Instituto Nacional de Matemática Pura e Aplicada

Robust Multivariate Estimates of Location and Scatter Applied to the Optimum Portfolio Selection Problem

Claudio Cardoso Flores

Advisor: **Beatriz Vaz de Melo Mendes**

Rio de Janeiro
May 2016

To my beloved son, Felipe Ganem Flores.

Acknowledgements

I would like to express my sincere gratitude to my advisor, Prof. Beatriz Mendes, for her friendship, for the continuous support provided to this work, for her patience to be my advisor for the second time, and mostly for unconditionally borrow part of her immense knowledge.

Also, it was a great honor and pleasure to be a student of Prof. Jorge Zubelli. His expertise in the Mathematical Finance field is unquestionable and, in my opinion, the course under his supervision is an inspiring example of a serious high level program in Finance.

Besides my distinguished professors, I would like to thank to my family, specially my wife Marcela Flores, for her love and understanding in periods I was absent. Without her, this work would not be possible.

I also indebted to my father Leandro Flores, my grandmother Léa Flores, my father-in-law Francisco Ganem, and my mother-in-law Márcia Ganem for their continuous friendship, love and support in all periods of my life. Needless to say that I would also like to thank to my son, Felipe Flores, for being the source of motivation, inspiration, love and everything good in my life.

Finally, to God Who gives me health, strength and permeates my live with such wonderful people.

Abstract

The optimum portfolio selection is at the core of utility maximization problems and, accordingly, it has been extensively investigated during the past decades.

Nowadays, although there are various methodologies available to portfolio managers, the most widely adopted one still relies on the traditional mean-variance approach, mainly because of its mathematical tractability. However, it is well known that mean-variance optimum portfolios can be heavily distorted due to the non-robustness of the classical mean and covariance estimates (e.g. sample mean and covariance matrix of asset returns). Under this approach, optimum portfolios may be composed by counterintuitive and/or extreme asset weights, may be unstable and sensitive to new information, and may perform poorly out of the sample. Practical consequences are: excessive transaction costs due to rebalancing policies and lack of adherence with investors views.

In this work we address this issue replacing the sample classical estimates of location and scatter as inputs in the portfolio problem by their robust counterparts. We propose the use of the high breakdown point, affine equivariant MVE, MCD, S and Stahel-Donoho estimators and compare the performance and stability of respective portfolios.

By dynamically determining breakdown points for the robust estimators and by employing a semi-parametric bootstrapping procedure, in order to formally address hypotheses tests, we find that robust portfolios present higher stability than the classical non-robust one and relative performance conditioned to the level of transaction costs in a financial market, possibly rewarding in less developed economies. Also, results prove to be robust both to the change of the analyzed portfolio and to modifications in the portfolio optimization restrictions. We find that S portfolios present the best stability profile and, accordingly, the best relative performance as well.

Keywords: Robust Estimation, Portfolio Optimization, Multivariate Robust Estimators.

Resumo

A seleção de portfolios ótimos encontra-se no cerne de problemas de maximização de utilidade e, dessa forma, vem sendo estudado extensivamente nas últimas décadas.

Atualmente, apesar de existirem diversas metodologias disponíveis aos gerentes de carteiras, sem dúvida a mais utilizada relaciona-se com a abordagem clássica de média-variância, principalmente por ser de fácil implementação e entendimento sob o ponto de vista matemático. No entanto, é bem conhecido o fato de que portfolios ótimos gerados de acordo com essa metodologia carregam distorções decorrentes da não-robustez dos estimadores clássicos (média e covariância amostrais). Sob essa abordagem, é possível que sejam verificados, entre outros: pesos extremos ou contra-intuitivos em certos ativos componentes dos respectivos portfolios ótimos, sensibilidade às novas informações advindas do mercado e performance inferior fora da amostra. Custos de transação excessivos em decorrência de políticas de rebalanceamento e perda de aderência aos objetivos dos investidores são algumas das consequências práticas.

Neste trabalho, abordamos esse problema substituindo os estimadores amostrais clássicos por alternativas robustas nos respectivos problemas de otimização. Mais especificamente, propusemos o uso de estimadores invariantes por transformações afins e com altos pontos de ruptura (MVE, MCD, S e Stahel-Donoho) e comparamos métricas de performance e estabilidade vis-à-vis os resultados obtidos sob a metodologia clássica. Além disso, determinamos os pontos de ruptura de maneira dinâmica e empregamos metodologia de bootstrapping semi-paramétrico para viabilizar a execução de testes de hipóteses.

Em linhas gerais, os resultados demonstram que os portfolios robustos são mais estáveis e que o excesso de performance relativamente ao portfolio clássico é condicionado aos custos de transação existentes no mercado, podendo ser recompensador em economias menos desenvolvidas. Os resultados obtidos se mostraram robustos à mudança do portfolio eficiente considerado na análise e a modificações nas restrições do problema de otimização correspondente. Entre os estimadores utilizados, consideramos que os portfolios gerados a partir da estimação S são os que apresentaram o melhor perfil de estabilidade e performance.

Palavras-chaves: Estimação Robusta, Otimização de Portfólios, Estimadores Robustos Multivariados.

