimators have actually been already defined in equation 3.4 from section 3.5.2. This means that the instruments are assumed to have a profile of linear risk. The weights given to the historical returns are equal. However, as markets are often heteroskedastic, there is an incentive to give more weight to close observation than older ones. This can be done through an Exponential Weighted Moving Average (EWMA) process (Galdi et al., 2007), so that the forecast in t is a linear combination of past innovations and variances. In particular,the conditional variance in RM is represented as follow:

$$h_t = \lambda h_{t-1} + (1-\lambda) x_{t-1}^2 \tag{4.4}$$

where x_t the realized return in t and λ is a decay parameter. The latter determines the relative weights that are applied to the observations (returns) and the past level of volatility. It is generally set to 0.94 for a daily frequency (Galdi et al., 2007; Trenca et al., 2011). h_t is incrementally computed with h_0 being the empirical variance. Accordingly, the VaR and ES are computed in the same way as presented in section 3.5.2 except for the variance. For a coverage rate α, the VaR is given by

$$VaR_t^{\text{R.M}}(\alpha) = -\mu - \sqrt{h_t}\,\phi(\alpha) \tag{4.5}$$

and consequently the ES is as follow:

$$ES_t^{\text{R.M}}(\alpha) = -\mu - \sqrt{h_t}\,\frac{\phi\left(\Phi^{-1}(\alpha)\right)}{1-\alpha} \tag{4.6}$$

This approach has the advantage to be relatively quick to implement and only requires the portfolio covariance matrix. Also, since the variance is conditional to its past realization, the risk estimated using Risk metrics is reactive to changes in market conditions. But, it is unsuitable for portfolios that have fat-tailed distributions and consequently greatly minimize the resulting VaR. This problem can be somehow overcome by introducing the so-called GARCH models. Moreover, because of the linear approximations, it often fails to model complex products, such derivatives and credit derivatives (Sollis, 2009).

4.3.2 GARCH processes

In the univariate case, the GARCH models are used to predict the conditional variances of the P&L distributions, from which the VaR is derived. It comes under several assumptions in terms of the distribution of the P&L. Under the normality assumption of the conditional distribution of the P&L, the corresponding VaR forecast associated to a coverage rate α is:

$$VaR_{t|t-1}(\alpha) = -\mu - \sqrt{h_t}\phi^{-1}(\alpha) \tag{4.7}$$

where h_t represents the conditional variance of the returns. To predict it, the GARCH model first assumes that the square of the perturbations follows an autoregressive process of order q. For example, let's say that X_t is a process that satisfies an ARCH(1) representation:

$$\begin{aligned} X_t &= \mu + \epsilon_t \\ \epsilon_t &= z_t\sqrt{h_t} \\ h_t &= \alpha_0 + \alpha_1\epsilon_{t-1}^2 \end{aligned} \tag{4.8}$$

$\alpha_0 > 0$ and $0 < \alpha_1 \leqslant 0$, and with z a white noise normally distributed (which means that $\mathbb{E}(z_t) = 0$ and $\mathbb{E}(z_t^2) = \sigma_z^2$).
The GARCH processes are Generalized ARCH processes in the sense it introduces P autoregressive terms in addition the shock ϵ_{t-1} to make the conditional variance h_t depend on its own past shocks. All in all, the GARCH(P,Q) is represented in the following way :

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i\epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} \tag{4.9}$$

such that $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$ for reasons of stationarity. The choice of numbers of the lags is of capital importance. The different models, i.e. with different (p,q) lag structures are compared using the AIC indicator since it has good properties for comparing forecast models (Alberg et al., 2008). Notably, the GARCH(1,1) often appears to be the best GARCH to model the behaviour of assets (McNeil et al., 2005). However, even if the tails of distributions are heavier than the normal distribution, they might remain too short (even with student-t innovations) (Tsay, 2010).

GARCH processes have the interesting property of being able to model volatility clustering (i.e. serial correlation) that often characterized financial time series (Sewell, 2011). However, the GARCH model assumes symmetry in the way shocks are accounted while it is known that negative shocks have a larger effect on the volatility than positive shocks (Wong et al., 2016). On the other hand, GARCH models are quite flexible in the sense that they can be made more complex according to the market conditions. For example, the Exponential GARCH (EGARCH) to capt asymmetry in the volatility of returns (Bucevska, 2012), the Integrated GARCH (IGARCH) to deal with non stationarity of returns (Mikosch and Stărică, 2004), the GARCH in the Mean (GARCH-M) to link level the volatility to the level of returns (Duță, 2018), etc. (implementations given in Sewell (2011)).

Overall, the GARCH models allow to design very flexible implementation of the volatility. Accordingly, this flexibility can be somehow transferred to the modelling of the VaR and ES obtained through previous described methods. On the other hand, GARCH processes for the conditional variance often add much variability, which is translated to the VaR and ES. In terms of capital to be added or removed, this variability is not enviable from this capital holder.

4.3.3 Monte-Carlo methods

This method consists in simulating a large number of times possible behaviors of the risk factors according to provided hypotheses. Accordingly, it is possible to derive a distribution of the P&L from which one can estimate a quantile using the same approach as in the historical methodology. To do so, it is required to consider a time horizon T broken down into increments Δt in order to discretize the continuous behaviour of the market. Therefore, the increment should be small enough to retrieve a proper estimation. Once done, one has to assume a distribution in a generator that will produce the simulated returns of the portfolio. For example, a typical model consists in considering the returns with a random part that follows a Brownian Motion W. It means that W_t is continued with independent increments such that $W_t - W_s \sim \mathcal{N}(0, t-s)$ for $(0 \leq s < t)$ and $W_0 = 0$, More specifically, in the standard **Monte-Carlo** (MC) approach, the return X_t is represented at time t by:

$$X_t = \frac{P_t - P_{t-1}}{P_{t-1}} := \mu \Delta t + \sigma \Delta W \tag{4.10}$$

P_t is the price at time t, μ is the average return, and σ is its standard deviation. From it, it is possible to derive paths for prices by simulating them. By doing so a large number of times, one can extract the quantiles from and derive a corresponding VaR and ES using the historical simulation methodology.

This approach allows to model complex evolutions of assets (e.g. derivatives) and is rather insensitive to extreme events. It is also extensible, e.g. by adding a jump size parameter λ in order to form what we would call the **Monte-Carlo jump** process. Paired with a count process dq that follows a Poisson process (Duffie and Pan, 2001), it turns equation 4.10 to:

$$X_t = \frac{P_t - P_{t-1}}{P_{t-1}} := \mu \Delta t + \sigma \Delta W + \lambda \, dq \tag{4.11}$$

In this work, the jump intensity is fixed as the lowest log return of the series (the Poisson's parameter), associated with a daily probability of occurrence of 1%. Adding a jump diffusion process allows to take into account risk of crises with large negative returns. They are simulated to inflate the tails and increase adverse events sensitivity.

The MC method is enviable since it is quite flexible according to what one thinks about the behaviour of the assets in the portfolio, regardless the presumed complexity of the portfolio (Sollis, 2009). The assets are not necessarily assumed to be normally distributed, as in the Risk

metrics. However, it has the disadvantage to be computationally expensive, which is worsened by the so called *curse of dimensionality* when the portfolio becomes very diversified (Pritsker, 1997). Regarding the VaR and ES per se, they suffer from bias caused by sample inefficiency and by inevitable numerical modifications (Yamai and Yoshiba, 2002a).

4.4 Discussion

While those techniques for the VaR and ES have been widely spread to estimate risk, they suffer from certain shortcomings that may become very costly. In particular, random variables might contain *extreme risks*, which is defined as the risk of extreme events occurring at a low probability level. Extreme risk is difficult to assess because of the common lack of data (By et al., 2011). Indeed, there are events that have not yet happened or have happened only very rarely. Therefore, standard statistical methods are sometimes inapplicable. For example, normal distributions are commonly used in these modeling techniques to describe reality although, in many cases, reality is closer to a fatter tailed distribution than a normal one (Bommareddy et al., 2014). This critical assumption of normality results in potentially inaccurate models, which can lead to tragic consequences (see The Black Swan theory in Taleb (2007)). Another important drawback of the conventional methods arises when it comes at describing the joint behaviour of the assets, resulting in the behaviour of the portfolio. It has been shown that the financial world is more and more interconnected, especially for banks (Moghadam and Vinals, 2010). Besides, this dependence increases in bearish markets (Hochrainer-Stigler et al., 2018). In such conditions, the correlation often fails to describe these dependence behaviours (four important reasons given in Rachev et al. (2009)). The regulator is particularly interested in risk measures that tackles these observed drawbacks (BCBS, 2019b).

Therefore, an extent to those conventional methods will be discussed in the following chapters of this work. Namely the **Extreme Value Theory** (EVT) to deal with very extreme events and **Copula** to deal with the interdependence structure of the assets in a portfolio.

CHAPTER 5

THE METHOD OF COPULAS

5.1 Introduction

As part of the RMP (as defined in 3.1.1), an accurate model of the portfolio loss distribution, on which the risk measure is estimated, is key for a financial institution willing to measure its market risk exposure. The method of Copula addresses the estimation of the loss distribution. They help model the dependence between individual risk factors, which would lead to a proper model of their joint distribution. In the previous methodologies, the loss or P&L distributions were only described in one dimension by taking weighted average of assets that compose the portfolio. In contrast, the copula allow to describe the portfolio joint distribution, aimed to reflect better the reality. Generally speaking, the knowledge of the joint distribution of random variables is sufficient to derive the marginals and the correlation structure among the random variables, but not vice versa (X. Li, 1999). To overcome that problem, the techniques of copula can be used. Introduced by (Sklar, 1959), the copulas allow to split the a (multivariate) CDF into two parts: (i) its margins $F_1, F_2, .., F_n$ (ii) its dependence structure C. It consequently enables to freely select that marginal distributions for the components of an *unknown* multivariate system (Ang and Chen, 2002).

This section is dedicated to the description of the best known copulas and how to estimate the VaR and ES from. Since the copulas are at the core of the thesis and constitutes an important sophistication to the risk estimation, the theoretical background provided is detailed.

5.2 Definition

Definition 5.2.1. *A copula function is a function that links or marries univariate marginals to their full multivariate distribution.*

For n uniform random variables $U_1, U_2, ..., U_n$, the joint distribution function C defined as

$$C(u_1, u_2, ..., u_n; \rho) = Pr\Big(U_1 \leq u_1, U_2 \leq u_2, ..., U_n \leq u_n\Big) \tag{5.1}$$

is the *copula function*, where ρ is the parameter of the copula and captures the information about the dependence of the variables of interest. The copula can be used to link the marginal distributions. Indeed, let's define the marginals $F_1, F_2, ..., F_n$ distribution functions we are interested in joining together. Then, using equation 5.1, we have:

$$\begin{aligned} C\Big(F_1(x_1), F_2(x_2), ..., F_n(x_n); \rho\Big) &= \Pr\Big(U_1 \leq F_1(x_1), U_2 \leq F_2(x_2), ..., U_n \leq F_n(x_n)\Big) \\ &= \Pr\Big(F_1^{-1}(U_1) \leq x_1, F_2^{-1}(U_2) \leq x_2, ..., F_n^{-1}(U_n) \leq x_n\Big) \end{aligned} \tag{5.2}$$

It can be shown via the Probability Integral Transform (PIT) that for any random variable x_i, $F_i(x_i) = U_i$ (it follows a standard uniform function, i.e. $F_i(x_i) \sim \text{U}[0,1]$). Accordingly, applying the respective inverse function F_i^{-1}, the equality becomes $F_i^{-1}(U_i) = x_i$. This can be applied to equation 5.2 and gives:

$$\begin{aligned} C\Big(F_1(x_1), F_2(x_2), ..., F_n(x_n); \rho\Big) &= \Pr\Big(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n\Big) \\ &= F(x_1, x_2, ..., x_n) \end{aligned} \tag{5.3}$$

From that result, the marginal distributions of X_i, $F_i(x_i)$, can be obtained.

(Sklar, 1959) showed that the latter equality can be used in the reverse way. That is, by knowing the marginals $F_1(x_1), F_2(x_2), ..., F_n(x_n)$, it exists a copula function $C(u_1, u_2, ..., u_n)$ to produce a proper multivariate distribution function. Sklar's theorem is general as it allows any joint distribution to be written in copula form. It follows that it is possible to describe a joint probability distribution function (PDF) from two components: the knowledge of the univariate marginal distribution functions (d.f.)'s and the copula to link them. Accordingly, their respective PDF can be written:

$$f\Big(x_1, x_2, ..., x_N\Big) = c\Big(F_1(x_1), F_2(x_2), F_N(x_N)\Big) \cdot \Pi_{i=1}^{N} f_i(x_i) \tag{5.4}$$

where c is the PDF of the copula.

5.3 Important properties of the Copula

$C : [0,1]^n \rightarrow [0,1]$ is a n-dimensional copula if:

- the copula is zero when at least one of its arguments in null: $C(u_1, ..., u_{i-1}, O, u_{i+1}, ..., u_n) = 0$
- the copula is u if each argument of C is equal to 1 except one: $C(1, ..., 1, u, 1, ...1) = u$
- C is n-non-decreasing

The *Fréchet-Hoeffding Theorem* states that for any Copula $C : [0,1]^n \to [0,1]$ and any $(u_1, ..., u_n) \in [0,1]^n$, the copula is bounded :

$$\max\left\{1 - n + \sum_{i=1}^{n} u_i, 0\right\} \leq C(u_1, ..., u_n) \leq \min(u_1, ..., u_n) \tag{5.5}$$

These bounds correspond to cases of extreme of dependency, i.e. comonotonicity (upper bound) and countermonotonicity (Haugh, 2016). These bounds are very useful in the sense that they can provide bounds to the VaR and ES (Rüschendorf, 2017)).

5.4 Dependence structure

As for now, a very important remark relates to the dependence parameter ρ that is introduced in the Copula definition (see equation 5.2 for example). The linear *Pearson* correlation coefficient ρ, which depends on the marginal distributions, is known to be inappropriate for financial series (Rachev et al., 2009). It only deals with linear dependence characteristic of elliptical distributions, although returns in financial markets often outline non-linear dependence (Aas, 2004). Hence, it is likely to induce misleading conclusion. Moreover, it requires finite variance while in thick tails of distributions leads the variance to tend to infinity. Most importantly, the correlation only give a degree of (linear) dependence but does not give its structure (Rachev et al., 2009). The copula-based approach aims to tackle this issue.

5.4.1 Kendall's tau and Spearman's rho

As a result, the *Kendall's tau* or *Spearman's rho* appear to be more suitable as they can be expressed in terms of the underlying copula alone (see Aas (2004) for details). In particular they are able to measure the *comonotonicity* as per definition 1.5.1. The Kendall's tau (τ) is the probability of comonotonicity minus the probability of countermonotonicity (opposed to comonotonicity) for a pair $(x_i, y_i), (x_j, y_j)$ of observations. Analogically, for independent random vectors $(X_1, Y_1), (X_2, Y_2)$ with a common joint distribution, the popular version of the Kendall's tau is:

$$\tau = \Pr\Big((X_1 - X_2)(Y_1 - Y_2) > 0\Big) - \Pr\Big((X_1 - X_2)(Y_1 - Y_2) < 0\Big) \tag{5.6}$$

Its empirical estimate $\hat{\tau}$ is

$$\begin{aligned} \hat{\tau}(X_1, X_2) &= \frac{\cdots}{\cdots} \\ &= \frac{4 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I\big((x_{1,i} - x_{1,j})(x_{2,i} - x_{2,j}) > 0\big) - n(n-1)}{n(n+1)} \end{aligned} \tag{5.7}$$

On the other hand, the Spearman's rho between two variables corresponds to the Pearson correlation between the rank values of those two variables:

$$\rho_S = \frac{S_{R_x,R_y}}{S_{R_x} \times S_{R_y}} \tag{5.8}$$

Whereas linear correlation measures the degree of linear dependence only, both τ and ρ_S are considered as measures of the degree of monotonic dependence between random variables (whether linear or not). They are therefore a more general concept than the linear correlation, which further emphasizes their use (Embrechts et al., 1999).

5.4.2 The tail dependence

The tail dependence is known as a model for the dependence structure that reflects more detailed knowledge of the risk management problem since it deals with the tails of distribution specifically. It ensues from the saying that, in time of market stresses, the correlations increase (Aas, 2004). Aware of this, Embrechts and McNeil (2000) introduced the so-called *coefficient of tail dependence*:

Definition 5.4.1. *For $X \sim F_X$ and $Y \sim F_Y$, the upper and lower tail dependence coefficient are respectively given by*

$$\lambda_u(X,Y) = \lim_{\alpha \to 1} \Pr\left(Y > F_Y^{-1}(\alpha) | X > F_X^{-1}(\alpha)\right)$$

$$\lambda_l(X,Y) = \lim_{\alpha \to 0} \Pr\left(Y \leq F_Y^{-1}(\alpha) | X \leq F_X^{-1}(\alpha)\right)$$

$\lambda_u(X,Y)$ represents the probability to observe a large Y provided that X is large. The higher $\lambda_u(X,Y)$, the more likely are large events to occur simultaneously. In contrast, Similarly, if $\lambda_l(X,Y)$ is high, low events tend to occur simultaneously. Note that when X and Y are independent, $\lambda_l(X,Y) = \lambda_u(X,Y) = 0$.

5.5 Implicit copula

Implicit copula is a category of copula that do not have an explicit expression for the copula function C. It is the dependence structure of a parametric multivariate distribution that implies the copulas. Typically, it refers to the elliptical copulas (e.g. the Gaussian and Student t-copula).

5.5.1 The Gaussian Copula

A very common example of copula is the Gaussian Copula. It was used to price and manage the risk of the CDO's before the financial crisis (Hull et al., 2005). It is defined as:

$$C_R^{\text{Gauss}}(u_1,...,u_n) = \phi_R\left(\phi^{-1}(u_1),\phi^{-1}(u_2),...,\phi^{-1}(u_n)\right)$$
$$c(u_1,u_2,...,u_n) = \frac{1}{\sqrt{det(R)}}\exp\Big(\frac{-1}{2}\begin{pmatrix}\phi^{-1}(u_1)\\ ..\\ \phi^{-1}(u_N)\end{pmatrix}^T(\frac{1}{R}-\mathbb{I})\begin{pmatrix}\phi^{-1}(u_1)\\ ..\\ \phi^{-1}(u_n)\end{pmatrix}\Big) \tag{5.9}$$

where R is the correlation matrix. Regarding the tail dependence coefficients λ_l^G and λ_u^G, they are null. It means that regardless the correlation ρ that is chosen, the extreme events should occur independently in X and Y. This suggests that the Gaussian copula does not have tail dependence (Malevergne and Sornette, 2003), which is often poorly representative of the behavior of financial markets, as shown by the 2008 crisis. This brings doubt about the use of the Gaussian copula in risk management of tail events.

The relation between the Kendall τ of two random variables X,Y and the linear correlation coefficient of Pearson is given as follow (McNeil et al., 2005, p. 215)

$$\tau(X,Y) = \frac{2}{\pi}\arcsin\big(\rho(X,Y)\big) \tag{5.10}$$

5.5.2 The Student t-Copula

In contrast with the Gaussian copula, the Student t copula depicts a positive tail dependence. And, since the t copula is symmetric, the dependence can be observed in both upper and lower tails of the distribution, i.e. λ_l^G and $\lambda_u^G > 0$. The t copula takes the form

$$C_{v,R}^t(u_1,...,u_n) = t_{v,R}\big(t_v^{-1}(u_1)+t_v^{-1}(u_2)+...+t_v^{-1}(u_n)\big) \tag{5.11}$$

where R is the correlation matrix and v accounts for the degree of freedom of the student t distribution. The higher v, the closer the t distribution to the Gaussian. Therefore, as v decreases and approaches 1, the stronger the tail dependence and the more inflated the tails. As far as we are concerned, this is a desirable feature.

The Kendall τ and the parameter of correlation ρ follow the same relation as for the Gaussian copula. Indeed, the parameter ρ is the same for all elliptical distributions.

5.6 Explicit copula

As the name says it, these kind of copula exhibit an explicit expression of the copula function. It is generally not the dependence structure of a parametric multivariate distribution. The

typical explicit copula refers to the so called *Archimedean copulas* whose the form is:

$$C_{1,\dots,n} = \psi^{-1}\big(\psi(u_1) + \dots + \psi(u_n)\big)$$

where $\psi : [0,1] \times \Theta \to [0,\infty)$ is the generator function of the copula (Θ is the parameter space) with ψ n-monotone on $[0,\infty)$. *Archimedean copulas*'s major advantage is to allow to model the dependence in arbitrarily high dimensions with only one parameter that governs the strength of dependence.

5.6.1 Clayton Copula

The Clayton copula is an asymmetric Archimedean copula, exhibiting greater dependence in the negative tail than in the positive. In terms of tail dependence coefficients, we have that $\lambda_l^G > 0$ and $\lambda_u^G = 0$. The generator of the Clayton copula is given by $\frac{1}{\theta}(t^{-\theta} - 1)$ and the copula function as follow

$$C_\theta^c(u_1,\dots,u_n) = (u_1^{-\theta} + \dots + u_1^{-\theta} - 1)^{-\frac{1}{\theta}} \tag{5.12}$$

where the parameter θ, $0 < \theta < \infty$, measures the strength of the dependence. Thus the Clayton copula interpolates between independence and perfect dependence. In the limit where $\theta \to 0$ we approach the independence copula. Equation 5.13 depicts the relation between the Kendall τ of two random variables X,Y and the Clayton copula's parameter (Doğan and Karakaş, 2017)

$$\tau(X,Y) = \frac{\theta}{\theta + 2} \tag{5.13}$$

Since the Clayton copula is specially dedicated to process lower tail dependence, its efficiency to capture joint extreme events in the lower tail of the P&L distribution is greater. Indeed, intuitively, as the tail dependence increases, the probability of simultaneous failure of financial instruments goes up, thereby increasing the magnitude of the loss expected. This feature fosters a strong case towards the relevance of this copula (Shamiri et al., 2011).

5.6.2 Gumbel Copula

The Gumbel copula, like the Clayton one, is an asymmetric Archimedean copula. It has one parameter θ with the generator $\psi(t;\theta) := \big(-\log(t)\big)^\theta$ so that the copula function becomes:

$$C_\theta^g(u_1,\dots,u_n) = e^{-\left((-\log u_1)^\theta + \dots + (-\log u_n)^\theta\right)^{\frac{1}{\theta}}} \tag{5.14}$$

where $1 < \theta < \infty$. As θ increases, so does the dependence in the Gumbel copula C^g. The relation between the Kendall τ and the Gumbel copula's parameter is as follow (Doğan and Karakaş, 2017)

$$\tau(X,Y) = \frac{\theta - 1}{\theta} \tag{5.15}$$

The Gumbel copula causes the distribution to have upper tail dependence $\lambda_u^G > 0$, whereas $\lambda_l^G = 0$. In the context of this work, since we are interested in losses occurring in left tail of the P&L distribution, it is already expected to find less relevance for this copula. However, it is shown to be sometimes useful, as in Chai and Guo (2011).

5.7 Estimation of the VaR and ES

The VaR and ES are estimated by extracting quantiles on the joint P&L distribution constructed using the copula function. This demands some prerequisites. The first step consists in selecting the marginal distributions of the random variables that constitute the portfolio, then to make use of the appropriate copulas in order to build the joint distribution. To do so, the margins parameters and the copula parameters are estimated and then assessed (e.g. least square fit, maximum likelihood,..). Typically, the copula parameter can be estimated using the following procedure:

1. Compute an empirical estimate of the Kendall's *tau* $\hat{\tau}$ (e.g. refer to equation 5.7). It ensures that the pre-existing dependence structure between the marginals is captured.
2. Use the relationship between the Kendall's *tau* and the copula parameter, i.e. either ρ for elliptical copula, or θ for Archimedean copula.

Then, the copula parameter is plugged into the copula function and we get

$$F(x_1, x_2, ..., x_n) = C\Big(F_1(x_1), F_2(x_2), ..., F_n(x_n)\Big)$$

where x_i is the i^{th} asset with associated distribution $F_i(x_i)$ with $i \in (1, ...n)$. As a results, once a copula $C^*$is selected and plugged, one can generate a large number of scenarios for the marginals $F_i(x_i)$ dependent via the copula. Then, $P := w_1 F_1(x_1) + w_2 F_2(x_2)$ is the log return observations of the portfolio. A quantile can be drawn for a VaR estimate over P and an average can be computed above it to obtain an ES estimate.