Contents

Index **vii**

1 Introduction **1**

2 Robust Statistics **4**

2.1 Introductory Concepts 4

2.1.1 Statistical Functionals 4

2.1.2 Differentiable Statistical Functionals 5

2.2 Measures of Robustness 7

2.2.1 Influence Function 7

2.2.2 Gross Error Sensitivity and Local Shift Sensitivity 10

2.2.3 Maximum Bias and Maximum Variance 11

2.2.4 Breakdown Point 11

2.3 Robust Multivariate Estimators 12

2.3.1 Introductory Concepts to the Multivariate Analysis 12

2.3.2 Minimum Covariance Determinant Estimator 14

2.3.3 Minimum Volume Ellipsoid Estimator 16

2.3.4 Tukey Biweight S-estimator 18

2.3.5 Stahel-Donoho Estimator 20

3 Modern Portfolio Theory **23**

3.1 The Mean-Variance Approach 24

3.2 The Bayesian Approach 29

3.3 Robust Estimation and Asset Allocation - Bibliographic Review 33

4 Empirical Analysis and Simulation **37**
4.1 Problem Modeling and Metrics for Comparison 37
4.2 Breakdown Point Calibration and Outlier Detection 41
4.3 Environment $\mathbf{E_1}$ - Real Dataset 44
4.3.1 $\mathbf{E_1S_1}$ - Buy-and-Hold Strategy on Observed Returns 48
4.3.2 $\mathbf{E_1S_2}$ - Rebalancing Strategy on Observed Returns 52
4.4 Environment E_2 - Simulated Dataset 56
4.4.1 $\mathbf{E_2S_1}$ - Buy-and-Hold Strategy on Simulated Data 61
4.4.2 $\mathbf{E_2S_2}$ - Rebalancing Strategy on Simulated Data 65
4.5 Sensitivity Analysis 70

5 Final Remarks **75**
5.1 Suggestions for Further Research 78

Bibliography **79**

List of Figures

4.1 Summary of the proposed modeling and metrics for comparison among portfolios. 40

4.2 Percentage of outliers identified ($RD_i^2 > \chi^2_{15,0.975}$) in the single estimation window available in the E_1S_1 strategy. Different breakdown point values, ε_n^* were assumed in the MVE estimation process. . . 43

4.3 Squared robust distances evaluated in the E_1S_1 estimation window, considering a breakdown value of 22%. The horizontal dashed line at 27.488 represents the critical value $\chi^2_{15,0.975}$. 43

4.4 Estimated Breakdown Point per each window in the E_1S_2 strategy. 44

4.5 Comparison between $E_1S_1P_1$, $E_1S_1P_2$ and $E_1S_1P_3$ efficient frontiers built upon 50 portfolios in each case. Robust portfolios were estimated based on a breakdown point of 22%. Short Selling is allowed. . 49

4.6 Comparison among $E_1S_1P_k$, $k = 1, ..., 5$ portfolio weights. Robust portfolios were estimated based on a breakdown point of 22%. Short Selling is allowed. 50

4.7 Comparison among estimates of Cumulative Excess Returns of MSR robust portfolios ($E_1S_1P_k\widehat{CER}_t$, $k = 2, ..., 5$, $t \in [t_{w+1}, T]$). 50

4.8 (left part) Comparison among estimates of Excess Maximum Drawdowns ($E_1S_1P_k\widehat{ED}_T$, $k = 2, ..., 5$) of MSR robust portfolios. (right part) Comparison among estimates of Excess Sharpe Index ($E_1S_1P_k\widehat{ES}_T$, $k = 2, ..., 5$) of MSR robust portfolios. 51

4.9 Comparison among estimates of Excess Total Norm Weights ($E_1S_2P_kE\widehat{TNW}_T$, $k = 1, ..., 5$) of MSR robust portfolios. 53

4.10 (left part) Comparison among estimates of Cumulative Excess Returns ($E_1S_2P_k\widehat{CER}_T$, $k = 2, ..., 5$) of MSR robust portfolios, considering $c = 0.001$. (right part) The same estimation, for $c = 0.1$. 54

4.11 (top) Comparison among estimates of Excess Maximum Drawdowns ($E_1S_2P_k\widehat{ED}_T$, $k = 2, ..., 5$) of MSR robust portfolios. (bottom left) Comparison among estimates of Excess Sharpe Ratios ($E_1S_2P_k\widehat{ES}_T$, $k = 2, ..., 5$) of MSR robust portfolios, considering $c = 0.001$. (bottom right) The same estimation, for $c = 0.1$. 55

4.12 Breakdown Points used in each of the 500 simulated estimation window for the strategy E_2S_1. We followed the procedure described in Section 4.2. 62
4.13 Sampling distributions for Excess Sharpe Ratio ($E_1S_1P_k\widehat{ES}_T$, $k = 2, ..., 5$) of MSR robust portfolios, considering 500 past simulated trajectories and the models specified in Tables 4.4 and 4.5. 63
4.14 Breakdown Points used in each of the 10500 simulated estimation window for the strategy E_2S_2. We followed the procedure described in Section 4.2. 65
4.15 Sampling distributions for Excess Total Norm Weights ($E_1S_2P_k\widehat{ETNW}_T$, $k = 2, ..., 5$) of robust portfolios, considering 500 past simulated trajectories and the models specified in Tables 4.4 and 4.5. 66
4.16 Sampling distributions for Excess Sharpe Ratio ($E_1S_2P_k\widehat{ES}_T$, $k = 2, ..., 5$) of MSR robust portfolios, considering 500 past simulated trajectories, $c = 0.001$ and the models specified in Tables 4.4 and 4.5. . 68
4.17 Sampling distributions for Excess Sharpe Ratio ($E_1S_2P_k\widehat{ES}_T$, $k = 2, ..., 5$) of MSR robust portfolios, considering 500 past simulated trajectories, $c = 0.1$ and the models specified in Tables 4.4 and 4.5. . . 69
4.18 Box plots for the sampling distributions of Excess Total Norm Weights ($P_k\widehat{ETNW}_T$, $k = 2, ..., 5$), considering MSR portfolios with and without short selling. 72
4.19 Box plots for the sampling distributions of Excess Total Norm Weights ($P_k\widehat{ETNW}_T$, $k = 2, ..., 5$), considering MSR and GMV portfolios with short selling. 72

List of Tables

4.1 *Selected stocks for this work and their corresponding economic sectors and industries.* . . . 45
4.2 *Descriptive statistics for the stock returns listed in Table 4.1 and the corresponding Jarque Bera test p-values. Exc. Kurt. means excess kurtosis values with respect to the Normal reference.* . . . 46
4.3 *Specification and results for the ADF test, including the best autoregressive order p according to the AIC and the test statistics with the corresponding significancy level.* . . . 47
4.4 *Best ARMA(m,n)-GARCH(p,q) specification for the stock returns listed in Table 4.1. In this table, we present the best set $\{m, n, p, q\}$ of autoregressive orders, the degrees of freedom of the standard-t distribution assumed to represent $\{z_t\}_{t\geq 0}$ behavior and the obtained p-values for the Ljung-Box (LB) autocorrelation test, considering 5 lags.* . . . 59
4.5 *Parameters estimates for the best ARMA(m,n)-GARCH(p,q) specifications presented in Table 4.4, or else, the intercepts (μ, ω) of the mean and the volatility equations, respectively, and the estimates for the autoregressive (a_1), moving average (b_1), ARCH (α_1) and GARCH (β_1, β_2) coefficients. All coefficients are statistically significant at 10% significance level.* . . . 60
4.6 *Descriptive statistics for the Cumulative Excess Returns ($E_1S_1P_k\widehat{CER}_T$, $k = 2, ..., 5$) and for the Excess Risk ($E_1S_1P_k\widehat{ED}_T$, $k = 2, ..., 5$) sampling distributions. 500 different simulated past realizations for 15 random variables were generated according to the models specified in Tables 4.4 and 4.5.* . . . 62

4.7 *Descriptive statistics for the Excess Sharpe ($E_1S_1P_k\widehat{ES}_T$, $k = 2, ..., 5$) sampling distributions. 500 different simulated past realizations for 15 random variables were generated according to the models specified in Tables 4.4 and 4.5.* 63

4.8 *Descriptive statistics for the Excess Total Norm Weights ($E_1S_2P_kE\widehat{TN}W_T$, $k = 2, ..., 5$) sampling distributions. 500 different simulated past realizations for 15 random variables were generated according to the models specified in Tables 4.4 and 4.5.* 66

4.9 *Descriptive statistics for the Excess Risk ($E_1S_2P_k\widehat{ED}_T$, $k = 2, ..., 5$) sampling distributions. 500 different simulated past realizations for 15 random variables were generated according to the models specified in Tables 4.4 and 4.5.* 68

4.10 *Descriptive statistics for the Excess After Cost Sharpe Ratio ($E_1S_2P_k\widehat{ES}_T$, $k = 2, ..., 5$) sampling distributions, considering $c = 0.1$. 500 different simulated past realizations for 15 random variables were generated according to the models specified in Tables 4.4 and 4.5.* 69

4.11 *Descriptive statistics for the Excess Total Norm Weights ($E_1S_2P_kE\widehat{TN}W_T$, $k = 2, ..., 5$) sampling distributions of MSR and GMV portfolios built with short selling (MSR-SS and GMV-SS) and MSR portfolios without short selling (MSR-noSS). 500 different simulated past realizations for 15 random variables were generated according to the models specified in Tables 4.4 and 4.5.* 73

CHAPTER 1

Introduction

It is no novelty that most classical statistical inferential methods rely explicitly or implicitly on a number of assumptions. On several occasions, gaussian distributions are considered to model observed data behavior, under exact parametric models premises, and the respective random variables are assumed to be independent and identically distributed (i.i.d). The theories of classical parametric statistics propose optimal procedures and estimators, but do not clarify the effects on their properties in the likely event that the suppositions made were violated. In fact, in most real cases, especially when financial variables are analyzed, assumptions made are at most approximations to reality and, indeed, have a negligible probability to be exactly observed. In a loose, non-technical sense, robust statistics seeks to provide solutions that emulate popular statistical methods, but which are not unduly affected by outliers or other small departures from model assumptions, while retaining good properties at them.

The optimum portfolio selection is at the core of utility maximization problems and, accordingly, it has been extensively investigated during the past decades. In general terms, it deals with the problem of seeking an optimal combination of securities that best suits a particular investor needs in an uncertain environment. Nowadays, although there are various methodologies available to portfolio managers, the most widely adopted one still relies on the mean-variance approach, pioneered by Markowitz, mainly because of its mathematical tractability. In simple terms, its main goal is to maximize expected return for a given level of risk and a given set of investment or budget constraints. However, this optimization problem is highly dependent on its inputs, the estimated measures of risk and return which, in turn, depend on the ways the investor choses to model, estimate, assess and manage uncertainty.