CHAPTER 6

THE EXTREME VALUE THEORY

6.1 Introduction

There is a problem in estimating the VaR, a sort of paradox. We are interested in the tail of a distribution and generally construct a metric using components from the center of the distribution. For example, the plug-in method discussed in subsection 3.5.2. The Extreme Value Theory (EVT) aims at overcoming this issue since it deals specifically with the extreme deviations from the median of probability distributions. It tries to answer the question: "How to assess the probability of events that are more extreme than any previously". Since the VaR and ES predominantly focus on the *tail-risk*, the EVT can bring valuable insight to improving the VaR/ES estimation. The question, of course, we first care about, is what data do we consider to be extreme. There are two main approaches. The first one, that is the most traditional is the *Block Maxima*. Basically, it consists in dividing chronologically the sample into m *blocks* of the same size and picks the maximum of each block (see section 6.2). The other approach, a more modern model, is the so-called *Threshold exceedances* model. From a sample, one extracts peak values exceeding a certain *threshold* (see section 6.3). In this setting, it is important to note that the definitions of the distribution functions are different because it deals with *sub-probabilities*, or conditional probabilities. Both approaches can be modelled in the multivariate universe, bringing into play the concept of copula.

6.2 The Block Maxima model

Let's consider a sample of returns with n observations $(X_i)_{i\in\mathbb{N}} \sim F_X$. The block maximum is simply $M_n := \max\{X_1, ..., X_n\}$.

6.2.1 Generalized extreme value distribution

Similarly to the central limit theorem (CTL) where the sum of n RV converge to a normal distribution when n converges to the infinity, there is a *convergence of the maxima*. More specifically, suppose we find normalizing sequences of real number $c_n > 0$ and d_n such that $\frac{M_n - d_n}{c_n}$ converges in distribution. That is:

$$\Pr\left(\frac{M_n - d_n}{c_n} \leq x\right) = \Pr\left(M_n \leq c_n x + d_n\right) = F^n(c_n x + d_n) \xrightarrow{n \to +\infty} H(x)$$

In such condition, F is said to be in the *maximum domain of attraction* of H $\left(F \in MDA(H)\right)$. Provided this condition, we have that for $n \to \infty$, $M_n \sim$ GEV, the Generalized Extreme Value distribution. This result is given by the so-called *Fisher-Tripper-Gnedenko* theorem.

Definition 6.2.1. *The d.f. of the standard GEV is*

$$H_\xi(x) = \begin{cases} e^{\left(-(1+\xi x)^{-1/\xi}\right)} & \text{for } \xi \neq 0 \\ e^{-e^{-x}} & \text{for } \xi = 0 \end{cases}$$

where ξ is the shape parameter of the GEV function. A set of three parameters is obtained by defining $x := \frac{x-\mu}{\sigma}$ where $\mu \in \mathbb{R}$ is a location parameter and $\sigma > 0$ is a scale parameter.

The GEV has three different configurations depending on the shape paramter ξ. Figure 6.1 provides a graphical representation (Gilli and këllezi, 2006).

1. when $\xi > 0$, H_ξ refers to the so-called *Fréchet* d.f. . It is heavy-tailed, which is a particularly nice feature when studying financial assets behaviours.
2. when $\xi < 0$, H_ξ refers to a *Weibull* d.f. featured by an upper finite endpoint.
3. when $\xi = 0$, H_ξ refers to the *Gumbel* d.f. characterized by an exponentially decaying tails. Hence it can also model distributions with heavier tails than the normal.. It is the only one that is not bounded.

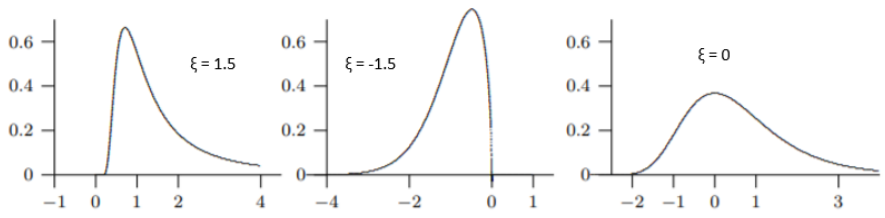

Figure 6.1: Densities for the Fréchet, Weibull and Gumbel functions.

The GEV distribution is similar to the central limit theorem (CLT) except that it can deal with variables that have an infinite variance. It especially means that the GEV function allows for inflated end tails.

In a risk management approach particularly, if X_i is a sequence of log-returns, one would rather be interested in the minimum values corresponding to the losses. It can be shown that the limiting theory for convergence of maxima encompasses the limiting behaviour of minima by using the fact that

$$\min(X_1,...,X_n) = -\max(-X_1,...,-X_n) \tag{6.1}$$

Similarly, the minima are going to converge if the d.f. of $-X_1,...,-X_n$, let's say $\tilde{F}(x) = 1 - F(-x)$, is in the maximum of domain of attraction of an extreme value distribution. Assuming $\tilde{F} \in MDA(H_\xi)$, we have:

$$\lim_{n\to\infty} \Pr\left(\frac{\max(-X_1,...-X_n)-d_n}{c_n} \le x\right) = H_\xi(x)$$
$$\Leftrightarrow \lim_{n\to\infty} \Pr\left(\frac{\min(X_1,...X_n)-d_n}{c_n} \le x\right) = 1 - H_\xi(-x)$$

Therefore, the distribution limits for minima will be of type $1 - H_\xi(-x)$.

An important, and nice, property from the Fréchet d.f. is its *regularly varying* function. This essentially means that the tails decay like a power function (e.g. exponential), which means that the end point of the distribution is infinite. This is particularly suitable for financial assets, as it is recognized that they are heavy-tailed distributed.

To fit the GEV, one estimates the three parameters (the scale, location and shape parameters) using the maximisation of the log likelihood of the density of the GEV distribution. The resultingVaR can be written as follow (Alentorn and Markose, 2008):

$$VaR_\alpha = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left(1 - \left(-\log(\alpha)\right)^{-\hat{\xi}}\right) \tag{6.2}$$

and the associated ES

$$ES_\alpha = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} + \frac{\hat{\sigma}}{\hat{\xi}\alpha}\int_0^\alpha (-\log v)^{-\hat{\xi}}\, dv \tag{6.3}$$

for $\xi \neq 0$, which suits the *Fréchet d.f.*. Another formula can be derived for $\xi = 0$ (see Alentorn and Markose (2008)).

6.3 The Threshold Exceedance model

Instead of taking maxima, this methodology is rather aimed to define a threshold u from which a loss is considered as extreme, and record all the observations that are above it. The whole distribution of the RV is broken down into a 'parent distribution' that features central tendencies, and a 'child distribution' that exclusively deals with the tails. For the latter, we refer to the so called *excess/residual life distribution over threshold* u. We have

Definition 6.3.1. *If X is a RV with a d.f. F, the excess/residual life distribution over the threshold u has a d.f.*

$$\begin{aligned} F_u(x) &= \Pr\Big(X - u \le x | X > u\Big) \\ &= \frac{\Pr(X \le x + u) + 1 - \Pr(X < u) - \Pr(X \le x + u \cup X > u)}{1 - \Pr(X < u)} \\ &= \frac{F(x+u) - F(u)}{1 - F(u)} \end{aligned}$$

for $0 \le x < x_F - u$, where $x_F \le \infty$, and since $\Pr(X \le x + u \cup X > u) = 1$

6.3.1 Generalized pareto distribution

In particular, it can be shown that such RV tends to follow the so-called Generalized Pareto Distribution (GPD), which has the following form (this refers to the Pickands-Belkema-de Haan theorem):

Definition 6.3.2. *The distribution function of the GPD is given by*

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \frac{\xi x}{\beta})^{-1/\xi} & \text{for } \xi \neq 0 \\ 1 - e^{\frac{-x}{\beta}} & \text{for } \xi = 0 \end{cases}$$

where $\beta > 0$, $x \ge 0$ when $\xi \ge 0$ and $0 \le x \le \frac{-\beta}{\xi}$ when $\xi < 0$. ξ and β are referred to the shape and scale parameters, respectively.

Similarly to the GEV, it is generalized in the sense that the shape parameter ξ value would change the type of distribution of $G_{\beta,\xi}$. When $\xi > 0$, it gives an ordinary Pareto distribution (*Fréchet*), when $\xi = 0$, it is an exponential distribution (*Gumbel*), and when $\xi < 0$, it is a *Weibull* d.f. which has an upper endpoint and is rather short-tailed. We would consequently be especially interested in GPD for which $\xi \ge 0$ since it denotes fatter tails.

6.3.2 Choice of the threshold

It requires a compromise between the variance and the bias. If the threshold is too low, the likelihood of including non-extreme values increases, which leads to bias. Conversely, a too high threshold would lead to so few data points that it might cause a large variance in the analysis and compromise the inference (McGreever, 2015). Plotting the *Mean residual life plot*, or the *Sample mean excess plot*, can help make an optimal decision. It plots u against the mean excess for a range of values of u. The plot should be linear above the threshold at which the GPD model becomes valid (McGreever, 2015). This gives an indication on a good threshold. Figure 6.2 provides an illustration of the *mean excess plot*.

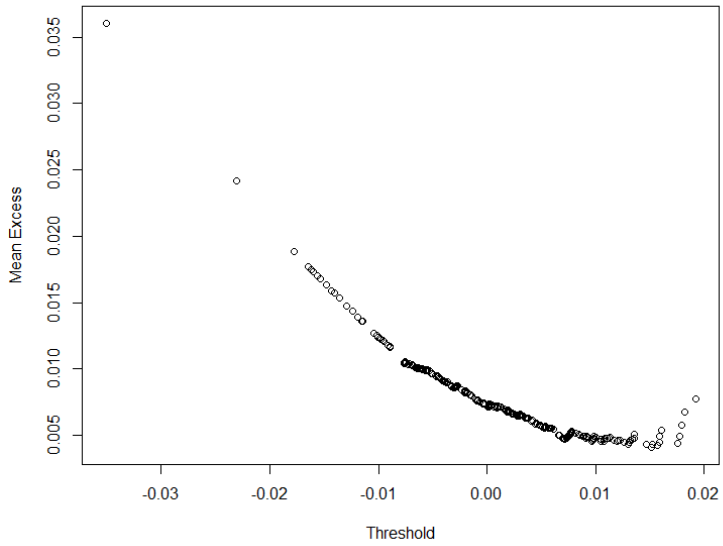

Figure 6.2: Example of a Mean excess plot.

Another interesting technique consists of fitting the GPD at a range of candidate threshold and then check the stability of the parameter estimates. This can be done through the so-called *parameter stability plot* that plots the parameter estimates against a range of values of u. A good threshold involves stable parameters estimates.

6.3.3 VaR and ES in POT

We would like to express the risk measures as a function of the GDP for values above the threshold, i.e. for $x \geq u$, so the associated CDF is $\Pr(X > x) = 1 - F(x) = \bar{F}(x)$ s.t.

$$
\begin{aligned}
\bar{F}(x) = \Pr(X > x) &= \Pr(X > x | X > u) \Pr(X > u) \\
&= \Pr(X - u > x - u | X > u) \Pr(X > u) \\
&= \big(1 - F_u(x-u)\big)\big(1 - F(u)\big) \\
&= \Big(1 - 1 + \big(1 + \frac{\xi(x-u)}{\beta}\big)^{-1/\xi}\Big)\bar{F}(u) \\
&= \bar{F}(u)\big(1 + \frac{\xi(x-u)}{\beta}\big)^{-1/\xi}
\end{aligned}
\tag{6.4}
$$

Accordingly, the VaR can be found straightforward as it simply is a quantile of the function. Equation 6.4 can be inverted to obtain a high quantile of the underlying distribution.

Proof

$$
\begin{aligned}
& y = \bar{F}(u)\big(1+\frac{\xi(x-u)}{\beta}\big)^{-1/\xi} \\
\Leftrightarrow & \Big(\frac{y}{\bar{F}(u)}\Big)^{-\xi} = 1+\frac{\xi}{\beta}(x-u) \\
\Leftrightarrow & \Big(\Big(\frac{y}{\bar{F}(u)}\Big)^{-\xi}-1\Big)\frac{\beta}{\xi}+u = x \\
\Rightarrow & F^{-1}(x) = u+\Big(\Big(\frac{x}{\bar{F}(u)}\Big)^{-\xi}-1\Big)\frac{\beta}{\xi}
\end{aligned}
$$

Therefore, for $\alpha > F(u)$, i.e. a loss exceeding the threshold u, we consequently have

$$
VaR_\alpha = q_\alpha(F) = u+\frac{\beta}{\xi}\Big(\big(\frac{1-\alpha}{\bar{F}(u)}\big)^{-\xi}-1\Big) \tag{6.5}
$$

Similarly, one can use the definition of the ES, relative to the VaR, to find its expression in the POT framework. This gives (McNeil et al., 2005, p.283) :

$$
ES_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 q_x(F)\,dx = \frac{VaR_\alpha}{1-\xi}+\frac{\beta-\xi u}{1-\xi} \tag{6.6}
$$

The parameters β and ξ can be easily estimated by maximum of likelihood. On the other hand, $\bar{F}(u) = \Pr(X > u)$ can be estimated by the frequency of points over the threshold, i.e. $\frac{N_u}{n}$. The estimate that can be plugged-in in the VaR and ES (see equation 6.4) is given by:

$$
\hat{\bar{F}}(x) = \frac{N_u}{n}\Big(1+\hat{\xi}\frac{x-u}{\hat{\beta}}\Big)^{\frac{-1}{\hat{\xi}}} \quad \text{with } x \geq u \tag{6.7}
$$

One should note that those equations stand provided that there is a sufficient proportion of sample values above the threshold to estimate $F(u)$ reliably.

6.4 Discussion

In this section, the two approaches of the EVT have been assessed. First of all, the block maxima method discussed in section 6.2 suffers from a considerable drawback. Indeed, it disregards many data, potentially meaningful for the analysis, since it only retains the maximum losses in large blocks, ignoring all others. Hence, it might be that in a same block, two large losses are significant. For that reason, it has been largely superseded in practice by the POT theory, where all the data over the threshold are used. This ensures relevant data are not being wasted provided that the threshold is high enough. This makes even more sense knowing that financial markets tend to depict clustering of volatility, especially regarding extreme events, what the threshold exceedances model is more likely to account for. Jumps in the time window may indeed ignore consecutive higher level of volatility. However, the POT model is entailed with the difficult determination of the threshold. Overall, the literature shows a preference in favor of the threshold exceedances model for financial risk assessment (McNeil et al., 2005).

CHAPTER 7

BACKTESTING FRAMEWORK

7.1 Introduction

Backtesting is mandatory in the banking sector. Banks can choose between a standard or internal model approach (IMA), whose the framework is dictated by the Basel Committee on Banking Supervision (BCBS). The regulator must ensure that the risk methodology adopted by the institution is conservative enough, and that the resulting capital reserves are robust to market shocks (Moldenhauer and Pitera, 2018). This conservativeness is assessed by the backtesting phase. Its definition is not uniform so it could take different aspects. Taking the definition from (Kerkhof and Melenberg, 2003), the backtesting is a final diagnostic check on a risk model carried out by the risk management team, through a set of statistical procedures aimed at inspecting whether real losses, observed ex-post, are in compliance with what has been predicted. Here, backtesting denotes the assessment of financial risk models, by means of historical data on risk forecast and P&L realizations (Roccioletti, 2016). There is a well-established regulatory backtesting procedure for the VaR, relating on the exceedance rate (first mentioned in section 3.5). It will be deeply explained and even extended with other important backtesting aspects. On the other hand, the backtesting framework with regard to the ES is still vague because of its novelty (recall it has been thought that the ES was not backtestable). As a result, we provide the theory of the most convincing ES backtesting framework we have found in the literature in section 7.3.

7.2 The VaR backtesting framework

There are three main backtesting processes: tests about *coverage* (exceedance procedure), *distribution* and *independence*. The most widely referred in the regulatory framework to backtesting

is the exceedance procedure which involves the so-called **Unconditional Coverage** (UC). Accordingly, the VaR backtesting involves systematically comparing the historical VaR forecast with the observed portfolio returns (Jorion, 2006, p. 139). Formally, the exceedance function can be described as dummy variables $I_t(\alpha)$ associated with the ex-post observations of a VaR exceedance at $\alpha\%$ at date t:

$$I_t(\alpha) = \begin{cases} 1 & \text{if } L_t > \widehat{VaR}_{t|t-1}(\alpha) \\ 0 & \text{otherwise} \end{cases} \tag{7.1}$$

where L_t corresponds to the observed loss at time t, and $\widehat{VaR}_{t|t-1}(\alpha)$ the Value-at-Risk estimation that prevails for time t. Then, the fraction of exceedances is recorded and compared with the confidence level α of the VaR. If we denote $\mathbf{x}$ the number of exceedances (i.e. when the observed loss exceeds the expected VaR) within a time interval $\{1, ..., T\}$, it is formally defined as

$$x = \sum_{t=1}^{T} I_t(\alpha) \tag{7.2}$$

Additionally, in the regulatory framework, two important assumptions underline the backtesting procedure:

1. The probability that a loss in excess of the expected ex-ante VaR is realized ex-post is precisely equal to the coverage rate:

   $$\Pr\left[I_t(\alpha) = 1\right] = \mathbb{E}\left[I_t(\alpha_t)\right] = \alpha$$

   Under this assumption, the exceedance rate, measured as the average number of exceedances, is a convergent estimator of the coverage. We expect that if our estimation method is reasonable, the proportion of exceedances would be close to α

   $$\frac{x}{T} \underset{T \to \infty}{\longrightarrow} \alpha$$

2. The assumption of **independence of exceedances** is respected when $I_t(\alpha)$ is independent of $I_{t-k}(\alpha)$. That is, when VaR exceedances observed at different dates are independently distributed.

When the two above assumptions are respected, the exceedance rate x follows, by definition, a Binomial distribution

$$x \sim \mathscr{B}in(T, \alpha)$$

with mean $\mathbb{E}(x) = \alpha T$ and variance is $Var(x) = \alpha(1-\alpha)T$. All of that allows to approximate a normal distribution (via the CLT) if T is large enough:

$$Z := \frac{x - \alpha T}{\sqrt{\alpha(1-\alpha)T}} \sim \mathscr{N}(0, 1)$$

From this relation, one can test the UC hypothesis. This is the so-called *Kupiec test*, where H_0 is in fact the unconditional coverage assumption

$$\begin{cases} H_0 : \mathbb{E}(I_t) = \alpha \\ H_1 : \mathbb{E}(I_t) \neq \alpha \end{cases}$$

Accordingly, under the null hypothesis, the number of exceedances is

$$x_{H_0} := \alpha T \tag{7.3}$$

Type I error occurs when a correct model is rejected, and **Type II error** occurs when a wrong model is failed to be rejected. Note that it is clear that, in risk management, type II error is much more costly for the firm.

The second backtesting dimension is the **Backtesting of distribution**. It aims at assessing the first step of the RMP, as explained in section 3.1. It assesses the goodness of fit of the overall loss distribution. A methodology is given in (Berkowitz, 2001). Beyond that, as a simple recipe, one can easily test the normality of a distribution via for instance *Shapiro Wilk* tests (Shapiro and Wilk, 1965) or Jarque-Bera tests. The former assesses whether the sample data have the skewness and kurtosis matching a normal distribution. Hence, the null hypothesis is a joint hypothesis of the skewness and the excess kurtosis being zero. The former will be used in this thesis.

The third and last aspect of the backtesting procedures relies on **independence backtesting**. Independence in the VaR measure's performance from a period to the next one can be assessed into two different ways. Either the independence of the exceedances, either the independence of the loss quantiles (Holton, 2003). While the former is assumed for the coverage rate and the latter is assumed for distribution tests, Basel did not introduce any independence formal test. There are no direct tests for independence so one has to rather focus on clustering, auto-correlation, etc. We are particularly interested in exceedance rates so we will describe its independence test. Exceedances independence can be tested via the so-called *Christoffersen's independent test*. This is essentially a likelihood ratio test (LRT) for consecutive exceedances for which the null hypothesis is that the probability that a non-exceedance occurs in t is independent from the the previous period (if there was an exceedance or not). This translate into the fact that, under H_0, $\hat{q}_0 = \hat{q}_1$, where

$$\hat{q}_0 = \Pr\left(I_t = 0 | I_{t-1} = 0\right)$$
$$\hat{q}_1 = \Pr\left(I_t = 0 | I_{t-1} = 1\right)$$

These probabilities can be estimated via frequencies. For $T+1$ periods, there are of course T consecutive observations, with $t_{(0,0)}$ pairs for which there are no exceedances, $t_{(0,1)}$ pairs

for which there are no exceedance and then exceedance, $t_{(1,0)}$ pairs for which there are an exceedance and then no exceedance and $t_{(1,1)}$ pairs for which there are both aexceedance. Therefore

$$\hat{q}_0 = \frac{t_{(0,0)}}{t_{(0,0)} + t_{(0,1)}}$$
$$\hat{q}_1 = \frac{t_{(1,0)}}{t_{(1,0)} + t_{(1,1)}}$$

Under H_0, using the fact that $T = t_{(0,0)} + t_{(0,1)} + t_{(1,0)} + t_{(1,1)}$, it can be easily shown that

$$\hat{q}_0^* = \hat{q}_1^* = \hat{q}^* = \frac{t_{(0,0)} + t_{(1,0)}}{t_{(0,0)} + t_{(0,1)} + t_{(1,0)} + t_{(1,1)}}$$

This builds the LRT to test the independence of exceedances (LR shown in subsection 14.5.1 in (Holton, 2003)).

7.3 The ES backtesting possibilities

As previously mentioned, Gneiting (2011) showed that the ES is not elicitable, meaning that no scoring functions are possibly drawn for the ES. Accordingly, it has been sometimes thought that the ES is not backtestable. However, this mostly means that *model comparison* for ES is compromised, while it is less sure that this conclusion can be extended to a *model testing* of ES (Acerbi and Szekely, 2014). Moreover, the VaR backtesting via exceedances test has nothing to do with scoring functions (Acerbi and Szekely, 2014), what suggests that it can be applied to the ES.

Multiple methods have been drawn, using different assumptions and methodologies. In particular, as summarized in Moldenhauer and Pitera (2018), parametric methods have been developed, and notably by Wong (2008) who suggested a parametric saddle-point method. Another is Righi and Ceretta (2013) who studied the truncation based ES backtests. In addition, Acerbi and Szekely (2014) propose three backtests models that are non-parametric but require simulations. In order to provide some theoretical background that proves the ES backtestability, we are subsequently presenting the two first ES backtesting models introduced by Acerbi and Szekely (2014). The first model requires the prior backtest of the VaR. It is based on the conditional definition of the ES:

$$\begin{aligned} & ES_{\alpha,t} = \mathbb{E}\Big[L_t | L_t > VaR_{\alpha,t}\Big] \\ \Leftrightarrow & \mathbb{E}\Big[L_t | L_t > VaR_{\alpha,t}\Big] - ES_{\alpha,t} = 0 \\ \Leftrightarrow & \mathbb{E}\Big[\Big(\frac{L_t}{ES_{\alpha,t}} - 1\Big)\Big| L_t > VaR_{\alpha,t}\Big] = 0 \end{aligned} \tag{7.4}$$

If the VaR passes the backtest phase, one can separately test the magnitude of the realized exceedances against the model predictions. It can be shown that the following statistic prevails:

$$Z_1(\mathbf{L}) = \frac{\sum_{t=1}^{T} \frac{L_t I_t}{ES_{\alpha,t}}}{\sum_{t=1}^{T} I_t} - 1 \tag{7.5}$$

with the indicator function $I_t := \mathbb{1}_{\{L_t > VaR_{\alpha,t}\}}$ and $\sum_{t=1}^{T} I_t$ consequently being the number of exceedances. Defining $P_t^{[1-\alpha]}$ as the conditional loss distribution fitted by the model at the $(1-\alpha)$ quantile, and $F_{L,t}^{[1-\alpha]}$ the realized loss distribution at the $(1-\alpha)-$quantile, the hypothesis tested by the statistic in Equation 7.5 can be written:

$$\begin{aligned} H_0 &: P_t^{[1-\alpha]} = F_t^{[1-\alpha]} \quad \text{(the model is exact in the tail)} \\ H_1 &: \widehat{ES}_{\alpha,t} \leq ES_{\alpha,t} \quad \text{(underestimation of risk)} \\ & \widehat{VaR}_{\alpha,t} = VaR_{\alpha,t} \end{aligned}$$

Under the null hypothesis, the realized tail is assumed to be the same as the predicted tail of the return distribution. The alternative hypothesis rejects Expected Shortfall without rejecting VaR. With regards to the statistic, the hypotheses can be written as:

$$\begin{aligned} \mathbb{E}_{H_0}[Z_1] &= 0 \\ \mathbb{E}_{H_1}[Z_1] &> 0 \end{aligned}$$

Acerbi and Szekely (2014) also propose to test directly the ES (not after the VaR), by taking the unconditional definition of the ES :

$$\begin{aligned} ES_{\alpha,t} &= \frac{1}{1-\alpha}\mathbb{E}\left[L_t I_t\right] \\ \Leftrightarrow \mathbb{E}[&\frac{L_t I_t}{(1-\alpha)\, ES_{\alpha,t}} - 1] = 0 \end{aligned} \tag{7.6}$$

With a similar methodology to the first test, one can construct the test statistic as follow:

$$Z_2(\mathbf{L}) = \frac{1}{T(1-\alpha)} \sum_{t=1}^{T} \frac{L_t I_t}{ES_{\alpha,t}} - 1 \tag{7.7}$$

Accordingly, the hypothesis are:

$$\begin{aligned} H_0 &: P_t^{[1-\alpha]} = F_t^{[1-\alpha]} \quad \text{(the model is exact in the tail)} \\ H_1 &: \widehat{ES}_{\alpha,t} \leq ES_{\alpha,t} \quad \text{(underestimation of risk)} \\ & \widehat{VaR}_{\alpha,t} \leq VaR_{\alpha,t} \end{aligned}$$

and with regard to the statistic, the hypothesis become:

$$\begin{aligned} \mathbb{E}_{H_0}[Z_2] &= 0 \\ \mathbb{E}_{H_1}[Z_2] &> 0 \end{aligned}$$

We reject the null hypothesis (RH0) if $p < \alpha$ or fail to reject the null hypothesis ($\bar{\text{RH0}}$) if $p > \alpha$.

Besides, a more straightforward manner to backtest the ES is one that is performed by backtesting the VaR at different confidence levels. Emmer et al. (2015) propose to estimate the ES using an interpretation of the ES as integrated VaR. However, they do not justify the number of levels to use. Taking $\alpha = 2.5\%$ in line with the regulatory framework, the ES can be approximated in the following way

$$ES_{2.5\%} \approx \frac{1}{5}\left[VaR_{2.5\%} + VaR_{2\%} + VaR_{1.5\%} + VaR_{1\%} + VaR_{0.5\%}\right] \tag{7.8}$$

This allows to easily backtest the ES by backtesting the VaR with the different levels α. However equation 7.8 rather provides an approximation of the value of the ES. As opposed, we would prefer to calculate the ES in its continuous form, as the area under the loss distribution beyond the VaR point (given per Eq. 2.5). It hence can be used in an exceedance rate procedure with the number of exceedances defined as

$$x = \sum_{t=1}^{T} I_t(\alpha) = \sum_{t=1}^{T} \frac{1}{1-\alpha} \int_{\alpha}^{1} \mathbb{1}_{\{L_t > VaR_t(x)\, dx\}} = \sum_{t=1}^{T} \mathbb{1}_{\{L_t > ES_t(\alpha)\}} \tag{7.9}$$

The number of exceedances determine the performance of the model with regard to a certain framework (refer to section 7.4 for an introduction to the Basel framework).

All of these methods do not rely on elicitability and globally work. However, they may require heavy mathematical skills, model assumptions, or large samples. All of them cannot reach the simplicity of the VaR backtesting.

7.4 The Basel framework for backtesting

The BCBS has drawn a very detailed framework that specifies the capital requirements procedures for financial institutions. In banks especially capital requirement should be met on a daily basis. In the Basel Framework (2019), it is stated that in measuring market risks, a choice between two broad methodologies, i.e. VaR and ES, is permitted, however subject to the approval of national authorities (BCBS, 2019a). Overall, banks have flexibility in devising the precise nature of their models, but they have to respect certain minimum standards for the purpose of calculating their capital requirement. Notably, the VaR is set at the 99^{th} percentile and the ES at the 97.5% confidence level. The historical observations that will be used for the VaR (and ES) estimations must be at least one year (i.e. 250 trading days).

In addition, a bank needs to perform backtesting to assess the quality of the estimation provided by its risk model. The backtesting method of Basel follows the same exceedance process as described in the backtesting coverage test. Indeed, it compares the VaR measure calibrated to a one-day holding period against each of the actual realization of the loss distribution, such that a exceedance occurs when the actual loss registered exceeds the corresponding VaR. It should then return the number of exceedances over 12 months (i.e. 250 trading days) (BCBS,

2019a). Depending on the number of exceedances recorded, the Basel framework classifies the model into three backtesting zones, distinguished by colours. The zones are defined according to the number of exceedances generated in the backtesting period. More specifically, we have (BCBS, 2019a)

1. *Green zone*. This corresponds to results that do not themselves suggest a problem regarding the quality or accuracy of the model. It is indeed associated with the least number of exceedances.

2. *Amber zone*. The number of exceedances start to be material such that the model deserves to be questioned.

3. *Red zone*. Classification corresponding to 10 or more exceedances. This indicates a substantial problem with the model.

The backtesting zones are defined according to the cumulative probabilities of obtaining a certain number of VaR exceedances under the model. As explained in Equation 7.3, under H_0, for $\alpha := 1\%$, the number of exceedances should equal $1\% \cdot 250 = 2.5$. Accordingly, the cumulative probability of obtaining 5 exceedances marks the limit between the green and amber zones, as it equals 95%.

Backtesting zones	Number of exceedances
Green	1
	2
	3
	4
Amber	5
	6
	7
	8
	9
Red	more than 10

Table 7.1: Backtesting zones for VaR, computed assuming $\alpha = 0.01$ and N = 250 observations

A similar classification based on exceedances rate procedure for the ES has been proposed by Costanzino and Curran (2015), which extends the backtesting model for VaR from the Basel Committee (table 7.1). We won't enter into the details, but it is based on a 97.5^{th}percentile, which translates the higher severity of the expected shortfall. It is aligned with the traffic light system, i.e. green zone if cumulative probability < 0.95. It gives the following table.

Backtesting zones	Number of exceedances
Green	0
	1.3929
	2.1131
	3.0276
	4.0520
	5.0622
	5.7048
Amber	5.7049
	6.9844
	8.5285
	9.8832
Red	more than 9.8833

Table 7.2: Backtesting zones for ES, computed assuming α = 0.025 and N = 250 observations

Part II

Practical and empirical analysis

CHAPTER 8

RISK ASSESSMENT OF A SIMULATED LOW AND HIGH RISK PORTFOLIO

8.1 Introduction

The type of portfolio a manager wishes to manage is crucial. For example, it is intuitive to think that the risk underlying sovereign bonds will be completely different from the one of crypto-currencies. Accordingly, we ask ourselves if the best RMP (among the ones described through Chapter 4,5 and 6) might differ depending on the risk profile of the financial series. We also spend some time at combining RMP, such as testing copula and EVT together. As a result, portfolios simulations allow to study *the relevance of the sophistication of the RMP with regard to the riskiness of the portfolio*. More particularly, we will assess in **i)** low risky simulations whether the sophistication is useless, or even detrimental, or rather beneficial to the performance of the backtesting, and in **ii)** high risky simulations whether the sophistication can substantially improve the risk estimation and if conventional methods remain good performers. In addition, we compare the stability of the VaR and ES estimates in order to evaluate their ability to produce qualitative estimations depending on the number of historical observations at hand. We subsequently present our assumptions underlying the concept of riskiness, both for the low and high risk environments.

First of all, it is necessary to make assumptions regarding the (discrete time) assets price model. It is generally agreed that Geometric Brownian Motions (GBM) have good properties to model assets prices. It relies on the so-called Stochastic Differential Equation (SDE) from

which it is the solution. The general **SDE** is given by:

$$\begin{aligned} dS_t &= \mu_t S_t dt + \sigma_t S_t dW_t \\ S_0 &= s_0 \end{aligned} \tag{8.1}$$

where μ_t and σ_t are the so called *drift* and *diffusion coefficient*, respectively. Note also that the solution S_t would be a continuous process.

Depending on the assumptions made, we can either derive a low risk asset (simple formulation of the factors) or, in contrast, a high risk asset (increased complexity of the factors). Moreover, portfolios of assets are considered. This involves a dependence structure needed to be described (e.g. the correlation).

In order to be simulated numerically, the SDE has to be discretized. For example, this can be done using the so-called *Euler's discretization scheme*. It replaces dt by small increments δt, we have

$$\begin{aligned} \delta S_t^{(n)} :=& S_{t+\delta t}^{(n)} - S_t^{(n)} \\ =& \mu\left(t, S_t^{(n)}\right) \delta t + \sigma\left(t, S_t^{(n)}\right) \delta W_t \end{aligned}$$

where $\delta t := W_{t+\delta t} - W_t \sim \sqrt{\delta t} Z$, with $Z \sim \mathbf{N}(0,1)$

8.2 Low risk simulations description

In an unrisky environment, we will consider that μ_t and σ_t are steady over time, i.e. they become deterministic so that the underscored t can be remove. The SDE can be rewritten as follows:

$$dS_t = \mu\, S_t dt + \sigma\, S_t dW_t \qquad \text{where } W_t \sim \mathbf{N}(0,t) \tag{8.2}$$

Here, W_t is a standard BM. It can be shown that the solution to this equation gives the standard GBM to model assets prices[1].

Definition 8.2.1. *A standard Geometric Brownian Motion is defined as:*

$$S_t = S_0 e^{(\mu - \frac{\sigma^2}{2})t + \sigma W_t}$$

In terms of the preceding observation, the standard GBM can be defined as follow:

$$S_{t_{i+1}} = S_{t_i} e^{\left(\mu - \frac{\sigma^2}{2}\right) + \sigma\sqrt{t_{i+1} - t_i}\left(W_{t_{i+1}} - W_{t_i}\right)}$$

[1] this can be indeed solved using the so called *Ito's calculus* theory but is out of the scope of this thesis.

The GBM as given in definition 8.2.1 is actually a *log-normally* distributed random variable[2]. Hence, the risk factors change taken for the empirical analysis are the log-returns. Making a bit of arithmetic, we get that

$$\begin{aligned} r_t &= \ln\left(\frac{S_t}{S_{t-1}}\right) = \ln S_t - \ln S_{t-1} \\ &= (\mu - \frac{\sigma^2}{2}) + \sigma(W_t - W_{t-1}) \end{aligned}$$

The term $\frac{\sigma^2}{2}$ is simply a convexity correction term (to balance the Jensen's effect ensued from). Moreover, by definition, $(W_t - W_{t-1}) \sim \mathbf{N}(0,1)$. In order words, this means that the log-returns can be described in the following way:

$$\begin{aligned} r_t &= \ln S_t - \ln S_{t-1} \\ &= \tilde{\mu} + \tilde{\sigma} Z_t \end{aligned} \tag{8.3}$$

where $Z_t \sim \mathbf{N}(0,1)$.
Regarding the dependence structure, we would consider the linear dependence represented by the linear correlation. It can be also shown that for the GBM $B := \rho W + \sqrt{1-\rho^2}W'$, then $d[W,B]_t = dW_t dB_t = \rho dt$. This allows to model the joint behaviour of the assets.

It follows that on the software R, we generate normal distributed data. In order to fit the regulatory framework, the historical data were generated over 250 trading days, and as many were generated for the purpose of the backtesting. We decided to simulate two stocks and to build the equally weighted portfolio made out of them. The individual stocks depict daily features that were strongly inspired from the BEL20 and IBEX35 in order no to take purely arbitrary parameters for our simulations. To have a common scale, it is assumed that both simulated stock S_1 and S_2 are such that their initial value (i.e. in $t = 0$) are $s_1^0 = s_2^0 = 100$. Moreover, the log-returns are assumed to have the following annual expectation and standard deviation, respectively:

$$\begin{aligned} \mu_1 = -2,3\%,\ \sigma_1 = 12,06\% \\ \mu_2 = -6,19\%,\ \sigma_2 = 13,05\% \end{aligned}$$

In addition, we have assumed a correlation level of 0.78 between both stocks, which corresponds to the correlation of our benchmark formed by the log-returns of the BEL20 and IBEX35.

The following graph illustrates one simulation from that methodology on a daily frequency over 500 days.

[2] with $\mathbb{E}(S_t) = S_0 e^{\mu t}$ and $Var(S_t) = S_0^2 e^{2\mu t}(e^{\sigma^2 t} - 1)$

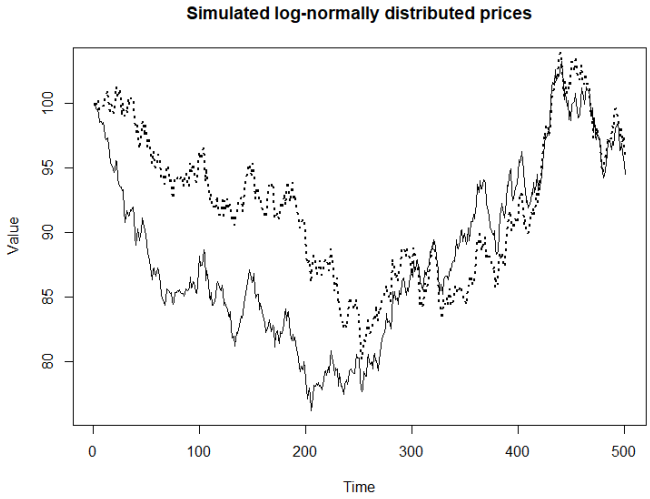

Figure 8.1: Illustration of the simulated correlated stock prices ($\rho = 0.711$).

A *Jarque-Bera test* is also performed to assess formally the normality of the log-returns.

8.3 High risk simulations description

8.3.1 Drivers of the portfolio riskiness

In the context of this work, a high risk portfolio is assumed to exhibit the following features:

1. A *high volatility* σ that reflects the inherent high instability and fluctuation of the price of the portfolio.

2. *Clusters of volatility* imply that large changes tend to be followed by large changes. This here refers to GARCH processes.

3. *Unpredictable large jumps* in the losses. Here we refer to jumps that might occur as explained in equation 4.11 in monte carlo with jumps.

4. A *positive skewness of the loss distribution*. In one word, it measures its asymmetry. Being positive means that the tails extends toward more positive losses than in the Gaussian case. Mathematically, the skewness γ_1 of a variable X is its third standardised moment:

$$\gamma_1 = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \tag{8.4}$$

5. A *kurtosis* greater than 3. It measures the heaviness of the tails. Being bigger than 3 means that the distribution exhibits fatter tails than the normal distribution. The kurtosis $Kurt$ of a

variable X is its fourth standardised moment:

$$Kurt[X] = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] \tag{8.5}$$

6. *A high tail dependence:* We here refer to subsection 5.4.2 for underlying theoretical background. This illustrates the chain reaction that might occurred in bearish environments. The most famous example is the CDO crisis in 2007 that lead to the global financial crisis.

8.3.2 Unbounded Johnson distributed log-returns

In an effort to produce high risk data including several features of riskiness presented in subsection 8.3.1, we propose to rely on the so called *Johnson's S_U-distribution*. The Johnson distribution has good properties in the context of risk management, in particular for VaR and ES because of its accuracy in tail fitting to the empirical distributions (Simonato, 2011). Indeed, X is Johnson distributed (unbounded) when

$$z := \gamma + \delta \sinh^{-1}\left(\frac{X-\xi}{\lambda}\right) \sim \mathbf{N}(0,1) \tag{8.6}$$

γ and δ are shape parameters, ξ is a location parameter, and λ is a scale parameter. The Johnson parameters can be estimated from the four first moments. It consequently offers flexibility on the skewness and kurtosis, as opposed to the normal distribution. Since it is commonly agreed that market returns usually encompass negative skewness and high excess kurtosis (Naguez and Prigent, 2017), we believe that the Johnson distribution may more suitably match the market reality.

Accordingly, one can easily define *Johnson stochastic processes*. By supposing $S_t := S_0 e^{X_t}$ and taking the inverse function of equation 8.6, we can assume the prices to behave as follow (Naguez and Prigent, 2017):

$$S_t = S_0 e^{\lambda \sinh\left(\frac{W_t-\gamma}{\delta}\right)+\xi} \tag{8.7}$$

where W_t is a standard BM. After discretization of the process, and adding random losses jumps, we can finally simulate asset prices with a high risk profile. Please note however that we do not add GARCH processes, nor high tail dependence between the assets (refer to the second and sixth drivers of subsection 8.3.1).

Concretely, we generated 500 log-returns with skewness $\gamma := -1$ and $Kurt := 8$. The difference with the low risk simulation is visible in Figure 8.2.

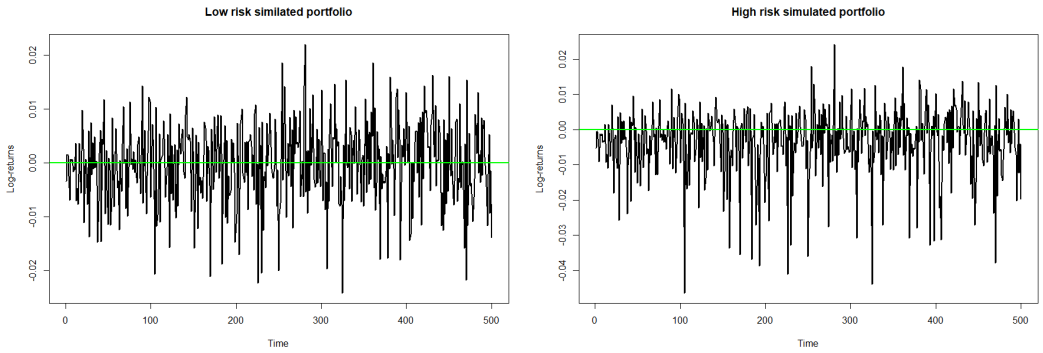

Figure 8.2: Illustration of the simulated low risk log-returns (left) and high risk log-returns (right). The high risk portfolio is populated with more and lower negative log-returns.

Particularly, Figure 8.3 illustrates the divergence between the high risk simulated log-returns and the low-risk ones via the empirical density of the simulations.

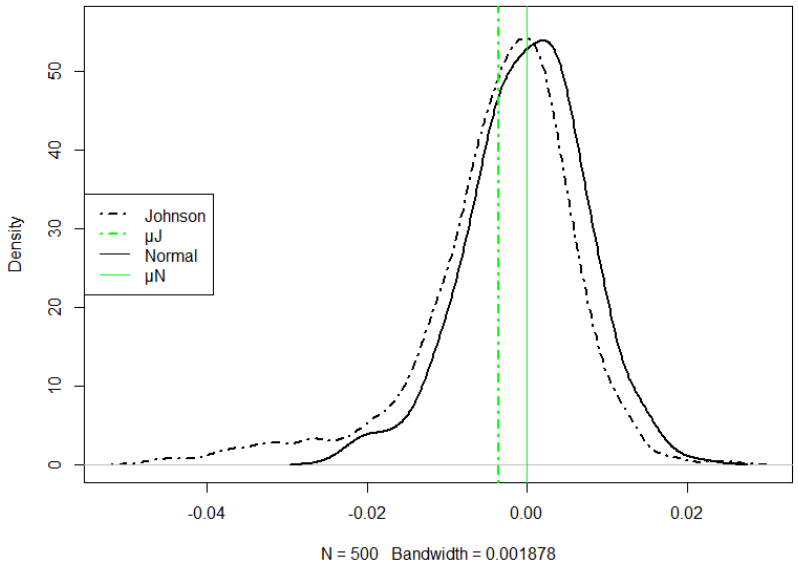

Figure 8.3: Comparison between the Johnson (J) and Normal (N) simulated densities. $\gamma_J = -1.27$; $Kurt_J = 5.9$, and $\gamma_N = -0.31$; $Kurt_N = 3.3$. μ and σ are the same in both simulations.

The null hypothesis stating the normality of the returns is strongly rejected, with a p-value $< 2.2e-16$ in the Jarcque-Bera test for the Johnson distributed simulation while the p-value for the normal simulations is 0.005609. Note also that the mean of the Johnson distribution is smaller since the skewness increases the probability to obtain low returns.

8.4 Copulas and EVT fitting and comparisons for the simulations

8.4.1 Fitting Copula model

The method of copula was used in order to derive the multivariate form of the portfolio distribution. As explain in chapter 5, it captures the dependence structure between the assets to produce their joint distribution. The choice of copula induces a different dependence structure, which is illustrated in Figure 8.4. Recall that we use Gaussian marginals with a linear correlation in the case of the low risk framework.

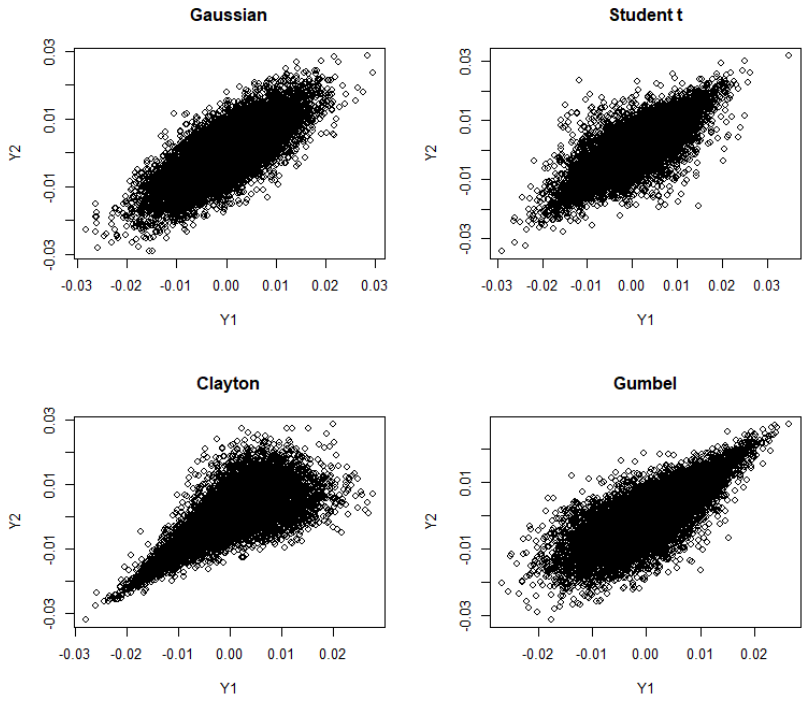

Figure 8.4: 10,000 simulated points from two normally assumed log-returns with dependence structure from Copula: (a) Gaussian, (b) t, (c) Clayton, and (d) Gumbel. $\tau = 0.515$.

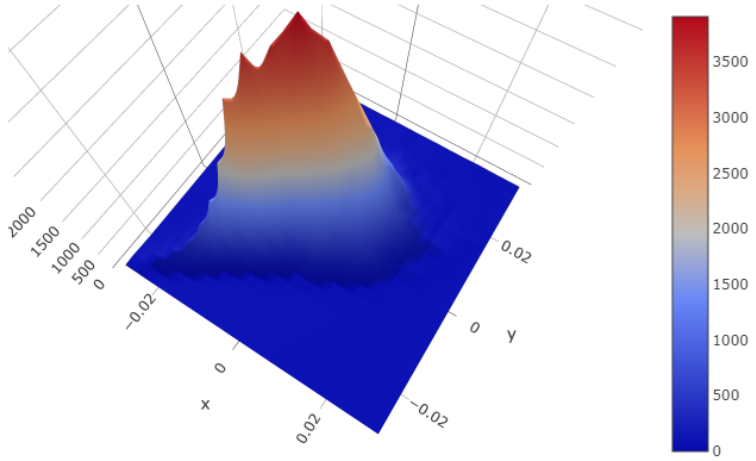

Figure 8.5: 3D plot of the probability density function under the Clayton copula. The left tail is particularly inflated.

Figure 8.4 clearly shows the different dependence structures that can arise depending on the copula used (Ane and Kharoubi, 2003). As expected and shown in Figures 8.6 and 8.7, in the context for which the log-returns are Gaussian marginals, the VaR and ES are the greatest using the Clayton copula to model the dependence as the probability of simultaneous negative returns is increased[3], leading to inflated lower tail (see Figure 8.5). The highest VaR and ES estimations are then given by the t copula, which is itself followed by the Gaussian copula. In contrast, the Gumbel copula delivers the smallest values since it barely deals with lower tail of the P&L distribution, whereby quantiles are extracted.