One of the most know criticism to the mean-variance approach regards the fact that the sample mean and the sample covariance structure of asset returns are assumed to properly represent the expected return and risk estimates, respectively, in the asset allocation problem. These are the maximum likelihood estimates of location and scatter, considering that i.i.d random variables follow the multivariate gaussian law. Under this hypothesis, they possess optimum properties. However, due to the stylized facts in finance, these characteristics are rarely observed in empirical data. More often than not, fatter tails than what would be expected in a normal distribution are present in unconditional distribution of asset returns. Also, they often exhibit strong serially conditional heteroscedasticity of the type characterized by ARCH or GARCH models.

In this sense, mean-variance optimum portfolios can be distorted due to the non-robustness of the sample mean and covariance estimates. Usually, these portfolios are composed by counterintuitive and/or extreme asset weights, are very unstable and sensitive to new information, and tend to perform poorly out of the sample. Possible consequences are: excessive transaction costs due to rebalancing policies and lack of adherence to the investors' views.

In this work we address this issue replacing the sample estimates of location and scatter, the mean-variance portfolio problem inputs, by robust alternatives. We propose the use of the high breakdown point, affine equivariant Minimum Volume Ellipsoid (MVE), Minimum Covariance Determinant (MCD), S and Stahel-Donoho multivariate estimators and compare the performance and stability of respective portfolios both using empirical data for collected securities and simulated random variables realizations. The process of portfolio weights estimation is known in the related literature as the two-step approach, in the sense that first we robustly compute the asset returns location and scatter estimates which, in a second instance, feed the optimum portfolio optimization.

As far as we know, there is not an extensive list of academic works in this theme. Most papers that we found examine one or two robust estimators in the context of portfolio optimization and there is a narrow space for controversy regarding their results.

In this sense, this work tries to distinguish from previous ones by simultaneously using a larger set of robust estimators than what can be found in the literature, providing a more complete comparative outlook of robust portfolios performance and stability. Also, breakdown points for the robust estimators are dynamically determined, in the sense that they are not set as fixed values regardless of the corresponding data sample used for estimation purposes. Allowing breakdown points to freely vary attempts to increase the estimator's level of efficiency, without sacrificing the level of robustness required by the existing contamination in each estimation sample. Finally, the traditional contamination model extensively used in past works is replaced by a resampling technic directly employed in the collected sample, in order to formally address hypotheses tests on the parameters of portfolios performance and stability. A sensitivity analysis is also provided with the purpose to cover a more varied range of investors and their respective investment policies and objectives.

After this brief introduction, the rest of this work is organized as follows: Chapter 2 presents general concepts in robustness and the multivariate estimators of location and scatter used in this work, alongside with their key properties and practical implementation issues. Chapter 3 introduces the main features of modern portfolio theory, focusing on the mean-variance framework, its problems, criticisms and alternatives. In its last section, we cite past studies that used robust statistics in asset allocation problems. Chapter 4 presents the performance and stability results of robust portfolios both in terms of empirical data and simulated one and a sensitivity analysis is provided that relaxes some of the decisions made through this work that might have influenced the results. Finally, Chapter 5 concludes the work.

CHAPTER 2

Robust Statistics

In this chapter we address the main concepts surrounding robust statistics, starting from introductory concepts and measures of robustness and gradually moving to the multivariate analysis and the respective robust estimators employed in this work. Sections 2.1 and 2.2 present key ideas on robustness that can be easily found in the related technical literature. Therefore, only in these sections, for the sake of convenience we chose to not explicitly reference every topic to its source, but the reader should find further details on [17], [14], [37], [26], [21] and [51].

2.1 Introductory Concepts

2.1.1 Statistical Functionals

Let $\{\Omega, \mathcal{A}, P\}$ be a probability space. A parametric model consists of a family $\mathcal{F}$ of distribution functions F_θ defined on the sample space Ω, where the unknown parameter θ belongs to some parameter space Θ.

In the classical theory of statistical inference, one adheres to this parametric model, while in robust statistics, the model $\mathcal{F} := \{F_\theta; \theta \in \Theta\}$ is a mathematical abstraction which is only an idealized approximation of the reality. Thus, procedures that still behave fairly well under deviations of this model must be in place.

In the one-dimensional case, let G_n be the empirical distribution associated with a particular sample $(X_1, ..., X_n)$, or else:

$$G_n(A) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{X_j \in A\}}, \qquad A \in \mathcal{A}$$

As estimators of θ we consider real-valued statistics $T_n = T_n(X_1, ..., X_n) = T_n(G_n)$, or else, a sequence of statistics $\{T_n; n \geq 1\}$, one for each possible sample size n. In the field of robustness one generally consider estimators which are statistical functionals[1], in the sense that $T_n(G_n) = T(G_n), \forall n$. This means that we assume that there exists a functional $T : \text{domain}(T) \to \mathbb{R}$, where domain$(T)$ is the set of all distributions for which T is defined.

If G is the true distribution function governing data behavior, example 1 illustrates the variance estimator as a statistical functional.

Example 1.

$$T(G) = V_G(X) = \int_{\mathbb{R}} x^2 dG(x) - (\mathbb{E}[X])^2$$
$$T(G_n) = V_{G_n}(X) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

If $(X_1, ..., X_n)$ are $i.i.d$ random variables at the true distribution G, then the Glivenko-Cantelli theorem states that $G_n(x) \xrightarrow{a.s} G(x)$. But nothing is said about the convergence of the functional $T(G_n)$, as it should tend to $T(G)$, when $n \to \infty$, with respect to some type of convergence considering an appropriate metric. Moreover, note that G does not even have to belong to the parametric model considered in the first instance. In fact, it will often deviate slightly from it. In this sense, we need to study the behavior of $T(G_n)$ in a neighborhood of G and we can perform this by an expansion of the statistical functional similar to the Taylor one. To analyze estimators behavior in a neighborhood of the assumed model is one of the main contributions of robust statistics.