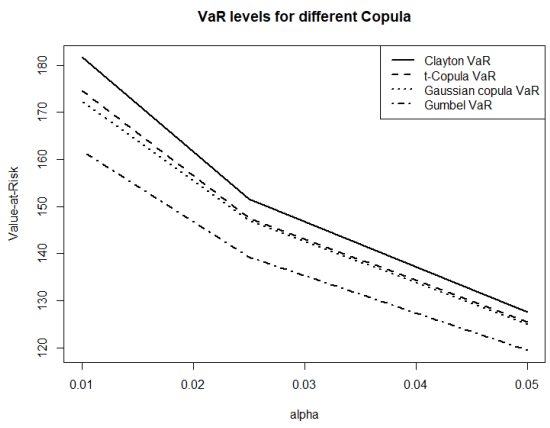

Figure 8.6: Levels of value-at-risk for the Copula at various levels alpha ($\alpha \in [0.01; 0.05]$).

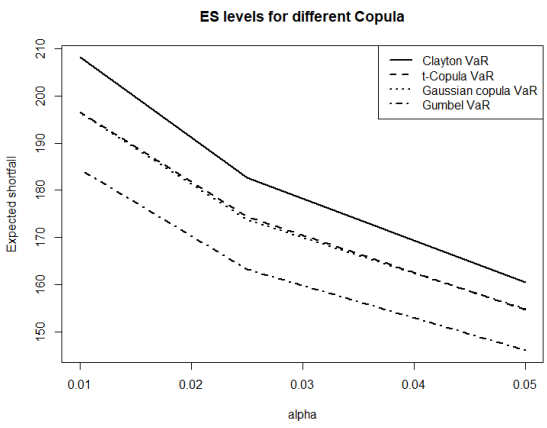

Figure 8.7: Levels of expected shortfall for the Copula at $\alpha \in [0.01; 0.05]$.

[3]this is the definition of the Clayton copula.

The same copula have been fitted to the risky data, however using Johnson SU-distributed marginals. The impact of the higher level of risk is clearly visible in the lower tail. Indeed, the presence of negative skewness breaks the symmetry and provokes a larger proportion of low returns. Yet the strong dependence in the lower tail implied by the clayton copula produces even better results.

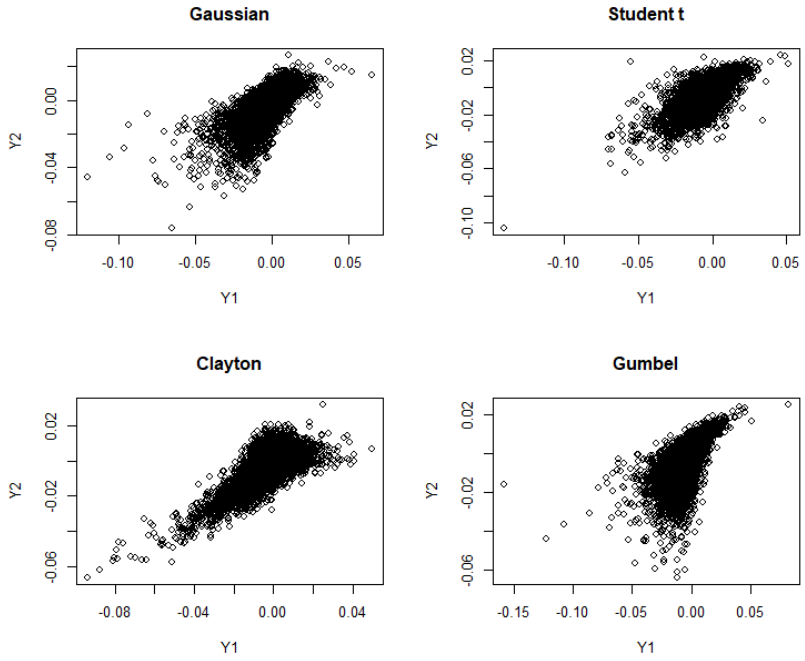

Figure 8.8: 10,000 simulated points from two Non-normal (Johnson) marginals Y1 and Y2 for the log-returns with dependence structure from respectively Copula: (a) Gaussian, (b) t, (c) Clayton, and (d) Gumbel. $\tau = 0.505$.

8.4.2 Fitting Extreme Value Theory model

In Figure 8.9, one can observe the working of both the blockmaxima and threshold exceedances methods at the heart of the EVT. The red points are the log-returns taken into consideration to fit the Generalized extreme value distribution and the Generalized pareto distribution respectively.

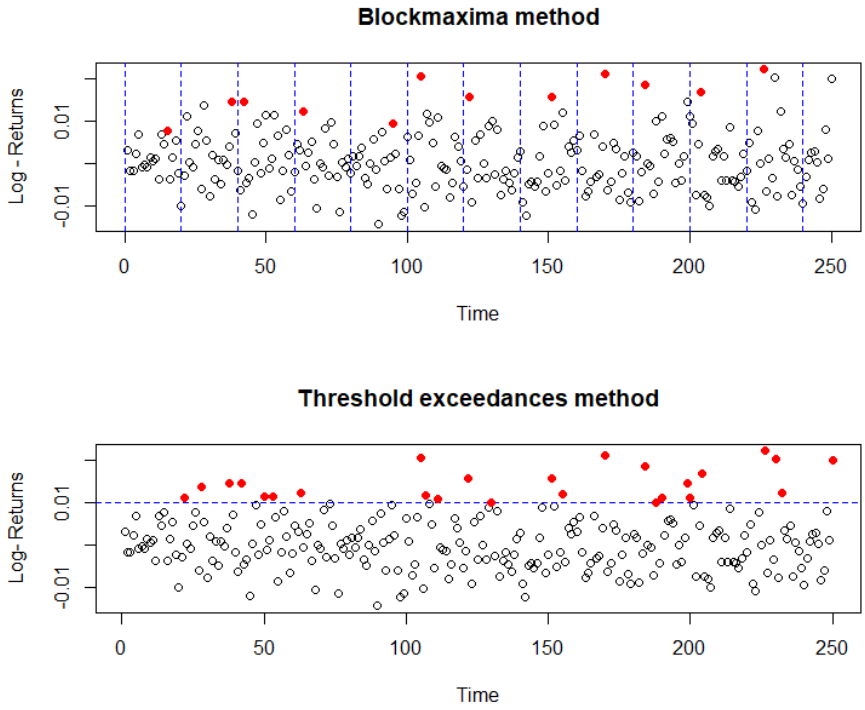

Figure 8.9: Method of blockmaxima and threshold exceedances over the low risk simulated data. The block window is defined to 20 days and the threshold to 0.01 to best fit the data.

The latter figure clearly shows the drawback of the blockmaxima method, to the regard of the number of relevant returns that it ignores. Also, if the sequence of blocks is such that outliers are selected in every block, the risk estimation produced under the blockmaxima method will skyrocket hugely at outrageous levels. As opposed, the threshold exceedances method always remains bounded by the threshold. It is thus more robust.

8.4.2.1 Threshold determination

In order to fit a GPD model, the threshold needs to be selected. A suitable threshold is obtained by plotting the *mean excess plot*.

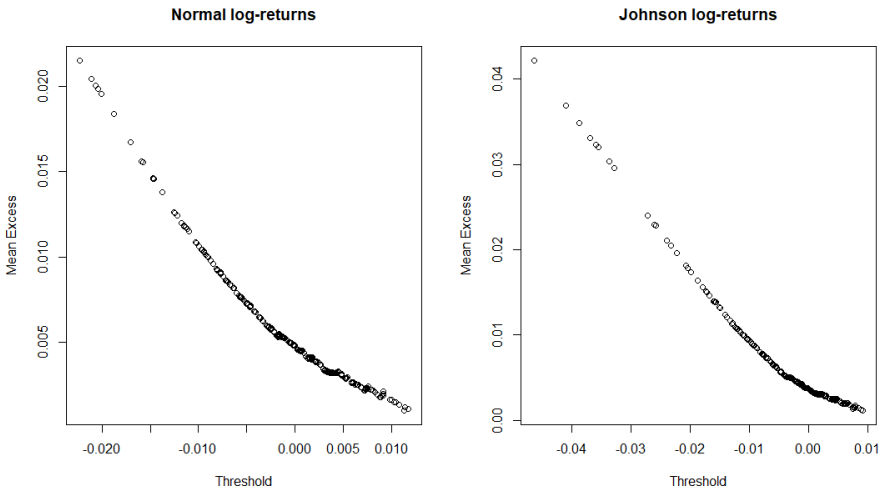

Figure 8.10: Mean excess plot for (a) Normal, and (b) Johnson simulated log-returns.

Figure 8.10 provides the mean excess plot for the series of Normal and Johnson log-returns, and suggest a threshold to be set at respectively -0.005 and -0.01, values from which the patterns get linear. According to the mean excess plot, these are the thresholds that would give the better trade-off between the bias and variance, thereby ensuring that we have enough extreme data. Although this technique is conditional to interpretations, it is often enough to determine a suitable threshold. As an example, Figure 8.11 depicts the model fit confronted to the empirical distribution for the GPD of the normal log-returns. Same graphical results can be obtained for the Johnson portfolio (refer to appendix A.2.2).

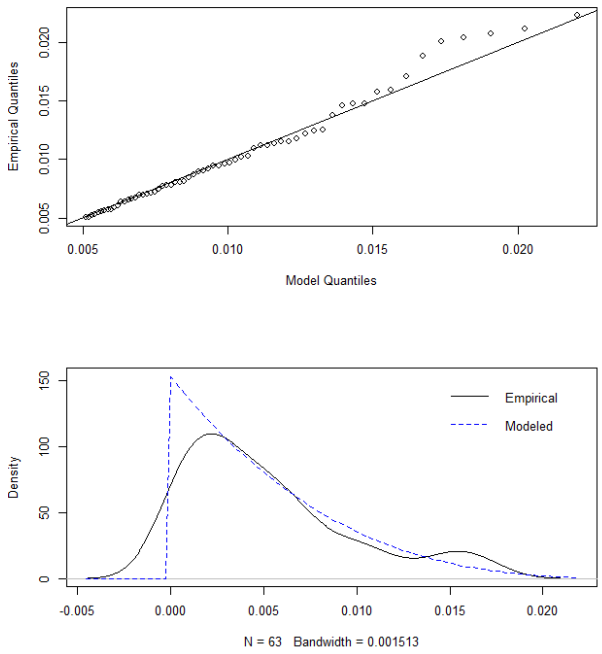

Figure 8.11: Diagnostic plots of fit for the GPD with Gaussian simulated log-returns (250 first observations). The threshold $u = -0.005$. The figure at the top represents the comparisons of quantiles and the figure at bottom shows the fitted GPD against the empirical distribution.

In Figure 8.11, one can observe that the quantiles plot is perfectly linear at lower quantile levels and start to deviate at higher quantile levels. This means that the model fit is extremely accurate when dealing with common returns, but slightly struggle with extreme ones. This is however to be expected given the scarcity of data. The distribution induced can also be observed on the second plot in Figure 8.11. The probability distribution matches the *Fréchet distribution* as described in theory for $\xi > 0$, and effectively depicts some deviation from the empirical distribution around the 0.015 return level. Afterwards both distributions however align. Overall, the diagnostic plots provide enough evidence to reasonably assume the good fit of the GPD model.

The EVT is also applied to copula.

8.4.3 Fitting GARCH model

Based on the AIC criteria, the best GARCH fit corresponds to a GARCH(1,1). Accordingly, it assumes that the continuously compounded return are described as follow:

$$r_t = \mu + \epsilon_t$$
$$\epsilon_t = z_t \, \sigma_t$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_1 \sigma_{t-1}^2$$

where ϵ_t is a shock, z_t a random variable *iid*, and σ_t the conditional volatility under the GARCH(1,1) representation. These returns, residuals from the GARCH fit, are used as the financial series in the case a GARCH specification is added to the estimation model.

8.5 Backtesting results and performance comparisons

In this section, we review the performance of all the VaR and ES estimation models that were presented in the Part I of this thesis. The performance assessment is conducted through an *exceedance procedure backtesting*, as intended in the Basel framework (see chapter 7). To estimate the VaR and ES, we consider the losses on a position in the portfolio. The current value of the position is assumed to be $V_t = 10,000$. The losses for this portfolio are given by $L_{t+1} = -V_t X_{t+1}$ where X_{t+1} denotes the daily log-returns of the portfolio (i.e. the risk factors change as recalled from section 1.3). Then, the VaR is identically used to estimate the maximum loss that can be incurred at the confidence level α: $\widehat{L}_{t+1} = -V_t \widehat{VaR}_{t+1|t}(\alpha)$. It is basically interpreted as an amount of economic capital. Please, remind that, in such framework, 250 days are used for the model estimation, then a rolling window of 1 day is run until the completion of the 250 days backtesting window. We would expect an accurate model to produce $\alpha\%$ of exceedances for the VaR. The VaR 1% and ES 2.5% are evaluated.

The following tables provide results for the low-risk and high-risk portfolios with respect to the considered estimation methods. We also define an *Average daily error* to take into account the bank interests.

(8.8)
$$\text{Average daily Error} = \left| \frac{\text{expected number of exceedances} - \text{realised number of exceedances}}{\text{time window}} \right|$$

It measures the absolute discrepancy between the observed and expected number of exceedances. For the bank, too many or too few failures is detrimental. Too many exceedances would lead to an underestimation of the risk that can fail to protect the financial institution against potential losses. Too few exceedances would mean that the risk is overestimated and that the bank is pledging more capital than needed, which is money that cannot be invested and reduces the performance of the bank overall. As a result, in the following, **the "Basel zone" better represents the regulator point of view, whereas the "Error" is more aligned with the bank interests**[4]. The *Average daily error* term is therefore a conservative measure because it accounts for both sides of the discrepancy: the over and under-estimation of risk. The backtesting results for the simulated log-returns are subsequently displayed, and additionally the plots

[4]However, note that withing the bank, conflicting interests might exist as well: a risk manager's job is to ensure the going concern of the bank is preserved, while the trading side seeks the highest return possible which means being subject to lower constraints on capital.

of the estimated VaR/ES through the hold-out period in order to observe the exact evolution of the estimations through time.

Estimation methods	VaR failures	ES failures	Basel zone	Average daily error (%)
HS	2	2	Green	0.2
HS-GARCH	2	2	Green	0.2
Risk Metrics	6	6	Amber	1.4
GARCH	6	6	Amber	1.4
MC	5	5	Amber	1
MC jump	3	2	Green	0.2
Clayton copula	4	4	Green	0.6
Gumbel copula	6	6	Amber	1.4
Gaussian copula	6	6	Amber	1.4
Student copula	6	6	Amber	1.4
GEV	1	1	Green	0.6
Clayton-GEV	0	0	Green	1
GPD	0	0	Green	1
GARCH-GPD	0	0	Green	1
Clayton-GPD	0	0	Green	1
Econ. Unbiased	6	6	Amber	1.4
Econ. Unbiased-GARCH	5	5	Amber	1

Table 8.1: Number of exceedances of the 99% VaR and 97.5% ES estimates calculated using various estimation methods, and for the series of **normal simulated log-returns** (low-risk). Expected failures=2.5.

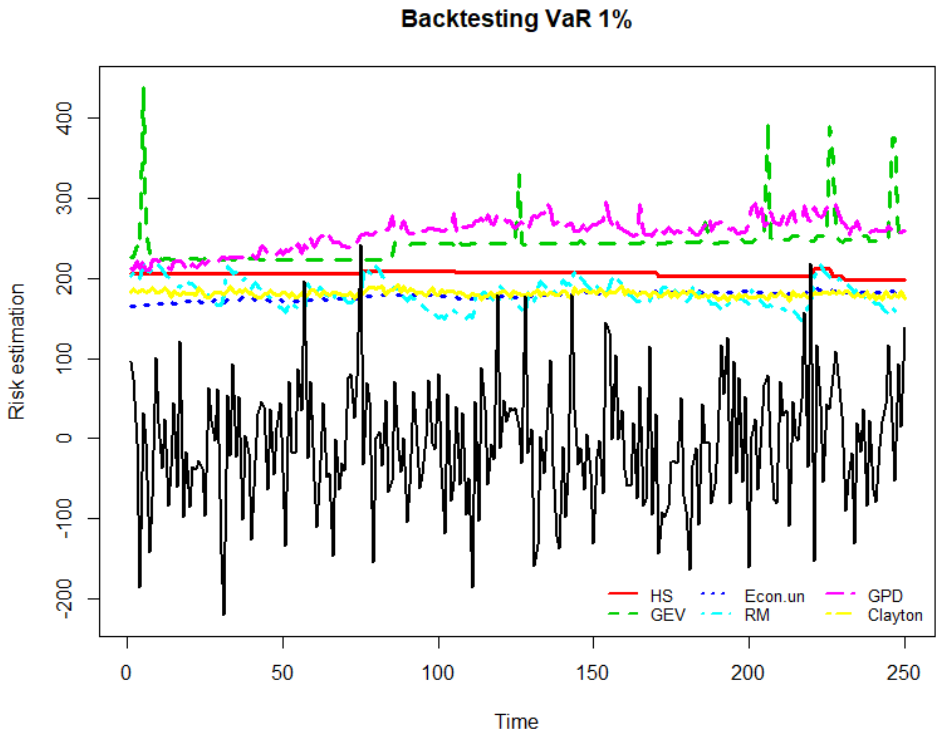

Figure 8.12: Daily losses for the backtesting period with risk measure estimate $VaR_{1\%}$, and exceedances for the methods in the legend. All backtesting results for VaR and ES to be found in Appendix A.1.1. *Portfolio: Normal*.

From the results collected in Table 8.1, we firstly notice that the $VaR_{1\%}$ and $ES_{2.5\%}$ behave in the same way. Besides, with this methodology, there are fewer methods favorable to the bank than to regulators. This is a consequence of the considerable risk overestimation induced by the EVT-based methods (zero exceedance). They result to be detrimental to the bank performance but meet the interests of the regulators by preconizing a large capital cushion. Note that, although the GEV has an associated error of 0.6, Figure 8.12 shows that it is not robust and thus inadequate for a risk manager either. Overall, in the regard of the failures number, we note that the estimation methods are narrowly dispersed for the normal log-returns (visible in appendix A.1.1). Moreover, one can observe that parametric methods are generally outperformed by non-parametric methods. Since the market environment is set to be constant, the risk metrics and the GARCH models perform very poorly, while the HS performance is largely fostered. Overall, this shows that the sophistication of the model with a conditional process might be actually detrimental if there is no volatility clustering. The MC jump method is also a good performer, providing evidence that the addition of a jump help align the estimation with the portfolio riskiness, and eventually withstand larger losses.

We then note that the best performing copula is the Clayton copula. As shown in Figure 8.5,

it increases the dependence in the lower tail of the risk-factors changes distribution thereby inflating the tail whereby the VaR and ES are calculated. It results in greater estimations than the other copulas, and a better exceedances rate in the present case. However, the benefit of it is mitigated by the low risk context. This shows that, in simple environment, the sophistication via the Clayton copula is not useful, but not detrimental to the backtesting performance neither, as against the EVT.

Another important fact is to observe the half-tone performance of the Econ. unbiased risk measure although we assumed a Gaussian framework. This may suggest that more observations are necessary to integrate the Gaussian pattern withing the simulation.

The subsequent table and plot display the results for the **Risky portfolio simulation**

Estimation methods	VaR failures	ES failures	Basel zone	Average daily error (%)
HS	2	2	Green	0.2
HS-GARCH	2	2	Green	0.2
Risk Metrics	10	10	Red	3
GARCH	7	7	Amber	1.8
MC	10	10	Red	3
MC jump	2	2	Green	0.2
Clayton copula	1	0	Green	0.6
Gumbel copula	2	2	Green	0.2
Gaussian copula	1	1	Green	0.6
Student copula	1	1	Green	0.6
GEV	0	0	Green	1
Clayton-GEV	0	0	Green	1
GPD	0	0	Green	1
GARCH-GPD	1	1	Green	0.6
Clayton-GPD	0	0	Green	1
Econ. Unbiased	10	10	Red	3
Econ. Unbiased-GARCH	12	12	Red	3.8

Table 8.2: Number of exceedances of the 99% VaR and 97.5% ES estimates calculated using various estimation methods, and for the series of **Johnson's S_U log-returns** (high-risk). Expected failures=2.5.

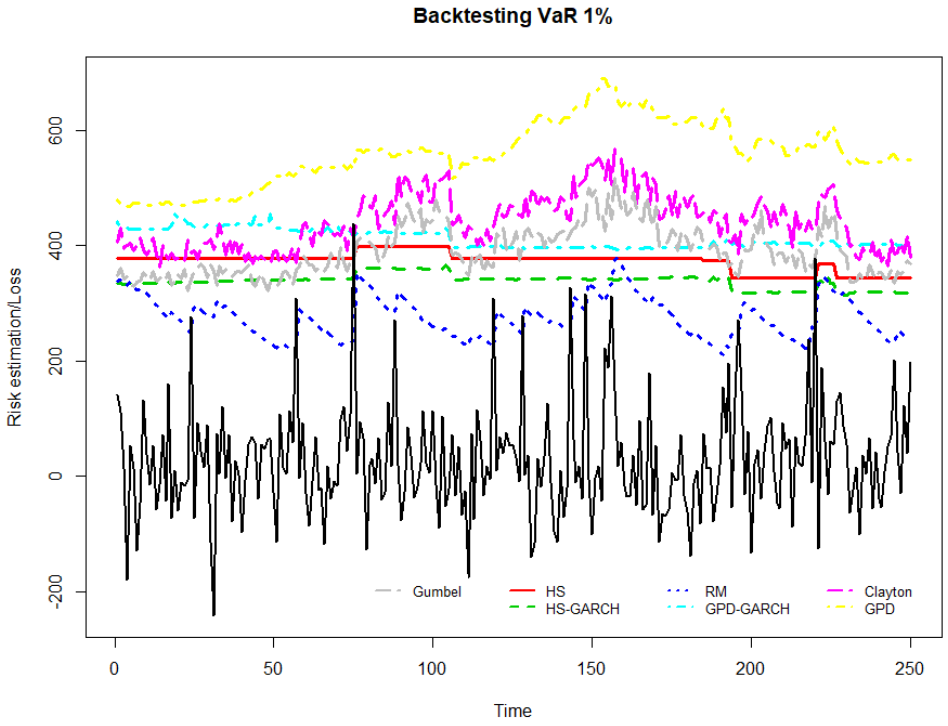

Figure 8.13: Daily losses for the backtesting period with risk measure estimate $VaR_{1\%}$, and exceedances for the methods in the legend. All backtesting results for VaR and ES to be found in Appendix A.1.2. *Portfolio: Johnson*.

Table 8.2 identically perform estimations from the 1% VaR and 2.5% ES. The risk estimation levels in Figure 8.13 are 1.5x up to 2x higher than for the normal portfolio. The riskiness of the *Johnson*-distributed portfolio correctly translates into the risk estimates. As far as we are concerned, most methods, except the parametric ones essentially, seem to provide satisfying results. This consolidates the irrelevance of the normality assumption when dealing with (risky) financial series. This would expose the bank's risk manager to a critical underestimation of the risk. Regarding the Econ. unbiased form of the VaR and ES, once more, because of the large sample period (250 days), we have a convergence of the unbiased to the plug-in distribution [5]. They both poorly perform. The simplistic assumptions of normality is consequently shown to be unsustainable for the regulator as well as the risk manager.

We then examine the sophisticated procedures. On the one hand, the performance of Copula has increased in presence of riskier log-returns in the sample. In particular, the Clayton copula performs better as the negative skewness increases the proportion of low returns. It is more conservative in this case and may get the favor of regulators. In general, all the considered copula provide a satisfying estimations. However, note that, although the higher risk of loss is the worse captured by the Gumbel copula, we might believe that, here, it matches the expected

[5] One may recall Equation 3.14

number of failures due to the steady high level of negative log-returns across the sample periods. On the other hand, the EVT models are again not performing ideally. Whereas one could have expected EVT to be one of the most suitable methods in this risky context, this is not reflected here by the risk manager perspective (no exceedance but a huge VaR estimate, see Figure 8.13). This can be inferred by looking at the time series (refer to Figure 8.2). The high risk portfolio is characterized by high level of negative log-returns over both the in sample and out of sample periods. This basically means that EVT run over the in sample inflates even more the lower tail and assumes too high levels of risk. It results that the risk is overestimated. The GARCH specification added to the model adjusts the risk estimation at reasonable levels. We may suppose that the GARCH compensates the EVT overestimation in a way.