2.1.2 Differentiable Statistical Functionals

If one wants to perform an expansion of the functional $T(\cdot)$ around G, analogous to the Taylor expansion, then the concept of a statistical functional derivative must be introduced. It happens that there is more than one possible definition for this derivative. In this work we consider two of them, the Gateaux derivative and the Fréchet derivative, and compare their properties from the statistical point of view.

[1]In general terms, a functional is a function from a vector space into its underlying scalar field, or a set of functions of the real numbers. In other words, it is a function that takes a vector as its input argument, and returns a scalar. Commonly the vector space is a space of functions, thus the functional takes a function for its input argument and it is sometimes considered a function of a function.

Definition 1. *Let $F, G \in \mathcal{F}$ and let $t \in [0,1]$. Then the probability distribution*

$$G_t(F) = (1-t)G + tF$$

is called the contamination of G by F in ratio t.

Definition 2. *A functional T is differentiable in the Gateaux sense in G in direction F, if there exists the limit*

$$T'_F(G) = \lim_{t \to 0^+} \frac{T(G + t(F-G)) - T(G)}{t} = \frac{d}{dt} T(G + t(F-G)) \bigg|_{t=0}$$

where $T'_F(G)$ is called the Gateaux derivative T in G in direction F.

Define the function $\varphi(t) = T((1-t)G+tF) = T(G+t(F-G))$, $0 \le t \le 1$. This function enables the evaluation of an estimator under a contaminated distribution. Then, the Gateaux derivative is simply its right derivative at the point $t = 0$, or else, $T'_F(G) = \varphi'(0^+)$. In other words, the Gateaux derivative of a statistical functional can be understood as an estimator's rate of variation as the original distribution is infinitesimally contaminated.

Continuing with the variance example provided in the last subsection, lets see the form of the Gateaux derivative of its estimator.

Example 2.

$$T(G) = V_G(X) = \mathbb{E}_G[X^2] - (\mathbb{E}_G[X])^2$$

$$\varphi(t) = T((1-t)G + tF) = \int x^2 d((1-t)G + tF) - \left[\int x d((1-t)G + tF \right]^2 =$$

$$= (1-t)\mathbb{E}_G[X^2] + t\mathbb{E}_F[X^2] - (1-t)^2(\mathbb{E}_G[X])^2 - t^2(\mathbb{E}_F[X])^2 - 2t(1-t)\mathbb{E}_G[X]\mathbb{E}_F[X] \Longrightarrow$$

$$\varphi'(t) = -\mathbb{E}_G[X^2] + \mathbb{E}_F[X^2] + 2(1-t)(\mathbb{E}_G[X])^2 - 2t(\mathbb{E}_F[X])^2 - 2(1-2t)\mathbb{E}_G[X]\mathbb{E}_F[X] \Longrightarrow$$

$$\lim_{t \to 0^+} \varphi'(t) = T'_F(G) = \mathbb{E}_F[X^2] - \mathbb{E}_G[X^2] - 2\mathbb{E}_G[X]\mathbb{E}_F[X] + 2(\mathbb{E}_G[X])^2$$

Definition 3. *A functional T is differentiable in the Fréchet sense in G if there exists a linear functional $L_G(F-G)$, such that*

$$\lim_{t \to 0} \frac{T(G + t(F-G)) - T(G)}{t} = L_G(F-G)$$

where $L_G(F-G)$ is called the Fréchet derivative of T in G in direction F.

In [17] the author demonstrates that if a functional is Fréchet differentiable then it is Gateaux differentiable as well, as the former implies higher strict conditions on functionals. In Section 2.2 we mainly adhere to the concepts of Gateaux differentiability, thus we will not explore here any further theoretical aspects behind Fréchet differentiability.

2.2 Measures of Robustness

Generally speaking, the following characteristics are desirable for any estimator:

- Consistency: A consistent sequence of estimators is a sequence that converges in probability to the real parameter. In other words, increasing the sample size increases the probability of the estimator being close enough to the unknown quantity being estimated;
- Asymptotic Unbiasedness Normality: An asymptotically unbiased normal estimator is an estimator whose distribution centered on the true parameter approaches a normal distribution with standard deviation shrinking in proportion to $1/\sqrt{n}$ as the sample size n grows;
- Efficiency: Among unbiased estimators, there often exists one with the lowest variance, called the minimum variance unbiased estimator (MVUE). In some cases, an unbiased efficient estimator also exists, which, in addition to having the lowest variance among unbiased estimators, satisfies the Cramér–Rao bound, an absolute lower bound for the variance. Later in this section we will explore this topic;
- Robustness: The next subsections present the most useful tools to measure the robustness of an estimator. They are: Influence Function; Maximum Bias and Maximum Variance; Gross Error Sensitivity and Local Shift Sensitivity; and Breakdown Point.

2.2.1 Influence Function

Definition 4. *The Influence Function (IF) of T at G is given by*

$$IF(x,T,G) = \lim_{t \to 0^+} \frac{T(G + t(\delta_x - G)) - T(G)}{t}$$

in those x where the limit exists and where δ_x is the Dirac distribution which gives mass 1 to $\{x\}$.