Finally, good performing models are again the MC jump and, surprisingly, the HS. The HS works well because of the steady market environment over the entire time frame.

8.6 Stability of the estimation of the VaR and ES

In the last section, we saw that even the unbiased estimators may poorly perform because it could not integrate the Gaussian distribution through 250 observations. To examine the influence of the number of observations, we set 100 samples and conduct estimates over a time window N varying from 50 to 5000 data (number of historical observations). The $VaR^{\text{plug-in}}$ and $ES^{\text{plug-in}}$ are estimated. Their mean over the samples has been observed to be very close to the true value (please refer to Appendix A.4). The standard deviations associated with the estimations (see Table 8.3) examine their deviation from the true estimate. The $VaR^{\text{plug-in}}_{1\%}$ and $ES^{\text{plug-in}}_{2.5\%}$ are investigated, as well as the $ES^{\text{plug-in}}_{1\%}$ for comparison.

		Time window size				
		50	250	500	1000	5000
RMP	$ES^{\text{plug-in}}_{1\%}$	21.549	10.419	8.254	5.474	2.339
	$VaR^{\text{plug-in}}_{1\%}$ (a)	19.200	9.362	7.433	4.923	2.106
	$ES^{\text{plug-in}}_{2.5\%}$ (b)	19.278	9.398	7.460	4.941	2.114
Difference	(b) - (a)	0.078	0.036	0.027	0.018	0.008

Table 8.3: Standard deviation associated with the estimation of the $ES^{\text{plug-in}}_{1\%}$, the $VaR^{\text{plug-in}}_{1\%}$ (a), the $ES^{\text{plug-in}}_{2.5\%}$ (b) over the samples and for different size of time window. The difference between (b) and (a) is also provided for comparison.

The estimations are slightly more volatile for the $ES_{2.5\%}$ compared to the $VaR_{1\%}$ but not significantly. We conclude that there is not significant difference in the quality of the estimations of the $VaR_{1\%}$ and $ES_{2.5\%}$, even for small window size. If we consider the same significance level for both risk measures (i.e. 1%), we end up with a more volatile ES, hence requiring an increasing number of historical observations to reach the same deviance as the VaR. At the

level α 1% and 2.5%, respectively, this is not the case. We assume that this result still holds for other estimation techniques.

8.7 Conclusions

Overall, this analysis over a low risk portfolio with normally distributed risk-factors changes and a high risk portfolio Johnson-distributed has shown that the risk estimation methods prevailing depends on the settings. If the series are rather stationary with a constant risk profile, conventional approaches such as HS capture the current behavior of the financial series is prone to be good performing. On the other hand, the EVT seems only judicious if shocks are expected in the market environment, prone to higher losses in the future. This could be expected if one observes many outliers in the last observations of the sample. On the other hand, the Clayton copula has demonstrated relevance in both simulation cases, and the copulas performance in general is more pronounced in the higher risk environment. Another key result is the risk underestimation of the Gaussian assumption under the risky assumptions we previously defined. We finally pointed out the equal performance of the $VaR_{1\%}$ and $ES_{2.5\%}$ in low risk and high risk environment.

This analysis over simulations has allowed to study risk estimation methods over data depicting a well-known and desired behavior. However, the reality of financial markets is often uncertain and unpredictable. Then, whereas these results encompass many valuable takeaways, they need to be compared with real financial assets. This is done is the next chapter.

CHAPTER 9

RISK ASSESSMENT OF A PRESUMED LOW AND HIGH RISK REAL PORTFOLIO

9.1 Introduction

In this chapter, we focus our analysis on real portfolios of financial assets and confront the results with the simulated data. In the same logic as for the simulated risk-factor changes, we opted for both what it is often believed to be a low risk and high risk portfolio. Nowadays, there is an increasing hype around Exchange-Traded Funds (ETF) and Cryptocurrencies. They theoretically show similar properties as a low and high risk profile, respectively, so they will be at the core of this practical analysis. Provided the previous results and interpretations, the importance of the data exploration and analysis has been brought to light in order to assess the risk profile of portfolios. Accordingly, the best estimation methods might vary with the market conditions to avoid under- or over-estimation of risk.

As such, we provide a technical sheet that breaks down the statistical points that should be covered by the risk manager while assessing the riskiness degree of a portfolio in the data exploratory analysis, in the light of our assumptions subsection 8.3.1. The evaluation of the risk magnitude of the portfolio at hand will help determine the most appropriate risk estimation method(s) to be used. Since we investigate all risk estimation methods in this work, all tests proposed in the technical sheet won't necessarily be performed (when not essential).

9.2 Technical sheet to assess the riskiness of a portfolio

1. **Broad view of the behaviour of the financial series**: this is the first thing to do. This is done through a graphical analysis of *historical prices* and *(log)-returns*. It gives relevant information on the first four moments (i.e. average return, volatility, asymmetry, presence of heavy tails). The *empirical histogram* as well as the *boxplot* should also be plotted in order to visualize the spread of the data. It is notably a great tool to spot outliers. Indeed, outliers may give reasons to use EVT, especially if theses outliers are recent. Indeed, they would give reasons to believe in stress market conditions, where EVT is particularly useful.

2. **Stationarity and auto-correlation analysis**: It allows to question the *iid* assumptions that usually underlie the models. The *ACF and PCF*, although being non-formal, remain crucial. Formally, the *Augmented Dirk Fuller* test and *Durbin Watson* tests can be used to test stationarity and autocorrelation, respectively.

3. **Distribution tests:** As mentioned in section 3.1, the RMP's first step is distribution fitting. As such, it is very important to check the fitting quality of the distribution estimations for parametric approaches. For example, this can be done through a Jarque–Bera test in order to assess the normality assumption of the log-returns. Other tests also exist for more complex distributions (e.g. the Kolmogorov–Smirnov test in Lopes (2011)).

4. **Dependence structure assessment:** As outlined earlier, the dependence structure is crucial in portfolio management. In order to visualise it, one should first plot the scatter plots of the variables with regard to each other. Moreover, the risk manager should also estimate the tail dependence between the assets in order to choose which copula to use.

9.3 Presumed low risky products: the Exchange-Traded Funds

9.3.1 Why to choose ETF as low risk profile

Exchange-Traded Fund or ETF are a collection of securities that tracks an underlying index. They may contain many different types of financial products to reach the index performance. For example, they can be constituted of stocks, bonds, commodities, etc. Similarly as stocks, they can be traded on stock exchanges. Their attractiveness has also been driven by their low associated transaction cost (Curcio et al., 2014). As a result, they have become increasingly popular in the financial world. For example, it went from a US market of $ 400 B in 2005 to $ 1000 B 5 years later (Ramaswamy, 2011) and it surged to $ 5 T at the end of 2018 (Evans and Wilson, 2018). That is one main reason that drove our choice to study that type of security. Moreover, it is low risky in the sense that, unlike the majority of financial products that are often leptokurtic and skewed (Rachev and Mittnik, 2000), the ETF products in the market often follow a Gaussian distribution. In particular, Petronio et al. (2014) observed that in many cases,

the Gaussian assumption is accepted (more than 40% of the stocks), whereas this hypothesis is usually rejected for usual stock (more than 90%).
In particular, the portfolio of the five ETF that we have chosen are the top ETF from the US market[1]. To be aligned with the Basel rules, we have retrieved two years of data, which amounts to 502 observations, from 2017-06-19 to 2019-06-17. We aim at putting ourselves in the shoes of an individual who would like to invest in the hyped market of ETFs, which, because of its definition, is known to be pretty secured.

9.3.2 Exploratory analysis of an equally weighted portfolio of ETF

1. **Broad view of the behavior of the financial series:**

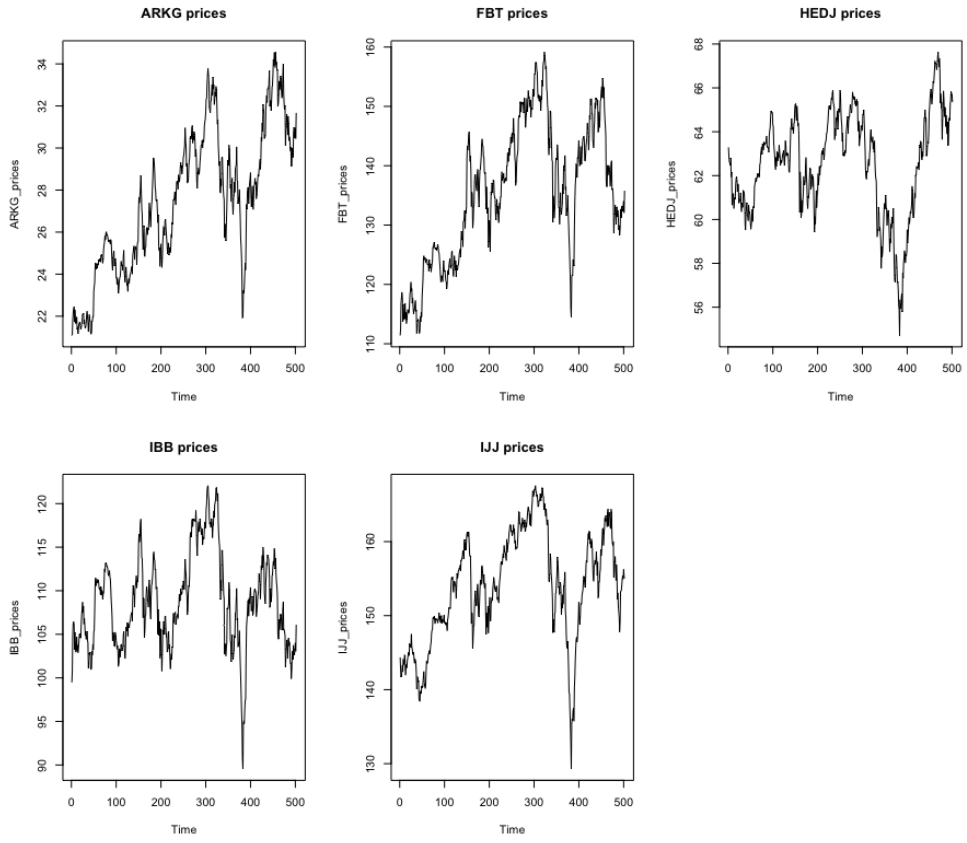

Figure 9.1: Daily prices of the ETFs constituting the studied ETF portfolio (502 observations).

Globally the ETF prices are more fluctuating than expected. They particularly incurred big losses at the same time, around the 400^{th} observation. This may jeopardize the Gaussian assumption, part of the low risk assumptions. Besides, the plots of the squared log-returns Figure 9.2 are from great help in order to first spot the series's stationarity and auto-

[1]Retrieved from Yahoo Finance. Accessed on the 18^{th} of June 2019

correlation. As the studied model for autocorrelation is GARCH (which does not take into account the sign of the shocks), the square of log-returns are taken. The log returns of the equally weighted portfolio are also studied:

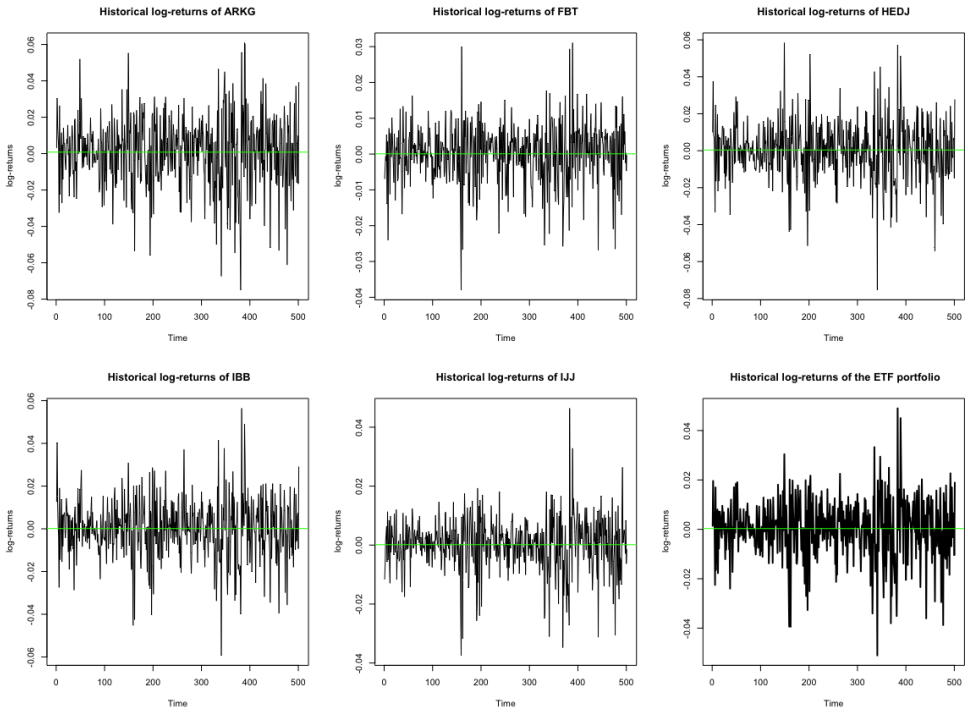

Figure 9.2: Historical daily log-returns of the ETFs and the equally weighted portfolio (bottom right).

They seem to depict 2 light periods of volatility, between the 100^{th} and 200^{th} observations and, more accentuated, between 300^{th} and 400^{th} observations. This might indicate the possibility for auto-correlation. This is further investigated through the ACF and PCF of the portfolio.

2. **Stationarity and auto-correlation analysis:**

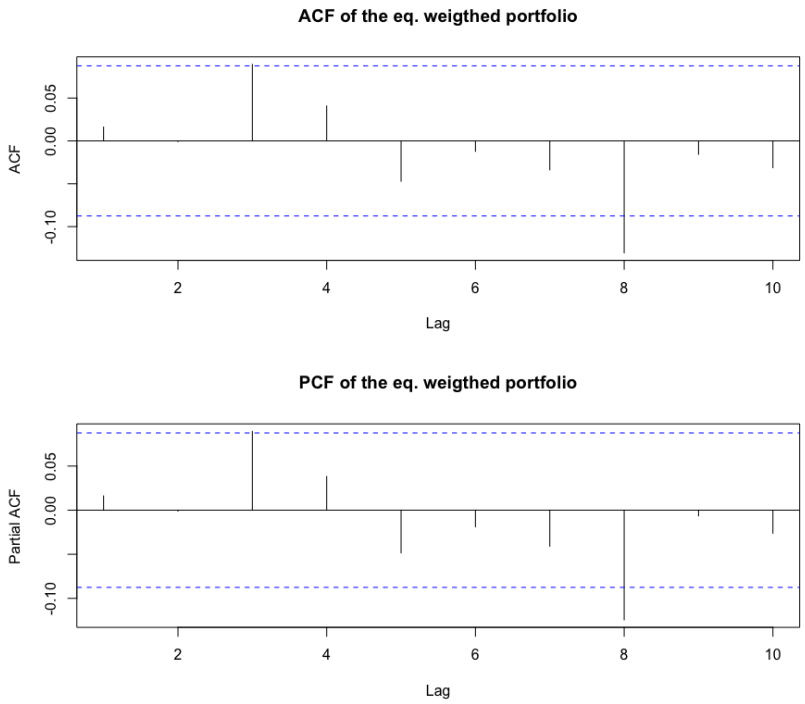

Figure 9.3: ACF and PCF of the log-returns of equally weighted portfolio of ETF.

As seen in Figure 9.3, we can already conclude that GARCH processes is likely to be detrimental to the model performance. Furthermore, it would add unnecessary variability in the cash to be added to the reserves, which is non-desirable from the risk manager perspective.

3. **Distribution tests:**
   We now explore if the Gaussian distribution is sustainable for the portfolio. The following Figure gives the Gaussian fitting to the portfolio and its associated boxplot:

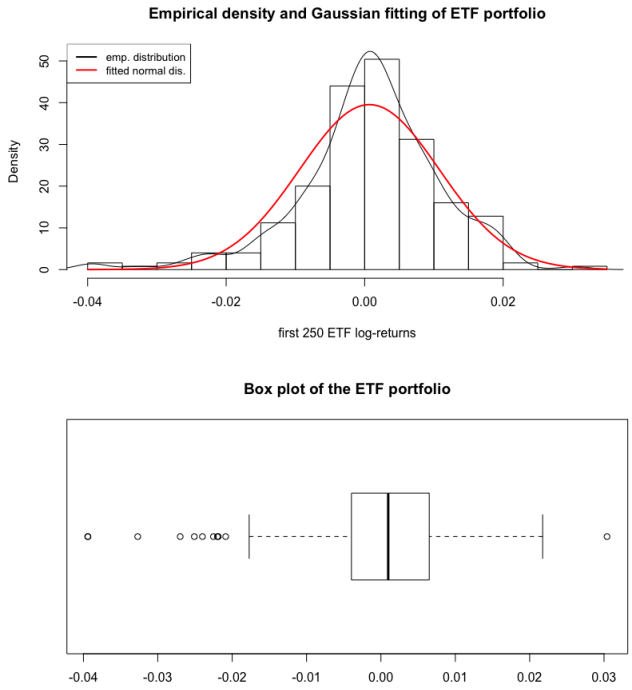

Figure 9.4: Gaussian fitting on to the log-returns of the equally weighted portfolio of ETFs (left) and its respective boxplot (right).

We graphically observe that the Gaussian fitting (in red) seems to be very different from the empirical distribution. We also observe outliers that heaven tails, particularly the left one. Moreover, it seems slightly negatively skewed.

In addition to the graphs, the formal measures are presented in Table 9.1. As foreseen, a non negligible excess kurtosis is observed, indicating higher probabilities of high losses. It is also negatively skewed, what marks a deviation from the Gaussian framework. The Jarque-Bera test strongly rejects the Gaussian assumption (p-value = 1.706e-13).

Moments	Empirical estimation
Mean	0.00031
Standard deviation	0.012
Skewness	-0.43
Kurtosis	5.14

Table 9.1: Values of the first four empirical moments of the sample for the ETF portfolio log-returns.)

4. **Dependence structure assessment:**

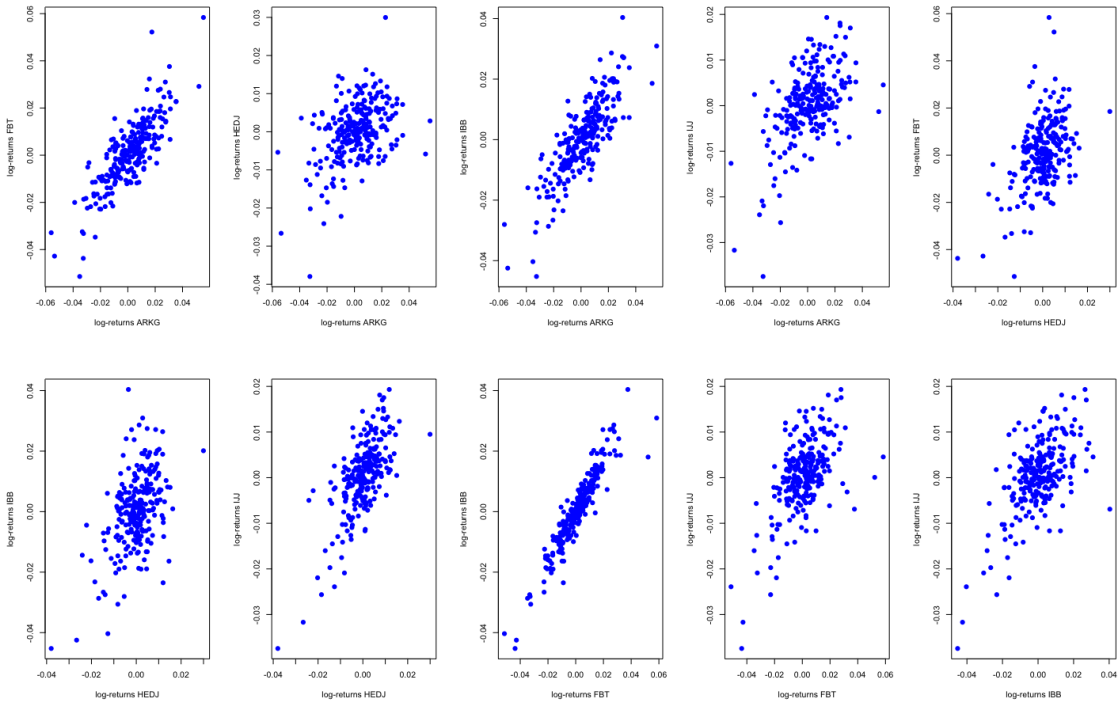

Figure 9.5: Scatter plots of the ETF margins against each other.

Although there are few observations in the tails, the points are relatively close to each other for very negative values. This suggests that the Clayton copula may be more suited than the other copulas.

From this exploratory analysis, we can already conclude that the VaR and ES estimated using simplistic methods may be preyed to underestimate the risk underlying the ETF portfolio. This will be ascertained in the next subsection.

9.4 Presumed high risky products: the Cryptocurrencies

9.4.1 Why to choose cryptocurrencies as high risk profile

Cryptocurrencies have been chosen because of their current craze. Crypto values are known to be very volatile. They can therefore be considered as risky assets. As such, we believe to put ourselves in the shoes of a non-risk averse manager. We especially constitute a portfolio made out of four currently leading cryptocurrencies, i.e. Bitcoin, Ethereum, Litecoin, and XRP. We retrieved 2 years of historical prices for the four cryptocurrencies constituting the portfolio, going from 05-2017 to 06-2019.

9.4.2 Exploratory analysis of an equally weighted cryptocurrencies portfolio

1. **Broad view of the behavior of the financial series:**

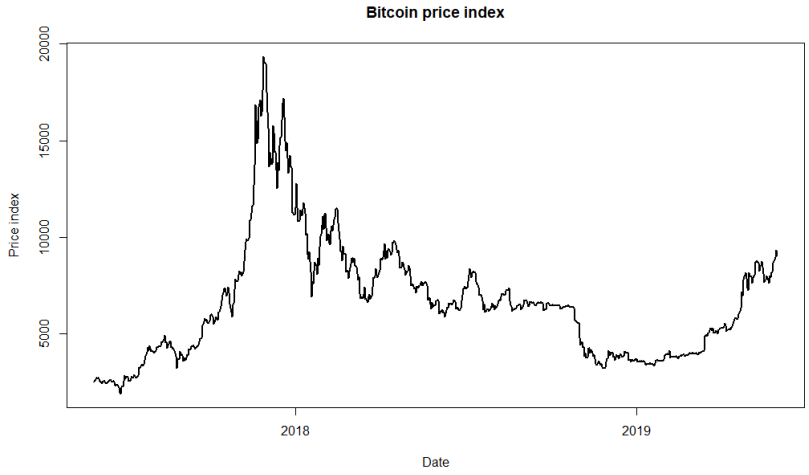

Figure 9.6: Price index of the Bitcoin against the dollar over the period from 05-2017 to 06-2019. The *Bitcoin bubble* is visible just before 2018.

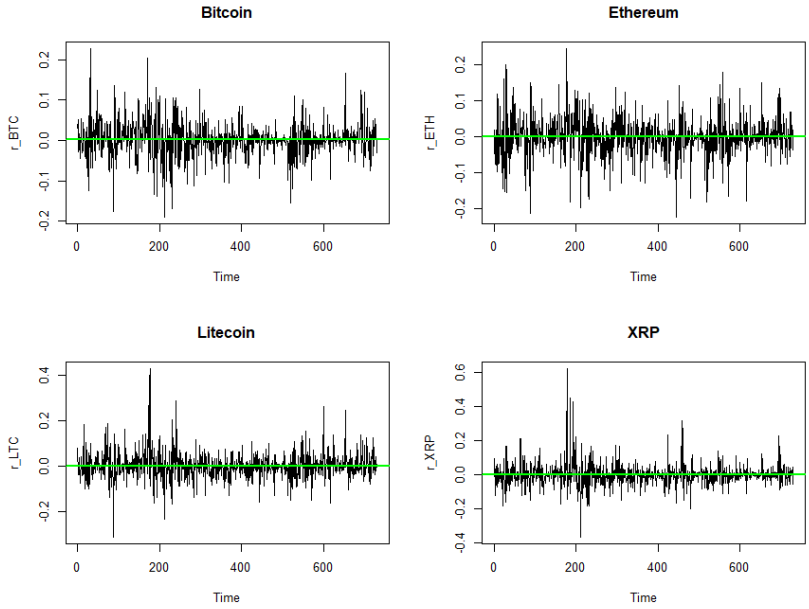

Figure 9.7: Time series of log-returns of the (a) Bitcoin, (b) Ethereum, (c) Litecoin, and (d) XRP for the period from 05-2017 to 06-2019. The mean appears in green.