Based on definition 2, it is clear that the Influence Function is the Gateaux derivative of an estimator when F is replaced by δ_x. In this respect, the IF measures the rate of variation of an estimator, when the original distribution is contaminated only in one single point. It describes the effect of an infinitesimal contamination at the point x on the estimator, standardized by the mass of the contamination.

Moreover, the Influence Function provides a valuable way to describe the properties cited in the beginning of this section, more specifically, the asymptotic normality, efficiency and the Cramér–Rao bound. To perform this, let H be some distribution "near" G. Employing the first-order Von Mises expansion[2] of the functional T at G evaluated in H it is possible to reach that

$$T(H) = T(G) + \int IF(x, T, G) d(H - G)(x) + \text{remainder} \tag{2.1}$$

If the reminder is considered as negligible, then the IF dictates the extent that the estimator is impacted when evaluated at a close, but different distribution. One important application of (2.1) uses the convergence properties of an empirical distribution G_n dictated by Glivenko-Cantelli theorem, as already mentioned before, in the sense that If the observations are *i.i.d.* according to G, the true distribution, then for sufficient large n we can replace H by G_n to get that

$$T(G_n) \approx T(G) + \int IF(x, T, G) d(G_n - G)(x) = T(G) + \int IF(x, T, G) dG_n(x) \tag{2.2}$$

where we used the clear fact that $\int IF(x, T, G) dG(x) = 0$.

However, note that the last term in (2.2) may be expressed as

$$\int IF(x, T, G) dG_n(x) = \mathbb{E}_{G_n}[IF(x, T, G)] = \frac{1}{n} \sum_{i=1}^{n} IF(X_i, T, G)$$

which by the central limit theorem is asymptotically normal.

Bringing these pieces together we can see the important relationship between the Influence Function and the asymptotic normality and the asymptotic variance of an estimator:

$$\sqrt{n}(T(G_n) - T(G)) \sim N(0, V(T, G)), \quad \text{where} \quad V(T, G) = \int (IF(x, T, G))^2 dG(x)$$

Also, for a pair of estimators $\{T_n; n \geq 1\}$ and $\{S_n; n \geq 1\}$ it is relatively simple to compute de Asymptotic Relative Efficiency (ARE), as

$$ARE_{T,S} = \frac{V(S, G)}{V(T, G)} = \frac{\int (IF(x, S, G))^2 dG(x)}{\int (IF(x, T, G))^2 dG(x)}$$

Regarding the Cramér–Rao bound we already know that it expresses a lower bound for the variance of estimators. In its simplest form, it states that the variance of any unbiased estimator is at least equal to the inverse of the Fisher information, $J(\cdot)$. An unbiased estimator that achieves this lower bound is said to be fully efficient.

[2] For more details on the Von Mises calculus, please refer to [44].

Suppose a sequence of estimators $\{T_n; n \geq 1\}$ for which the corresponding functional T is consistent, in the sense that if we estimate the true parameter θ using the empirical distribution G_n by the rule $\hat{\theta}_n = T(G_n)$, then the functional applied to the true distribution would recover the true parameter, $\theta = T(G)$.

For a fixed $\theta^* \in \Theta$ and the corresponding distribution $G^* = G_{\theta^*}$, the Fisher Information is

$$J(G^*) = \int \left(\frac{\partial}{\partial \theta} (ln g_\theta(x))_{\theta^*} \right)^2 dG^*$$

where g_θ is the density of G_θ.

It can be proved, as it is for example in [14], that

$$\int IF(x, T, G^*)^2 dG^*(x) \geq \frac{1}{J(G^*)}$$

Finally, it is also possible to calculate the absolute asymptotic efficiency of an estimator, which is given by

$$e := \frac{1}{V(T, G^*) J(G^*)}$$

As we saw in this section, Influence Functions are the cornerstone of the infinitesimal approach as they shed light on the behavior of estimators after a single point contamination. Moreover, through the past decades, these useful tools became more and more popular allowing, among others: the investigation of several local robustness properties, the deeper understanding of particular estimators and the formulation of new estimators with pre-specified characteristics.

We end this subsection evaluating the Influence Function for the expected value using its corresponding functional $T(G) = \int x dG(x)$. Suppose, without loss of generality, that the real parameter is $\theta = 0$. In this case, the Influence Function takes the form:

Example 3.

$$IF(x, T, G) = \lim_{t \to 0+} \frac{\int x d[(1-t)G + t\delta_x](x) - \int x dG(x)}{t} =$$
$$= \lim_{t \to 0+} \frac{(1-t)\int x dG(x) + t \int x d\delta_x(x) - \int x dG(x)}{t} = \lim_{t \to 0+} \frac{tx}{t} = x$$

because $\int x dG(x) = 0$ by the previously adopted premise. It is not difficult to show that the arithmetic mean enjoys most properties of a good estimator, except by the fact that it is not robust to deviations of the assumed model. In fact, example 3 demonstrated that its Influence Function is not bounded at all and even a single outlier can cause serious damage to the respective estimate. The next measures of robustness deepen the understanding of this kind of problem.

2.2.2 Gross Error Sensitivity and Local Shift Sensitivity

Definition 5. *The Gross Error Sensitivity (γ^*) of a functional T under a distribution G is*

$$\gamma^* = \underset{x}{sup}|IF(x,T,G)|$$

Being the maximum absolute value of the Influence Function under G, the Gross Error Sensitivity measures the worst influence which a small amount of contamination of fixed size can have on the value of the estimator, considering all points x that this contamination may occur. In this sense, it may be understood as an upper bound on the asymptotic bias of the estimator and, as so, a finite γ^* is always desirable, in which case the estimator is called B-robust.

Definition 6. *The Local Shift Sensitivity (λ^*) of the functional T under distribution G is*

$$\lambda^* = \underset{x}{sup}\left|\frac{IF(y,T,G) - IF(x,T,G)}{y-x}\right|$$

Intuitively, the Local Shift Sensitivity measures the effect of shifting the contamination slightly from the point x to some neighboring point y, standardized by the difference between x and y.

The example below illustrates the difference between γ^* and λ^* when considering the functional $T(G) = \mathbb{E}_G[X]$.

Example 4.

$$T(G) = \mathbb{E}_G[X] \Longrightarrow IF(x,T,G) = x - \mathbb{E}_G[X] \Longrightarrow$$
$$\gamma^* = \underset{x}{sup}|IF(x,T,G)| = \underset{x}{sup}|x - \mathbb{E}_G[X]| = \infty$$
$$\lambda^* = \underset{x}{sup}\left|\frac{IF(y,T,G) - IF(x,T,G)}{y-x}\right| = \underset{x}{sup}\left|\frac{y - \mathbb{E}_G[X] - (x - \mathbb{E}_G[X])}{y-x}\right| = 1$$

which means that the arithmetic mean is not B-robust, as the worst possible contamination can potentially lead the estimator out of bounds but, on the other hand, the mean is not sensible to local changes.

2.2.3 Maximum Bias and Maximum Variance

Definition 7. *Assume that the true distribution function G lies in some family $\mathcal{F}$. The Maximum Bias ($b(\mathcal{F})$) is*

$$b(\mathcal{F}) = \sup_{F \in \mathcal{F}} |T(F) - T(G)|$$

The family $\mathcal{F}$ can take various forms. For example, it can be a Levy-neighborhood

$$\mathcal{F}_\varepsilon = \{F; \forall t, G(t-\varepsilon) - \varepsilon \leq F(t) \leq G(t+\varepsilon) + \varepsilon\}$$

or, as usually considered in the robustness analysis, it can be the ε-contaminated neighborhood of the true distribution function G, that is:

$$\mathcal{F}_\varepsilon = \{F; F = (1-\varepsilon)G + \varepsilon H\}, H \text{ arbitrary}$$

However, independently of the model $\mathcal{F}$ considered, the maximum bias is always evaluated considering all possible distributions in a contaminated environment. In a similar way, it is possible to define the Maximum Variance.

Definition 8. *Assume that the true distribution function G lies in some family $\mathcal{F}$. The Maximum Variance ($v(\mathcal{F})$) is*

$$v(\mathcal{F}) = \sup_{F \in \mathcal{F}} V(F, T)$$

2.2.4 Breakdown Point

Definition 9. *Let $X^{(0)} = (X_1, ..., X_n)$ be a random sample and consider the corresponding value $T_n(X^{(0)})$ of a functional T. Then, replace any m components by arbitrary values, possibly very unfavorable, even infinite. The new sample after the replacement denotes $X^{(m)}$, and let $T_n(X^{(m)})$ be the pertaining value of the estimator. The Breakdown Point (ε_n^*) of estimator T_n for sample $X^{(0)}$ is the number*

$$\varepsilon_n^*(T_n, X^{(0)}) = \frac{m^*(X^{(0)})}{n}$$

where $m^*(X^{(0)})$ is the smallest integer m, for which

$$\sup_{X^{(m)}} ||T_n(X^{(m)}) - T_n(X^{(0)})|| = \infty$$

In other words, the Breakdown Point is the smallest part of the observations that, being replaced with arbitrary values, can lead T_n up to infinity. For instance, the sample mean ($\bar{X}$) can be completely upset by a single outlier. If any data value $X_i \to \pm\infty$, then $\bar{X} \to \pm\infty$. This contrasts with the sample median, which is little affected by moving any single value to $\pm\infty$. In fact, the median tolerates up to 50% of "bad" contamination before it can be made arbitrarily large. So, in the context of this subsection, we say that the sample median has a Breakdown Point of 50% whereas that for the sample mean is 0%.

2.3 Robust Multivariate Estimators

Multivariate location vectors and scatter matrices estimation is a cornerstone in multidimensional data analysis, as this process generates results that are often used as inputs in subsequent inferential methods. The most common estimators of multivariate location and scatter are the sample mean and the sample covariance matrix. They are easily understood and computed by practitioners and are optimal estimators in a multivariate gaussian context. However, they are extremely sensitive to the presence of even a few outliers. Since the field of robustness became popular among statisticians and researchers, different authors have proposed various robust estimators. In this section, we present the robust alternatives chosen to be used in this work.

In this section we are considering that the multivariate dataset is composed by random vectors $\mathbf{X}_i \in \mathbb{R}^p$, $i = 1, ..., n$, where n is the sample size and p is the problem dimension.

2.3.1 Introductory Concepts to the Multivariate Analysis

In the multivariate analysis, it is important to present three more definitions.