The prices of cryptocurrencies ("crypto") have fluctuated considerably past years, and especially in the period from 2017 to 2018 where the cryptocurrencies market witnessed

significant shocks in line with the so-called *Bitcoin bubble* (Antonio Nunez et al., 2019). As illustrated in Figure 9.6, whereas the price of Bitcoin exceeded $19,300 in Dec 2017, its value was more than halved to $7000 in Feb 2018 (Lu, 2018). Historical log-returns have been derived from and are illustrated in Figure 9.7.

As it can be seen, considerable variations in the log-returns have occurred, and in particular around the 200^{th} observation that constitutes the bubble. The equally weighted portfolio made of the four crypto depicts the following time series of risk-factor changes:

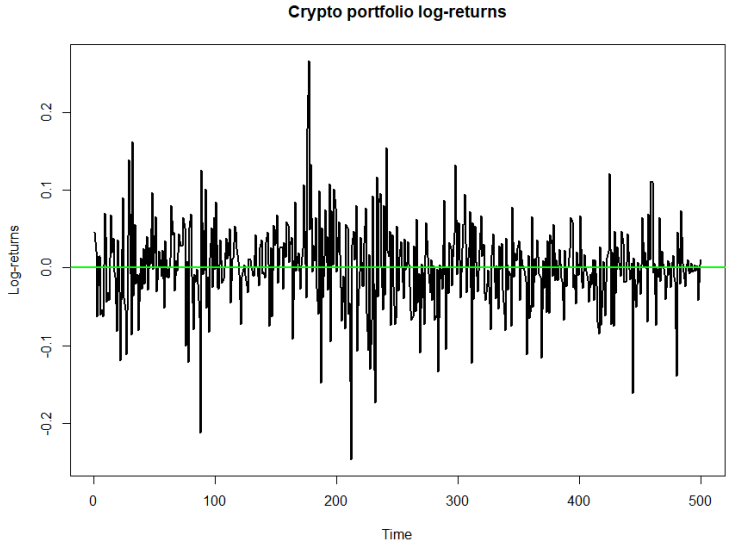

Figure 9.8: Time series of log-returns of the crypto equally weighted portfolio. The mean appears in green.

The riskiness of the cryptocurrencies is thus mainly driven by a huge variance, which means that their prices can change dramatically over a short time period in either direction. This is partially explained by the huge uncertainty about the crytpo future's value and the perception of their intrinsic ability to be a store value of money, their sensitivity to bad news, and the concerns about security exceedances, among others.

2. **Stationarity and auto-correlation analysis:**

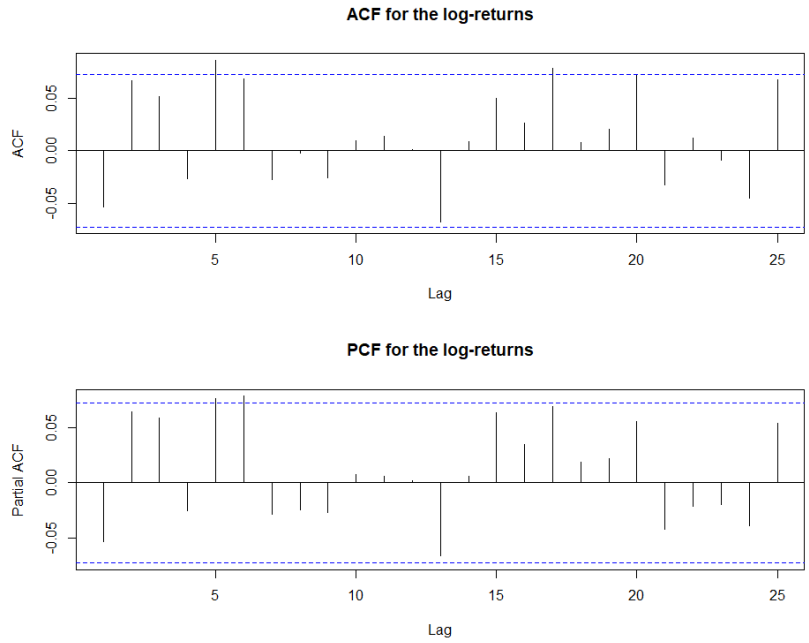

Figure 9.9: ACF and PCF of the log-returns of the crypto portfolio.

The plot of ACF and PCF of the portfolio's log-returns in Figure 9.9 shows little autocorrealtion. This may modestly indicates the presence of volatility clustering. In that context, the benefit from a GARCH specification in the model may be more substantial than in previous cases.

3. **Distribution tests:**

   The four first moments of the log-returns of the crypto portfolio are given by:

Descriptive measures	Values (daily)
Mean	0.0005
Standard deviation	0.051
Skewness	-0.11
Kurtosis	5.68

Table 9.2: Descriptive statistics for the crypto portfolio's log-returns.

   As expected, the measure of volatility is extremely high compared to common financial series. The average return is relatively high as well. It is to be noted that the crypto portfolio is very slightly skewed, and it is characterized by a greater kurtosis than normal, thereby producing larger tails. Besides, the null hypothesis of the Jarque Bera test supposing the normality of the series has been strongly rejected, as the p-value < 2.2e-16.

4. **Dependence structure assessment:**

No detailed analysis on the dependence structure between the cryptocurrencies has been carried out. Since all cryptocurrencies considered have been affected by the burst of the crypto. bubble (led by the Bitcoin), we expect a stronger dependence in the lower log-returns that would foster the use of the Clayton copula.

9.5 Backtesting results and observations

9.5.1 ETF portfolio

In the context of EVT, given the predominance of the GPD over the GEV, we decided to only use the former. We fitted a GPD over the ETF portfolio, which can be found in appendix A.2.3, and concluded that a threshold of 0 is adequate. Table 9.3 provides the backtesting results for the selected estimation models.

Estimation methods	VaR failures	ES failures	Basel zone	Average daily error (%)
HS	5	5	Amber	1
HS-GARCH	5	5	Amber	1
Risk Metrics	8	8	Amber	2.2
GARCH	8	8	Amber	2.2
MC	9	8	Amber	2.6
MC jump	4	3	Green	0.6
Clayton copula	9	8	Amber	2.2
Gumbel copula	17	16	Red	5.4
Gaussian copula	12	12	Red	3.8
Student copula	12	12	Red	3.8
GPD	1	1	Green	0.6
GARCH-GPD	1	1	Green	0.6
Clayton-GPD	2	2	Green	0.2
Econ. Unbiased	9	7	Amber	1.8
Econ. Unbiased-GARCH	6	6	Amber	1.4

Table 9.3: Number of exceedances of the 99% VaR and 97.5% ES estimates calculated using various estimation methods, for the **ETF portfolio's log-returns**. Expected failures=2.5.

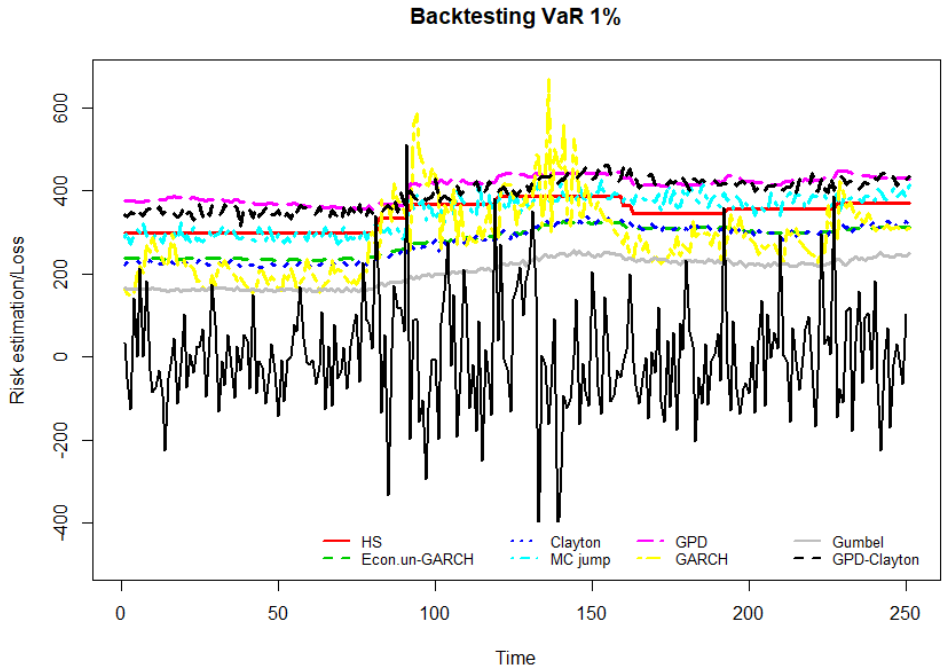

Figure 9.10: Backtesting of the **ETF** portfolio: Daily losses for the backtesting period with risk measure estimate $VaR_{1\%}$, and exceedances for the methods in the legend. All backtesting results for VaR and ES to be found in Appendix A.1.3.

The performance of the ES is similar to the VaR over the ETF portfolio. The ES produces less exceedances for some methods, indicating a higher severity that may meet the favor of the regulator. Only four estimation methods are considered *green* according to Basel, and as many are suitable for the bank.

Generally speaking, the performance of the conventional methods decreases considerably. For example, the HS is a robust estimation method so that it is unable to integrate the market change that occurs in the hold-out period. It is inadequate in periods of volatility fluctuation. One may also remark that the performance of the copulas, however to a lesser extent for the Clayton copula, is poor. The greater amplitude of negative log returns in the out of sample is poorly accounted for. The Clayton copula gives superior results since it naturally inflates the lower tail. It also enlightens the presence of dependence for bigger losses. The Gumbel is the worse performing copula.

Moreover, as expected, the GARCH specification does not benefit in terms of number of exceedances. Besides, as shown in yellow in Figure 9.10, it is extremely fluctuating, which is not desirable in a bank's perspective.

In contrast, the EVT sophistication produces the most satisfying results. It may fulfill the regulator and risk managers interests at the same time since it almost reaches the expected number of failures. The GPD appears to be highly valued when lower returns populate the

out of sample period. It is the only method that is able to respond the losses occurring across the backtesting, and should therefore be preferred when outliers are recently observed. For example, taking a $\$1,000,000$ worth portfolio in ETF at that period, and a maximal loss of $\$51,059$[2], table 9.4 shows the non absorbed loss[3] for each method. As expected, unlike other risk estimation methods, the GPD absorbs most of the shock.

Estimation methods	Risk estimate	Non absorbed loss
HS	$\$38,727$	$\$12,332$
Gumbel	$\$25,683$	$\$25,376$
Clayton	$\$34,012$	$\$17,047$
GPD	$\$44,875$	$\$6,184$
GPD-Clayton	$\$46,180$	$\$4,879$

Table 9.4: Non absorbed loss over a $\$1,000,000$ worth portfolio in ETF, and a maximal loss of $\$51,059$.

9.5.2 Cryptocurrencies portfolio

As for the ETF portfolio, the GDP is ideally fitted for a threshold of 0. Refer to appendix A.2.4 for details. Table 9.5 provides the backtesting results for the selected estimation models and Figure 9.11 shows the estimated VaR with regard to the realised losses in the backtesting period.

[2] Around the 90^{th} observation in Figure 9.10

[3] Non absorbed loss = Effective loss incurred - Estimated economic capital

Estimation methods	VaR failures	ES failures	Basel zone	Error
HS	2	2	Green	0.2
HS-GARCH	2	2	Green	0.2
Risk Metrics	7	7	Amber	1.8
GARCH	7	7	Amber	1.8
MC	2	2	Green	0.2
MC jump	1	0	Green	0.6
Clayton copula	2	2	Green	0.2
Gumbel copula	3	3	Green	0.2
Gaussian copula	2	2	Green	0.2
Student copula	2	2	Green	0.2
GPD	0	0	Green	1
GARCH-GPD	0	0	Green	1
Clayton-GPD	0	0	Green	1
Econ. Unbiased	2	2	Green	0.2
Econ. Unbiased-GARCH	2	2	Green	0.2

Table 9.5: Number of exceedances of the 99% VaR and 97.5% ES estimates calculated using various estimation methods, and for the series of **crypto portfolio's log-returns**. Expected failures=2.5.

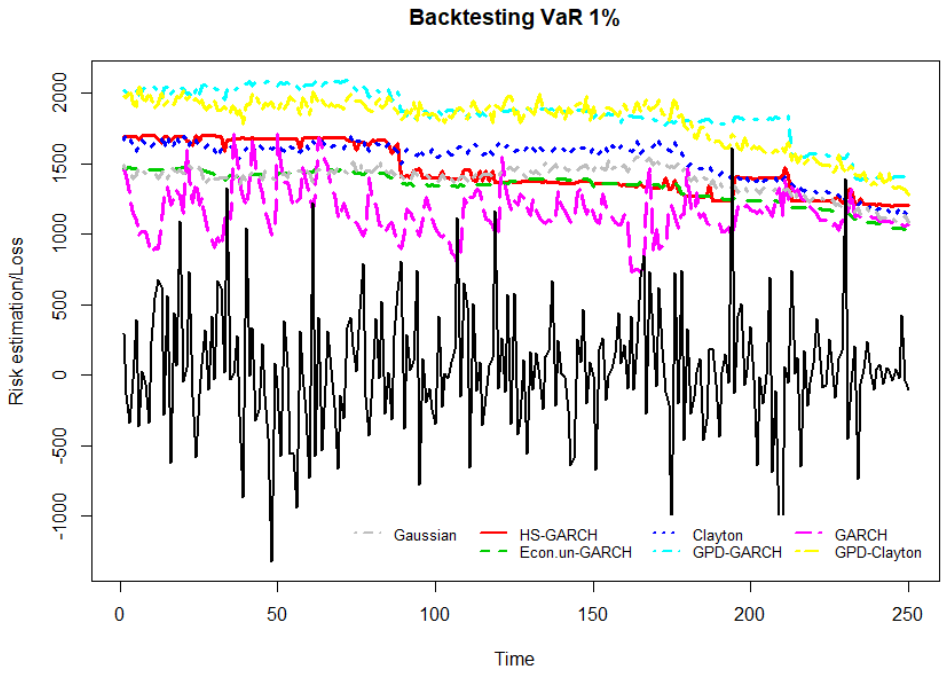

Figure 9.11: Daily losses for the backtesting period with risk measure estimate $VaR_{1\%}$, and exceedances for the methods in the legend. All backtesting results for VaR and ES to be found in Appendix A.1.4. *Portfolio: Cryptocurrencies*.

Once more, the same equivalence between the $VaR_{1\%}$ and $ES_{2.5\%}$ is observed. Moreover, it can been seen that the Basel committee would approve most of the models while the bank would be tempered. At the first sight, except the RM, GARCH and EVT specifications, all the methods seem suitable.

More concretely, it is first observed that the GARCH process is detrimental. There is not sufficient volatility clustering in the series. The RM approach, that also models conditional variance, poorly performs. In addition, while it was initially not expected, the GPD specification produces a large overestimation of the risk (no exceedances), what results in unreasonably high economical capital. Whereas it may be wished by the regulator, it is inadequate for the bank. This finds essence in the exploratory analysis. We showed the presence of the *cryptocurrencies bubble*'s burst within the historical series in which many extreme returns were observed. This has naturally inflated the left tail of the distribution. It is even more inflated by the GPD fitting, which results in a risk overestimation that is detrimental for the management of the bank.

We observe that the conventional methods, as for example the HS and MC, pass the backtesting phase. This can also be explained by the presence of more extreme log returns in the in sample than in the hold-out period. We would however stress the necessity to remain extremely attentive to the behaviour of the crypto-currencies. The steps done in the exploratory analysis should act as indicators to changes in the market conditions that can, as earlier shown, dramatically change the models performance and turn some estimation methods out of relevance. This is even more true as cryptocurrencies are known unstable.

Then, the copula specification seems to work well. In light of the (potential) instability of the portfolio, we would however opt for the Clayton copula as the reference. It indeed performs well in the backtesting period, and is known to be the most reliable in case of big losses. A comparison between the Clayton and Gaussian copula in Figure 9.11 also shows the superior severity of the Clayton copula, because it inherently inflates the lower tail of the P&L. It is also interesting to note how the copula estimates are narrowing as far as the outliers disappear from the estimation sample.

As a last important result, the Economically unbiased VaR and ES perform good in the backtesting phase. The same reasons as mentioned just above applied here. To go a bit further, we have studied the influence of the window size (on which the model is fitted) on its backtesting performance against its plug-in counterpart (see Figure 9.12). We point out that the $VaR^{\text{plug-in}}$ frequently produces more failures than the Econ. unbiased VaR until a 200 window size. The ES estimation is steadily higher. It means that the Econ. unbiased form has no prominent advantage over a 250 backtesting time window here, but is noticeably better the smaller the window size gets. It gives another argument to use the economically unbiased estimators because it requires fewer observations to deliver good results in backtesting. Refer to Appendix A.3 for a similar analysis regarding the ETF portfolio.

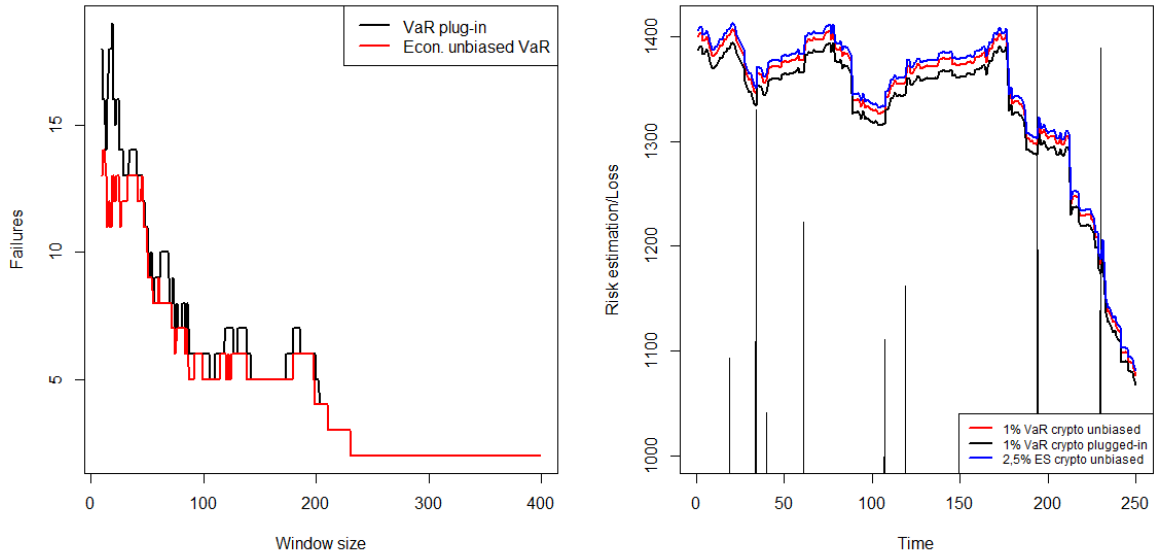

Figure 9.12: A comparison between the $VaR_{1\%}^{\text{Plug-in}}$ and the Economically unbiased $VaR_{1\%}$ in backtesting for the **crypto. portfolio**.

9.6 Conclusions

The ETF and crypto. portfolios analysis and backtesting have brought an interesting learning experience. They showed that the normality assumption is not sustainable in real cases, even for ETF's log-returns that are theoretically prone to the acceptance of the Gaussian assumption. This illustrates the general sophistication of the markets.

We mostly observe that a greater number of risk estimation methods are acceptable in the crypto. case than it is for the ETF. This brings us to a crucial ascertainment: it is the market regime variation that mostly drives the relevance of one risk estimation technique. If a surge in the negative log returns is to be observed, it is likely that a sophisticated method like the EVT will be extremely required. If the regime of the out of sample is safer than the in sample, sophistication (EVT essentially) brought to the estimation may lead to risk overestimation. The study shows that since the risk of the ETF portfolio increases in the out of sample, few estimation methods are acceptable, and the use of the EVT is required. A greater number of methods are acceptable for the crypto. portfolio because of the inclusion of the crypto bubble that produces large negative log-returns within the historical series, never reached in the out of sample. In that aspects, we argue that the data exploratory phase is shown to be extremely useful to assess the riskiness of the portfolio and determine the most appropriate RMP to use.

Finally, it can be mentioned the slight outperformance of the ES relative to the VaR.

Conclusion

Considering the woes of the financial actors in QRM before the Global financial crisis, we examined ways to avoid this happening again. Financial institutions often resorted to simplistic models to quantify risk in a rapidly changing environment. This had greatly underestimated the risk. That is what led us to this thesis, in which we adopt the risk manager view within the regulatory framework. To investigate the best current risk measures, we confronted the use of the ES against the VaR as the standard risk indicators with a systematic comparison to the variance. We also examined if the sophistication of the models, via the copulas and the EVT always provide (more) satisfying results when the portfolio riskiness varies. In the theoretical perspective (Part I), we highlighted the Basel framework in which the ES should continue to gain ground over the VaR because of its theoretical properties. The former indeed better considers the tail-risk, and is coherent while the latter is not. Furthermore, the ES biggest presumed issue, non-elicitability, does not actually prevent its backtesting. Moreover, its higher sensitivity to new data is not really an issue because it might help foresee future large unexpected losses. It however remains important to provide a procedure as simple and elegant as the VaR backtesting procedure. In Part II, we investigated the influence of the portfolio riskiness through a practical analysis conducted in a two times procedure (i) we first executed risk estimations over a low risk and high risk simulated log-returns, respectively normally and Johnson distributed, (ii) we then applied the same risk estimation methods over real portfolios, composed of ETF and Cryptocurrencies, that theoretically reflect the two different risk profiles. The performance of the models was compared in term of the number of exceedances over a 250 backtesting window, both in the view of the regulator and the bank. It reasonably lets us think that, if sophisticated risk estimation methods, like the EVT, would have been more considered during crisis periods, financial institutions may have been able to parry extreme losses thereby avoiding the disastrous repercussions on the economy. The Clayton copula is the most flexible sophistication considering low and high risk changing environments. The conventional methods mainly suffer from environment changes or turbulence. Overall, **it is the market regime variation that mostly drives the relevance of one risk estimation technique**.

The results that gave us this opinion are here summarized:

- The ES entails more theoretical desirable properties than the VaR: (*i*) it better accounts for tail-risk and answer the important *what-if* question, ensuring a greater risk sensitivity (*ii*) it satisfies the subadditivite axiom and is a coherent risk measure. Despite its proven non-elicitability, it is now proven to be backtestable. Also, we demonstrated that at the level 2.5%, the stability of the ES estimate is similar to the $VaR_{1\%}$. This supports the move to an ES-based Basel risk assessment procedure. Empirically, we showed an equivalence in the backtesting results between the $VaR_{1\%}$ and $ES_{2.5\%}$, both generally producing the same number of exceedances. The $ES_{2.5\%}$ has even delivered fewer exceedances in higher risk environments, thereby mitigating potential losses.

- The Economical unbiasedness property of the risk estimator has been shown fully effective within the Gaussian framework in theory. It offers levels of performance superior to the Gaussian plug-in VaR that systemically underestimates the risk. Nonetheless, the Gaussian plug-in VaR is economically consistent to the economical unbiased VaR. As a result, the latter offers good performance regardless the sample size. However, in the Basel framework, for which the time window is 250 (trading) days, the difference of estimate is so small that the performance is similar in the backtesting phase.

- The parametric methods of VaR and ES calculation based on the normal assumption are largely outperformed by the non-parametric methods. Normality does not offer a proper consideration of the tail-risk inherent to financial series. To this regard, the Risk metrics and GARCH methods, that rely on a conditional form of volatility, should be considered with a *t* or *Johnson* innovation. That may allow a better response to market changes and avoid too many exceedances, as shown by McNeil et al. (2005). In general, the GARCH specification can produce estimations closer to the financial series that would enable the risk manager to add no more than the economic capital required.

- The historical simulation approach, quite surprisingly, performs relatively well in the low and high risk simulations. This suggests that it is worth using in steady conditions of the market. However, in the context of a rise in market riskiness, as illustrated in the ETF case, the HS poorly performs. This also shows the importance of the exploratory analysis of the portfolio series.

- The Clayton copula is generally the copula to advocate, given its flexibility to the market risk environment. The strong dependence it induces in the lower tail allows a better consideration of the tail-risk. All other copula cannot achieve that standard of performance. Generally speaking, the performance of the Clayton copula is superior in period of higher risk. It provides less benefits when the risk is low. Notably it can poorly respond outliers in the out sample period if there were not in the in sample. In that aspect, the EVT is a great alternative/complement. In particular, **the Generalized Pareto distribution gives the most conservative risk estimations** among all the methods considered. It produced the least

number of exceedances. From a regulatory point of view, it should be the favourite method. On the other hand, when dealing with stationary time series where no increase in volatility is expected in the near future, it leads to risk overestimation. This may be detrimental for the bank because it diminishes its performance. If early adopted, it would however preserve the bank from a crash if adverse market movements were to be encountered. Overall, the EVT must be used when one may expect large losses in the near future, or during market stress (like in 2008 and beyond) since it would be the only method to respond properly.

Limitations and further possible research

It is to be noted that the result of the thesis are dependent on the financial time series selected and simulated. The results could have been different within a different setting. Also, we conducted our estimations over 250 trading days that better fits the regulatory framework. Extending the time window of the in sample could have led to different conclusions. In particular we argue that a short time window does not allow to properly differentiate the reactivity of the VaR and ES to high volatility periods since the $VaR_{1\%}$ only leaves 2 observations apart in the setting considered. Over more data, we may expect a poor risk sensitivity of the VaR compared to the ES, as a result of their different consideration of the tail. This should be explored.

We end up this thesis with several future research suggestions that are worth investigating:

1. One could think about including Machine learning principles to improve the predictive effects of the models, like the Markov Chain Monte-Carlo and Long short term memory (deep learning) methodologies.

2. In our effort to model stochastic volatility which manifestly help improve the estimations in certain circumstances, whereas we have confined our interest to rather simple GARCH models, it would be judicious to add asymmetric innovation to GARCH in order to give more weights to negative log-returns.

3. The economical unbiasedness property could also have been examined for different levels of standard deviation, and for all other estimation methods presented in this work. This might be conducted in future research.

4. The number of assets constituting the portfolios could be increased. This may bring diversification benefit and refine distributions. This may also challenge the power of the Copula to capture the whole dependence structure between the assets. However, this is to the price of a higher computer intensity.

5. We point out that there exists always room to sophisticate the models: e.g. dynamic copula, t-innovation in the parametric model, etc. These methods could also be considered but would have extended the scope of this thesis too large.

6. The *expectile*, introduced in chapter 1, despite its poor interpretation in term of capital requirement, is the only risk measure to be coherent and elicitable at the same time. Its future in the field of banking regulation is not yet clear, but is certainly worth being investigated.

APPENDIX A

APPENDIX

A.1 Results of the backtesting

A.1.1 Normal-distributed portfolio

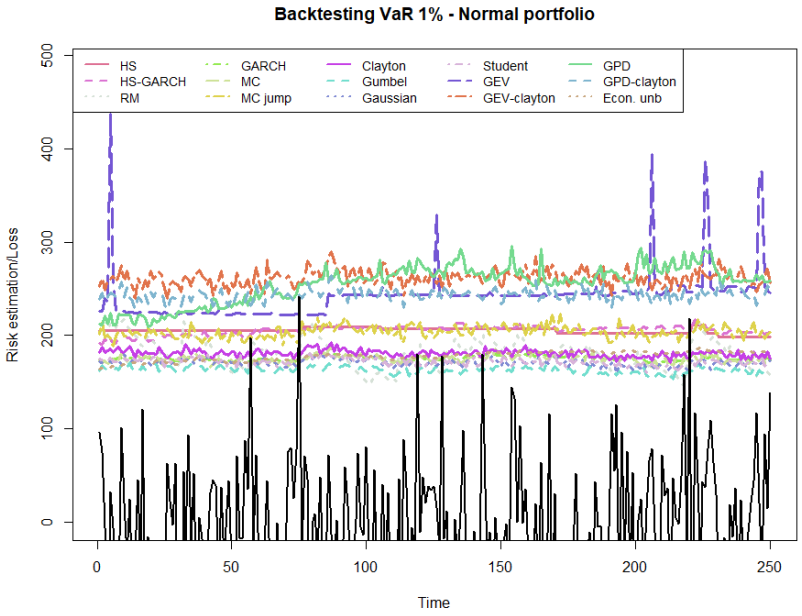

Figure A.1: Backtesting results of the $VaR_{1\%}$ for all the considered risk estimation methods. Only the negative log-returns (loss side) are plotted. While the EVT-based methodologies tend to overestimate the risk, the difference in term of the number of failures between the other methods is quite narrow. We can also point out the irregularity of the GEV, which can be assumed to be a noticeable pitfall of the method.

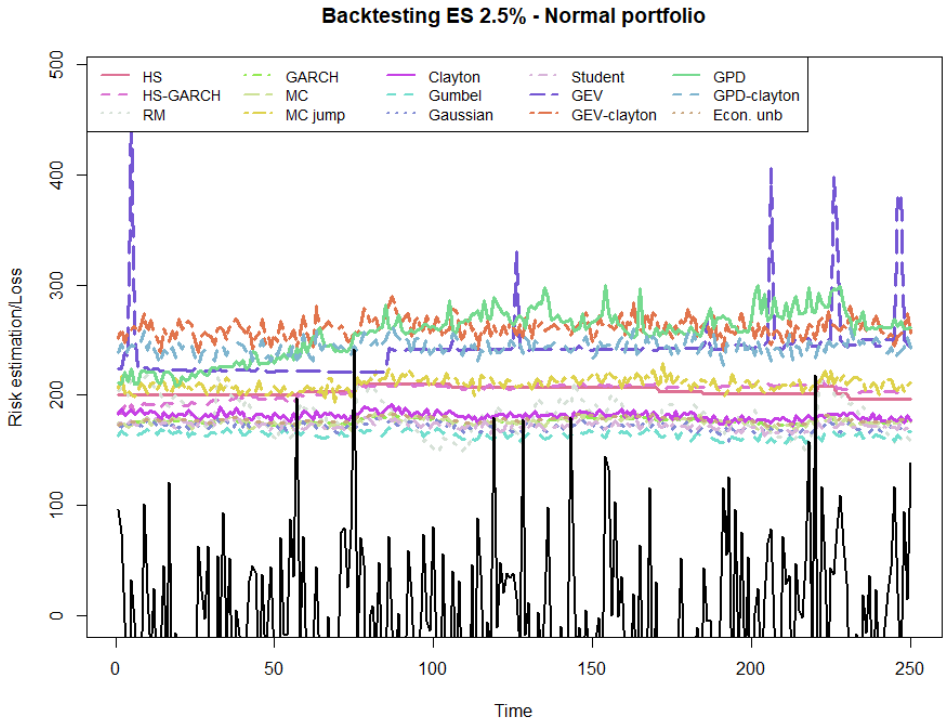

Figure A.2: Backtesting results of the $ES_{2.5\%}$ for all the considered risk estimation methods. Only the negative log-returns (loss side) are plotted. The results obtained for the expected shortfall at the 2.5% significance level are very close to the VaR 1%. We may therefore argue that the higher severity of the ES (α can still be increased) is desirable, as well as its relevant properties.

A.1.2 Johnson-distributed portfolio

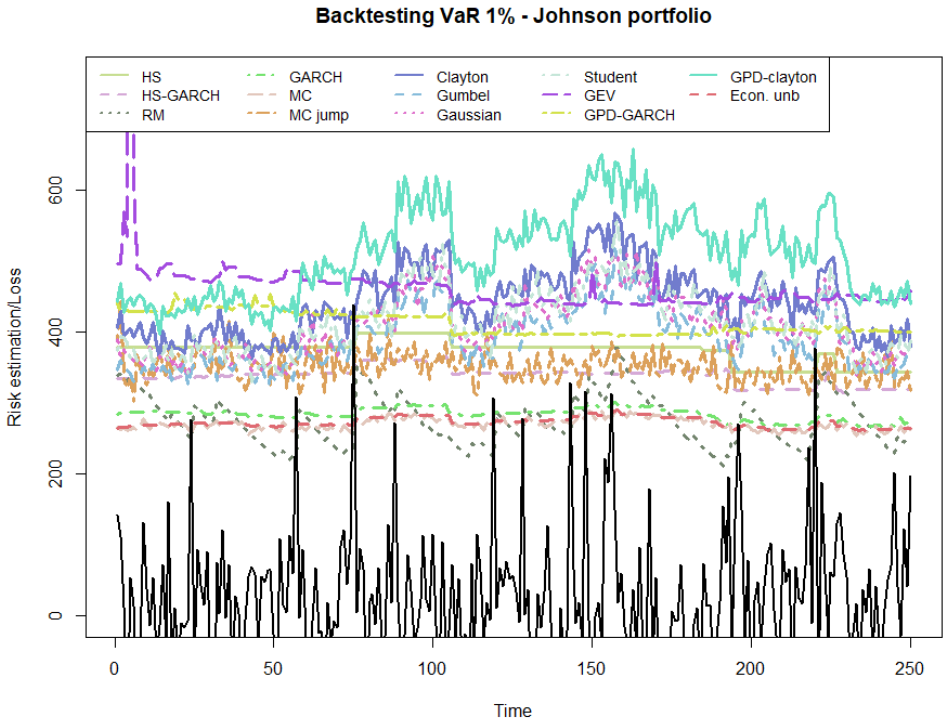

Figure A.3: Backtesting results of the $VaR_{1\%}$ for all the considered risk estimation methods. Only the negative log-returns (loss side) are plotted. We remark that the different estimation methods are more scattered than for the normal log-returns. The riskiness is a key determinant for the relevance of certain methods. The copula are good performer, with the Clayton providing a superior estimation, followed by the t, Gaussian, and Gumbel copulas in order. Parametric methods on the behalf of the RM, GARCH and MC produce poor results. The EVT is the most conservative, prudential method.

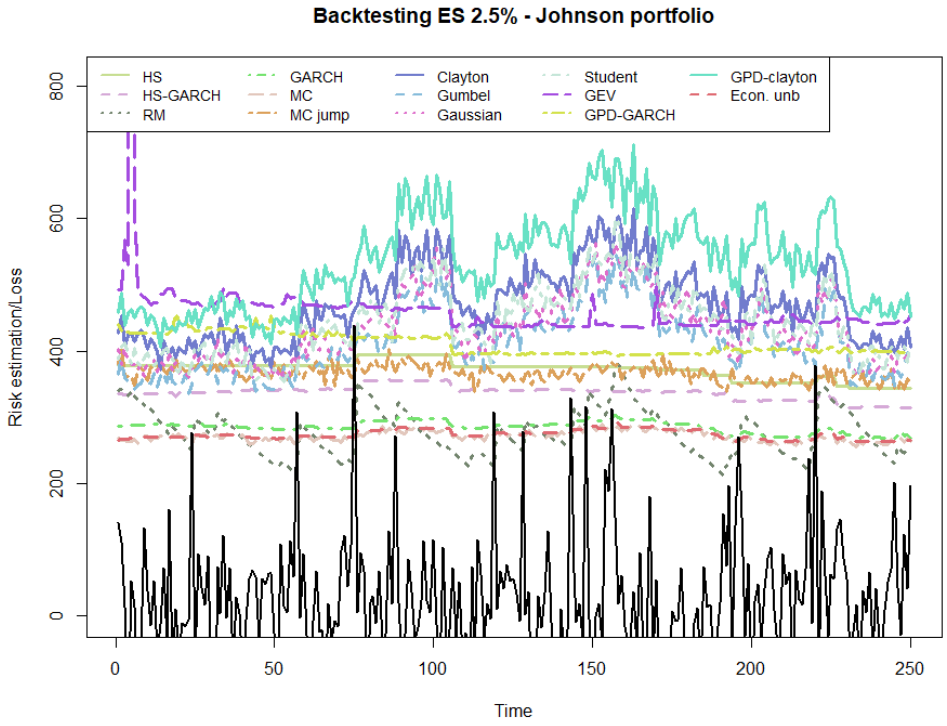

Figure A.4: Backtesting results of the $ES_{2.5\%}$ for all the considered risk estimation methods. Only the negative log-returns (loss side) are plotted. The results are similar to those of the VaR 1%. The estimations may, however, sometimes be slightly greater. This is inherent to the behavior of the ES that looks further in the tail. Tail where we find more low returns in this portfolio.

A.1.3 ETF portfolio

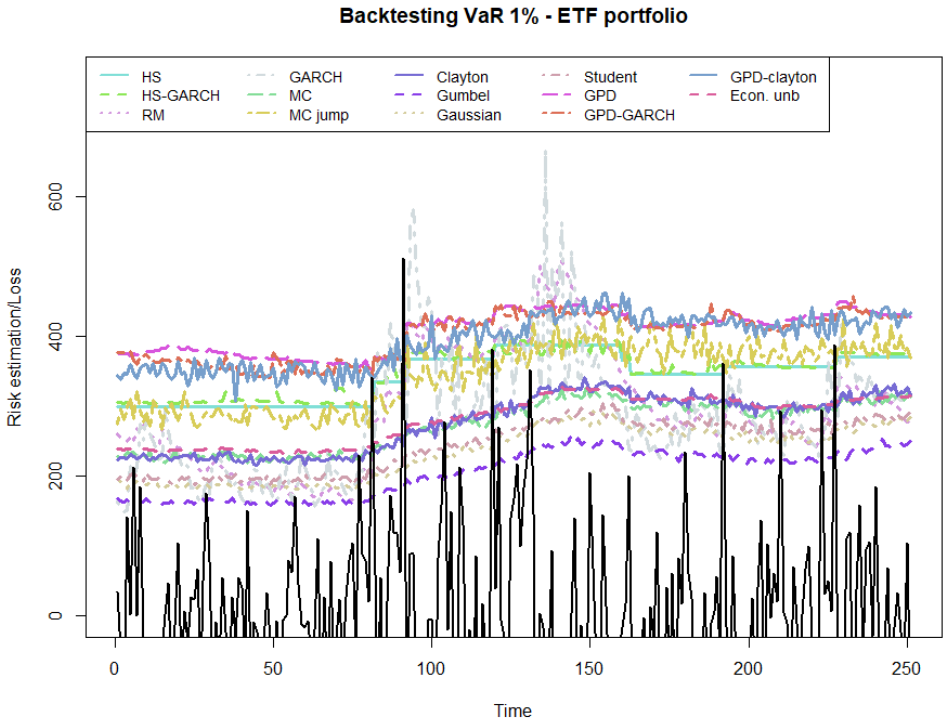

Figure A.5: Backtesting results of the $VaR_{1\%}$ for all the considered risk estimation methods. Only the negative log-returns (loss side) are plotted. The variability out of sample symbolized by the few peaks around the 100^{th} and 200^{th} observations leads to a powerful EVT (GPD) that outperforms all other estimation methods. It is also interesting to note the superior performance of the Clayton copula compared to the other copula. The difference is more pronounced here. The GARCH and RM are found very reactive to market changes but produce many hits since there is no clustering of volatility. Parametric methods are again bad performers, while the HS(-GARCH) and MC with jump are satisfying.

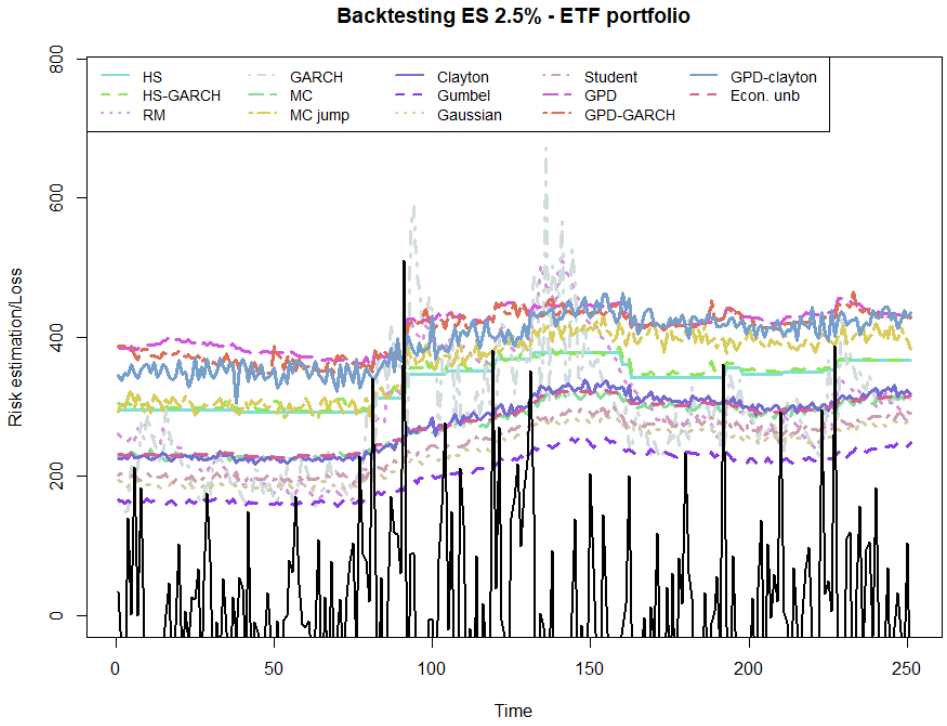

Figure A.6: Backtesting results of the $ES_{2.5\%}$ for all the considered risk estimation methods. Only the negative log-returns (loss side) are plotted. The backtesting results are very close to those of the VaR 1%.

A.1.4 Cryptocurrencies portfolio

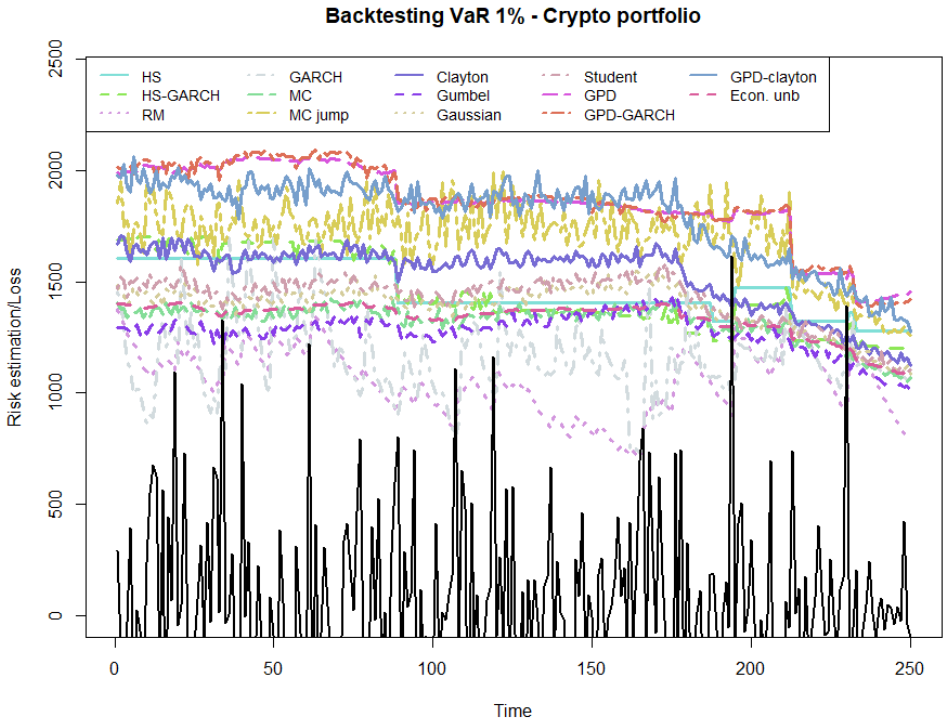

Figure A.7: Backtesting results of the $VaR_{1\%}$ for all the considered risk estimation methods. Only the negative log-returns (loss side) are plotted. The taking into account of the *Bitcoin bubble* produces high estimations for, more or less, all the methods. The GPD remains the methods associated with the lowest number of failures, however slightly overestimating the risk in this case (the adjustment at the end is good point). The superiority of the Clayton copula against other copula is even more established here. The Gumbel copula is pretty low. It is also interesting to highlight how the copula estimates are narrowing as far as the outliers disappear from the estimation sample.

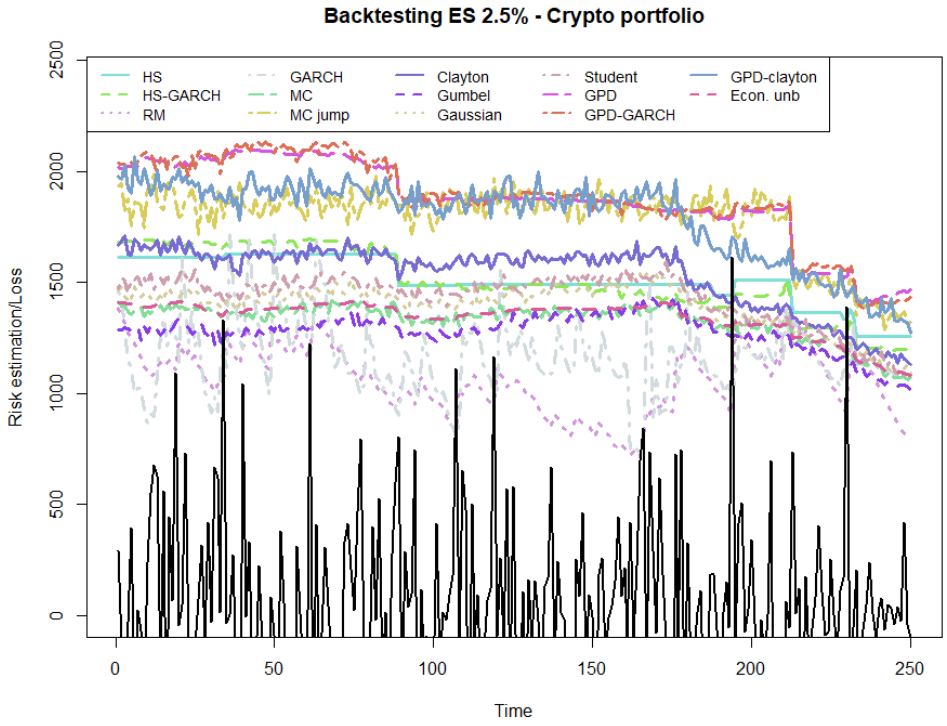

Figure A.8: Backtesting results of the $ES_{2.5\%}$ for all the considered risk estimation methods. Only the negative log-returns (loss side) are plotted. The backtesting results are very close to those of the VaR 1%.

A.2 GPD fitting

A.2.1 Normal-distributed portfolio

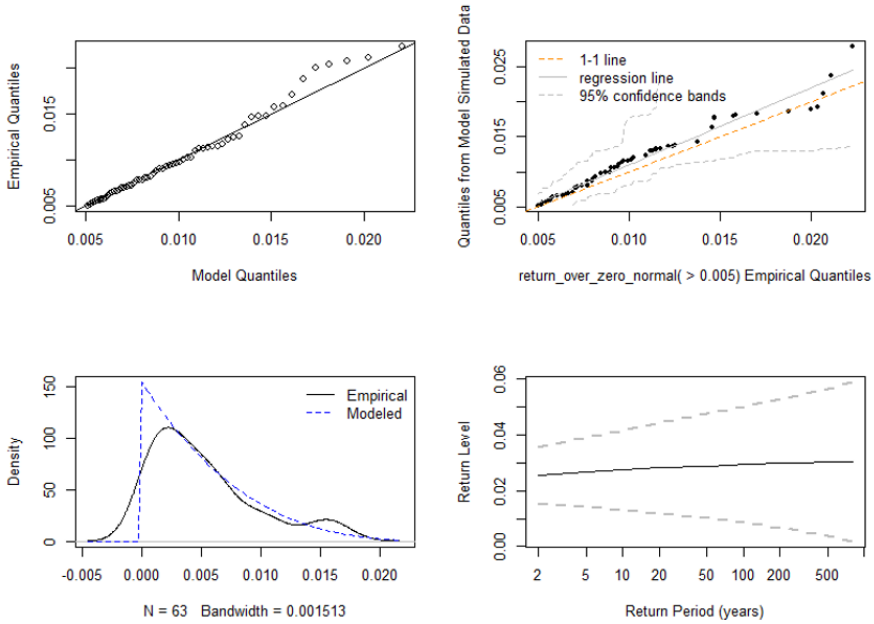

Figure A.9: Diagnostic plots of fit for the GPD with Normal log-returns. The threshold $u = -0.005$.

A.2.2 Johnson-distributed portfolio

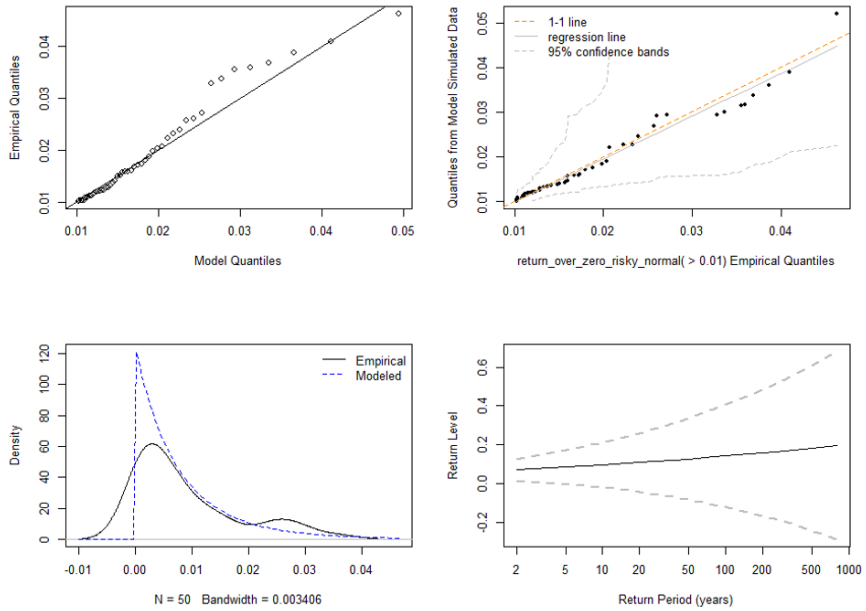

Figure A.10: Diagnostic plots of fit for the GPD with Johnson log-returns. The threshold $u = -0.01$.

A.2.3 ETF portfolio

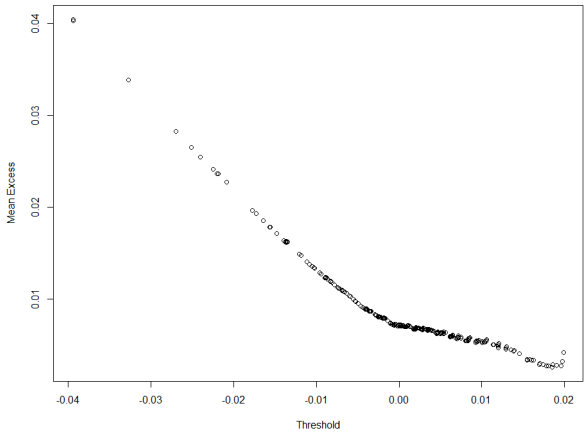

Figure A.11: Mean excess plot for the ETF's log-returns.

The mean excess plot suggests a threshold to be set at 0. This seemingly provides are rather good fit for the GPD on the ETF portfolio.

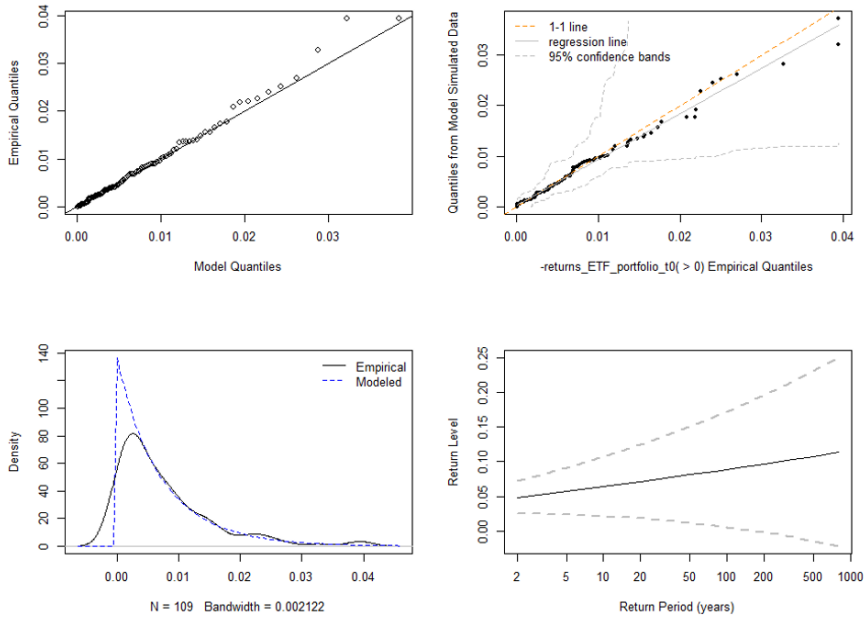

Figure A.12: Diagnostic plots of fit for the GPD with the ETF's log-returns. The threshold $u = 0$.

A.2.4 Cryptocurrencies portfolio

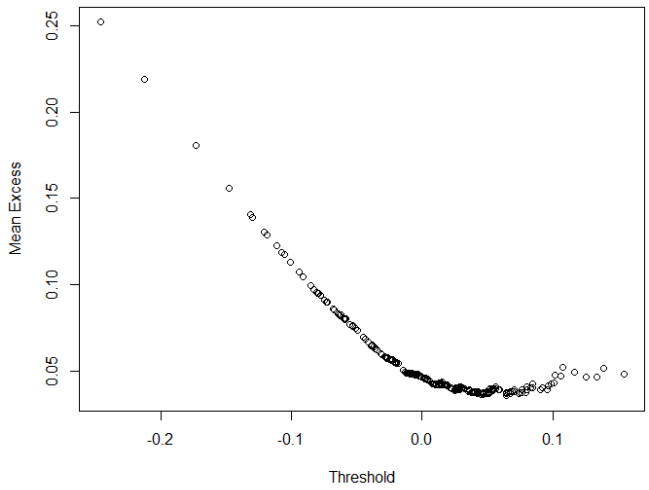

Figure A.13: Mean excess plot for the log-returns of the crypto portfolio.

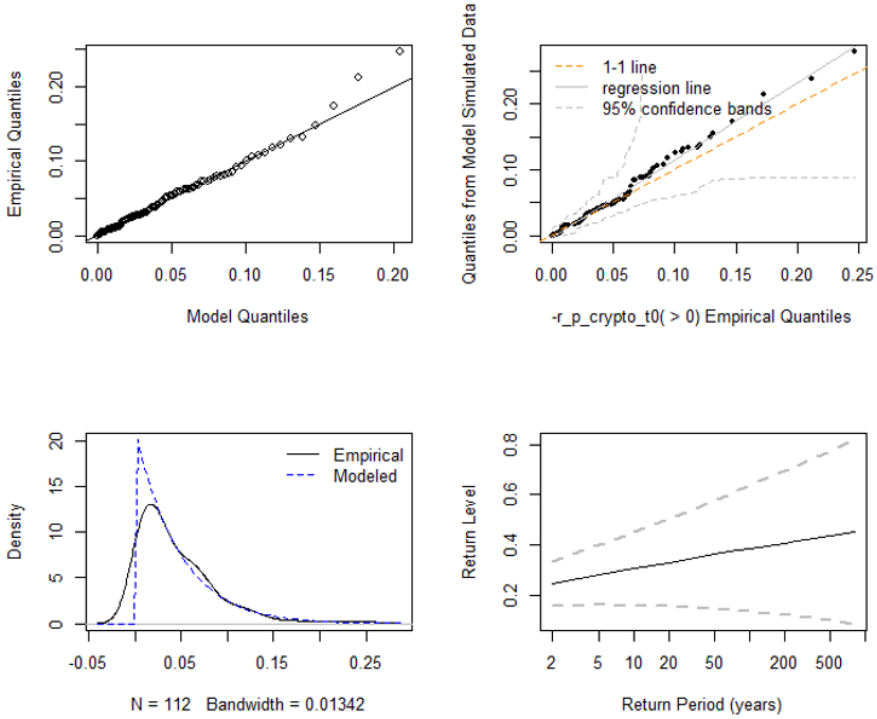

Figure A.14: Diagnostic plots of fit for the GPD with the crypto portfolio's log-returns. The threshold $u = 0$.

The mean excess plot provided in figure A.13 would suggest a threshold to be set around 0 in order to fit the GPD model. Doing so, it can be observed that the model is suitably fitted (see figure A.14), even if deviation at high quantiles occur (top left plot). This is however logic due to the scarcity of data at that levels.

A.3 Economical unbiasedness: sensitivity analysis for the ETF portfolio

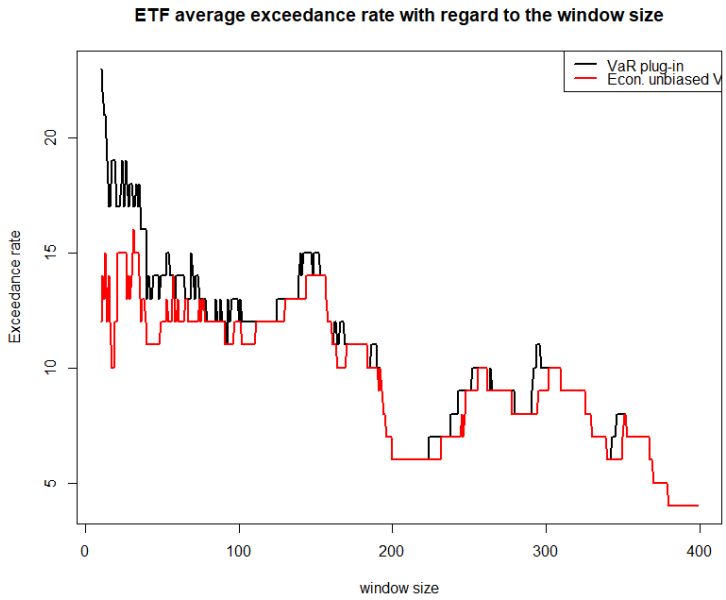

Figure A.15: A comparison between the $VaR_{1\%}^{\text{Plug-in}}$ and the Economically unbiased $VaR_{1\%}$ in backtesting for the **ETF portfolio**. Over an increasing time window size, from 10 to 400 days, we recorded the number of exceedances of the plug-in and econ. unbiased $VaR_{1\%}$, respectively. The results obtained are similar as for the crypto-currencies portfolio, whose analysis is provided 9.5.2. We observe that the plug-in VaR returns more violations as far as the time window considered for the backtesting is small. After what it converges to the economical form (around a window size of 200). Over a 250 days backtesting window, as per the Basel regulation, no real difference is to be noted.

A.4 Stability of the VaR and ES: samples mean and discrepancy with the true value

In the present appendix, we investigate the discrepancy between the samples mean and the true value of the risk measure considered. Table A.1 reports, for every time window size (number of historical observations) and risk measure considered, the difference between the mean over the sample and the true risk measure's value. We observe that the samples mean is very close to the true value, regardless the time window size and for a sample size fixed to 100.

		Time window size				
		50	250	500	1000	5000
RMP	$ES_{1\%}^{\text{plug-in}}$	-2.307	0.308	0.503	-0.942	-0.426
	$VaR_{1\%}^{\text{plug-in}}$ (a)	-2.042	0.273	0.438	-0.886	-0.405
	$ES_{2.5\%}^{\text{plug-in}}$ (b)	-2.051	0.274	0.441	-0.887	-0.405

Table A.1: Difference between the samples mean and the true value of the risk measure for the $ES_{1\%}^{\text{plug-in}}$, the $VaR_{1\%}^{\text{plug-in}}$, and the $ES_{2.5\%}^{\text{plug-in}}$ for different time window size.

BIBLIOGRAPHY

Aas, K. (2004). Modelling the dependence structure of financial assets: A survey of four copulas. *Norwegian Computing Center: Applied research and development*.

Acerbi, C. and Szekely, B. (2014). Backtesting expected shortfall. *Risk*, 27.

Acerbi, C. and Tasche, D. (2002). Expected shortfall: A natural coherent alternative to value at risk. *Economic Notes*, 31(2):379–388.

Alberg, D., Shalit, H., and Yosef, R. (2008). Estimating stock market volatility using asymmetric garch models. *Applied Financial Economics*, 18(15):1201–1208.

Alentorn, A. and Markose, S. (2008). *Generalized Extreme Value Distribution and Extreme Economic Value at Risk (EE-VaR)*, pages 47–71. Kontoghiorghes, E., Rustem, B., and Winker, P.

Ane, T. and Kharoubi, C. (2003). Dependence structure and risk measure. *The Journal of Business*, 76:411–438.

Ang, A. and Chen, J. (2002). Asymetric correlations of equity portfolios. *Journal of Economics*, 63(3):443–494.

Anosov, D. (1994). Measure space. *Encyclopedia of Mathematics*, 1(1).

Antonio Nunez, J., I. Contreras-Valdez, M., and A. Franco-Ruiz, C. (2019). Statistical analysis of bitcoin during explosive behavior periods. *PLoS ONE*, 14.

Artzner, P., Delbaen, F., Jean-Marc, E., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9:203 – 228.

Basu, D. (1955). *On Statistics Independent of a Complete Sufficient Statistic*, volume 4, pages 377–380. Sankhyā.

BCBS (2019a). The basel framework. *Bank for international settlements*.

BCBS (2019b). Explanatory note on the minimum capital requirements for market risk. Technical report, Bank of international settlements.

Berkowitz, J. (2001). Testing Density Forecasts, with Applications to Risk Management. *Journal of Business & Economic Statistics*, 19(4):465–474.

Berkowitz, J. and M. O'Brien, J. (2001). How accurate are value-at-risk models at commercial banks. *SSRN Electronic Journal*.

Bommareddy, S., Beleidy, S. E., Narapareddy, S., and Yoner, N. (2014). Applications of fat tail models in financial markets.

Bucevska, V. (2012). An empirical evaluation of garch models in value-at-risk estimation: Evidence from the macedonian stock exchange. *Business Systems Research Journal*, 4:1–16.

By, J.-J., H J Cai, J., Einmahl, L., and Haan, D. (2011). Estimation of extreme risk regions under multivariate regular variation. *The Annals of Statistics*, 39:1803–1826.

Chai, S. and Guo, C. (2011). Copula-based dependence analysis of u.s. stock index and futures time series in financial crisis. In *International Business and Management*, pages 927–935.

Cont, R., Deguest, R., and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, 10:593–606.

Cornuéjols, G. and Tutuncu, R. (2007). *Optimization Methods in Finance*. Mathematics, Finance and Risk (Book 5). Cambridge University Press, 1 edition.

Costanzino, N. and Curran, M. (2015). A simple traffic light approach to backtesting expected shortfall. *Risk*.

Curcio, R., I Anderson, R., and Guirguis, H. (2014). Stock price volatility of banks and other financials emanating from the inception of leveraged, inverse and traditional etfs. *The Journal of Index Investing*, page 140512050812007.

Danielsson, J. and al. (2005). Subadditivity re–examined: the case for value–at–risk. *RiskResearch*.

David Miles, J. Y. and Marcheggiano, G. (2012). Optimal bank capital. *The Economic Journal,*, 123:1–37.

Doğan, M. and Karakaş, A. (2017). Archimedean copula parameter estimation with kendall distribution function. *Journal of the Institute of Science and Technology*, 7:187–198.

Duffie, D. and Pan, J. (2001). Analytical value-at-risk with jumps and credit risk. *Finance and Stochastics*, 5(2):155–180.

Duță, V. (2018). Using the symmetric models garch (1.1) and garch-m (1.1) to investigate volatility and persistence for the european and us financial markets. *Studii Financiare (Financial Studies)*, 22(1):64–86.

Embrechts, P. and al. (2014). An academic response to basel 3.5. *Risks 2(1):25-48*.

Embrechts, P. and McNeil, A. (2000). Correlation and dependence in risk management: Properties and pitfalls. *Rev. Econ. Stat.*, 86.

Embrechts, P., McNeil, D., and Straumann, E. (1999). Correlation: Pitfalls and alternatives. *RISK*, pages 69–71.

Embrechts, P. and Wang, R. (2015). Seven proofs for the subadditivity of expected shortfall. *Dependence Modeling 3(1)*.

Emmer, S., Kratz, M., and Tasche, D. (2015). What is the best risk measure in practice? a comparison of standard measures. *Journal of Risk*, 18:31–60.

Evans, R. and Wilson, C. (2018). How etf became the market.

Francioni, I. and Herzog, F. (2012). Probability-unbiased value-at-risk estimators. *Quantitative Finance*, 12(5):755–768.

Fu, J. C. (1975). The rate of convergence of consistent point estimators. *Ann. Statist.*, 3(1):234–240.

Galdi, F., Leonel, M., and Pereira, L. (2007). Value at risk (var) using volatility forecasting models: Ewma, garch and stochastic volatility. *Brazilian Business Review*, 4:74–94.

Gilli, M. and këllezi, E. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics*, 27:207–228.

Gneiting, T. (2011). Making and evaluating point forecasts. *J. Amer. Statist. Assoc. 106 (494)*.

Hammersley, J. (1964). *Monte Carlo methods*. Methuen's monographs on applied probability and statistics. Methuen.

Haugh, M. (2016). An introduction to copulas. *Quantitative Risk Management*.

Hiusman, R., Koedijkand, K., and Pownall, A. (1998). Var-x: Fat tails in financial risk management. *Journal of Risk*, pages 47–61.

Hochrainer-Stigler, S., Pflug, G., Dieckmann, U., Rovenskaya, E., Thurner, S., Poledna, S., Boza, G., Linnerooth-Bayer, J., and Brännström, Å. (2018). Integrating systemic risk and risk analysis using copulas. *International Journal of Disaster Risk Science*, 9(4):561–567.

Holton, G. (2003). *Value-at-risk: Theory and Practice*. Academic Press advanced finance series. Academic Press.

Hull, J., White, A., and L Rotman, J. (2005). The perfect copula: The valuation of correlation-dependent derivatives using the hazard rate path approach. Technical report, University of Toronto.

Jorion, P. (2000). *Value at Risk: The New Benchmark for Controlling Market Risk*. McGraw-Hill Inc., second edition.

Jorion, P. (2006). *Value at Risk, 3rd Ed.: The New Benchmark for Managing Financial Risk*. McGraw-Hill Education.

Kerkhof, J. and Melenberg, B. (2003). Backtesting for risk-based regulatory capital. *Journal of banking & finance*, 28.

Kratz, M. (2017). Discussion of "elicitability and backtesting: Perspectives for banking regulation". *The Annals of Applied Statistics*, 11:1894–1900.

Kusuoka, S. (2001). On law invariant coherent risk measures. *RIMS Kokyuroku*, 3.

Liu, J. and Yang, X. (2012). The convergence rate and asymptotic distribution of the bootstrap quantile variance estimator for importance sampling. *Adv. in Appl. Probab.*, 44(3):815–841.

Lopes, R. H. C. (2011). *Kolmogorov-Smirnov Test*, pages 718–720. Springer Berlin Heidelberg, Berlin, Heidelberg.

Lu, L. (2018). Bitcoin: Speculative bubble, financial risk and regulatory response. *Butterworths Journal of International Banking and Financial Law (2018)*, 33:178–182.

Malevergne, Y. and Sornette, D. (2003). Testing the gaussian copula hypothesis for financial assets dependences. *Quantitative Finance*, 3(4):231–250.

McGreever, L. (2015). Estimating the value at risk: A dynamic extreme value theory approach. Master's thesis, Amsterdam School of Economics, Faculty of Economics and Business, University of Amsterdam. Supervisor: Dr. S.U. Can , Second reader: Prof. R.J.A. Laeven.

McNeil, A., Frey, R., and Embrechts, P. (2005). Quantitative risk management: Concepts, techniques and tools. *Princeton University Press*.

Mikosch, T. and Stărică, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the igarch effects. *The Review of Economics and Statistics*, 86(1):378–390.

Modigliani, F. and Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *The American Economic Review*, 48/3:261–297.

Moghadam, R. and Vinals, J. (2010). Understanding financial interconnectedness. Discussion paper, INTERNATIONAL MONETARY FUND.

Moldenhauer, F. and Pitera, M. (2018). Backtesting expected shortfall: a simple recipe? *SSRN Electronic Journal*.

Naguez, N. and Prigent, J.-L. (2017). Optimal portfolio positioning within generalized johnson distributions. *Quantitative Finance*, 17:1–19.

Newey, W. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–47.

Oanea, D.-C. and Gabriela, A. (2015). Value at risk prediction: The failure of riskmetrics in preventing financial crisis. evidence from romanian capital market. *Procedia Economics and Finance*, 20:433–442.

P. Richardson, M., Boudoukh, J., and F. Whitelaw, R. (1997). The best of both worlds: A hybrid approach to calculating value at risk. *RISK*, 11.

Petronio, F., Biglova, A., and Ortobelli, S. (2014). Optimal portfolio performance with exchange traded funds. *Ekomicka revue - Central European Review of Economic issues*, 17:5–12.

Pitera, M. and Schmidt, T. (2018). Unbiased estimation of risk. *Journal of Banking & Finance*, 91(C):133–145.

Pritsker, M. (1997). Evaluating value at risk methodologies: Accuracy versus computational time. *Journal of Financial Services Research*, 12:201–242.

Rachev, S. and Mittnik, S. (2000). *Stable Paretian Models in Finance*. Financial Economics and Quantitative Analysis Series. Wiley.

Rachev, S. T., Stein, M. A. B., and Sun, W. (2009). Copula concepts in financial markets. In *Copula Concepts in Financial Markets*.

Ramaswamy, S. (2011). Market structures and systemic risks of exchange-traded funds. *BIS working papers*.

Righi, M. and Ceretta, P. (2013). Individual and flexible expected shortfall backtesting. *Journal of Risk Model Validation*, 7:3–20.

Roccioletti, S. (2016). *Backtesting Value at Risk and Expected Shortfall*. Springer Gabler, Wiesbaden.

Rüschendorf, L. (2017). Improved hoeffding–fréchet bounds and applications to var estimates. In Úbeda Flores, M., de Amo Artero, E., Durante, F., and Fernández Sánchez, J., editors, *Copulas and Dependence Models with Applications*, pages 181–202, Cham. Springer International Publishing.

S. Lambert, N., M. Pennock, D., and Shoham, Y. (2008). Eliciting properties of probability distributions. In *Proceedings of the 9th ACM conference on Electronic commerce*, pages 129–138.

S. Wang, S., R. Young, V., and H. Panjer, H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, 21:173–183.

Sewell, M. (2011). Characterization of financial time series.

Shamiri, A., Hamzah, N., and Pirmoradian, A. (2011). Tail dependence estimate in financial market risk management: Clayton-gumbel copula approach. *Sains Malaysiana*, 40.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.

Simonato, J.-G. (2011). The performance of johnson distributions for computing value at risk and expected shortfall. *The Journal of Derivatives*, 19.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.

Sollis, R. (2009). Value at risk: A critical overview. *Journal of Financial Regulation and Compliance*, 17:398–414.

Taleb, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable*. Random House Group.

Trenca, I., MUTU, S., and Dezsi, E. (2011). Advantages and limitations of var models used in managing market risk in banks. *Finance : Challenges of the Future*, 1.

Tsay, R. (2010). *Analysis of Financial Time Series*. CourseSmart. Wiley.

Wackerly, D. D., III, W. M., and Scheaffer, R. L. (2008). *Mathematical Statistics with Applications*. Duxbury Advanced Series, seventh edition.

Wong, W. (2008). Backtesting trading risk of commercial banks using expected shortfall. *Journal of Banking & Finance*, 32:1404–1415.

Wong, Z. Y., Chin, W., and Tan, S.-h. (2016). Daily value-at-risk modeling and forecast evaluation: The realized volatility approach. *The Journal of Finance and Data Science*, 2.

X. Li, D. (1999). On default correlation: A copula function approach. *Journal of Fixed Income*, 9.

Yamai, Y. and Yoshiba, T. (2001). Comparative analyses of expected shortfall and value-at-risk under market stress. *IMES Discussion Paper Series 2001-E-12*.

Yamai, Y. and Yoshiba, T. (2002a). Comparative analyses of expected shortfall and value-at-risk: Their estimation error, decomposition, and optimization. *Monetary and Economic Studies*, 20(1):87–121.

Yamai, Y. and Yoshiba, T. (2002b). On the validity of value-at-risk: Comparative analyses with expected shortfall. *Monetary and Economic Studies*, 20:57–85.