Definition 10. *Consider the problem of estimating $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$, the multivariate location and scale parameter, by means of estimators $\mathbf{t}_n$ and $\mathbf{C}_n$, respectively. We say that $\mathbf{t}_n$ and $\mathbf{C}_n$ are affine equivariant if:*

$$\mathbf{t}_n(\mathbf{AX}_1 + \mathbf{b}, ..., \mathbf{AX}_n + \mathbf{b}) = \mathbf{At}_n(\mathbf{X}_1, ..., \mathbf{X}_n) + \mathbf{b}$$
$$\mathbf{C}_n(\mathbf{AX}_1 + \mathbf{b}, ..., \mathbf{AX}_n + \mathbf{b}) = \mathbf{AC}_n(\mathbf{X}_1, ..., \mathbf{X}_n)\mathbf{A}'$$

for all nonsingular $p \times p$ matrices $\mathbf{A}$ and $\mathbf{b} \in \mathbb{R}^p$.

The importance of affine equivariance regards the fact that in many practical occasions a linear transformation might be applicable to an estimator and, in these cases, one might want the estimators properties to be preserved. It makes the analysis independent of the variables measurement scale as well as translations or rotations of the data, for example. Not all robust estimators possess this property. According to [26], although desirable, affine equivariance is not a mandatory property, and may in some cases be sacrificed for other attributes such as computational speed. Another important result for affine equivariant estimators can be found in [10] which stated that no affine equivariant estimator can achieve a breakdown point higher than $\frac{n-p+1}{2n-p+1}$ which asymptotically tends to 50%.

Also, in the multivariate context, it is common to consider the location estimates as center of symmetry, considering a particular data cloud shape. However, while the symmetry and the center of symmetry are uniquely determined in the univariate model (e.g. the sample median), their extension to the multivariate model is not straightforward and can be made in several possible ways. Usually, robust estimators are constructed based on specific types of symmetry: spherical and elliptical.

Definition 11. *The distribution of a random vector* $\mathbf{X} = (\mathbf{X}_1, ..., \mathbf{X}_n)$ *is spherically symmetric about the location parameter* $\boldsymbol{\theta}$*, if* $\mathbf{X} - \boldsymbol{\theta}$ *is orthogonally invariant, i.e, for any orthogonal* $p \times p$ *matrix* $\mathbf{U}$*, the distribution of* $\mathbf{U}(\mathbf{X} - \boldsymbol{\theta})$ *is the same as the distribution of* $\mathbf{X} - \boldsymbol{\theta}$*.*

In these cases, as it is better explained in [21], if a density function exists, it has the form $g(||\mathbf{X} - \boldsymbol{\theta}||)$, where $|| \cdot ||$ stands for the Euclidean norm and $g(\cdot)$ is a non-negative function. Note that the class of spherically symmetric distribution includes, for instance, the multivariate normal distributions, the standard multivariate t and logistic distributions.

Definition 12. *A random vector* $\mathbf{X} = (\mathbf{X}_1, ..., \mathbf{X}_n)$ *follows an elliptically symmetric distribution with parameters* $\boldsymbol{\theta}$ *and* $\boldsymbol{\Sigma}$ *if it is affinely equivalent to that of a spherically symmetric random vector* $\mathbf{Y} = (\mathbf{Y}_1, ..., \mathbf{Y}_n)$*:*

$$\mathbf{X} \overset{d}{=} \mathbf{A}'\mathbf{Y} + \boldsymbol{\theta}$$

where $\mathbf{A}$ *is a nonsingular* $p \times p$ *matrix such that* $\mathbf{A}'\mathbf{A} = \boldsymbol{\Sigma}$*.*

For example, if $\mathbf{Y}$ is distributed according to a p-variate normal $N(\mathbf{0}, \sigma^{\mathbf{2}}\mathbf{I})$ and, therefore, spherically symmetric, then $\mathbf{X}$ is p-variate normal $N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \mathbf{A}'\mathbf{A}$. Moreover, its characteristic function and the respective density, when it exists, has the form:

$$\begin{aligned} \varphi_{\mathbf{X}}(\mathbf{t}) &= \mathbb{E}\big[e^{i\mathbf{t}'\mathbf{X}}\big] = e^{i\mathbf{t}'\boldsymbol{\theta}}\Psi(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}) \\ f_{\mathbf{X}}(x) &= |\boldsymbol{\Sigma}|^{-\frac{1}{2}} g((\mathbf{X} - \boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\theta})) \end{aligned} \tag{2.3}$$

for some scalar function $\Psi(\cdot)$ and a non-negative function $g(\cdot)$.

In this work we are dealing with financial data and, in this context, it is important to note that many authors have already showed that the Gaussian distribution is not an option for data behavior modeling. See for example, [12] and [20] regarding the dependence structure of multivariate time series. Moreover, as it is clear from Definitions 11 and 12, the spherical symmetry is a special case of the elliptical symmetry. In robustness and nonparametric studies using financial data, elliptical distributions are a commonly acceptable alternative retaining the workability of the normal distribution, as stated in [35].

The next subsections are devoted to the explanation of the robust estimators used in this work.

2.3.2 Minimum Covariance Determinant Estimator

The Minimum Covariance Determinant (MCD) estimator is one of the first affine equivariant and highly robust estimators of multivariate location and scatter proposed in the related literature. Although introduced in 1984, its popularity has increased among statisticians only after the computationally efficient FAST-MCD algorithm was proposed in [41]. Since then, the MCD has been applied in numerous fields such as medicine, finance, image analysis, and chemistry.

Definition 13. *The MCD estimators of location* $\mathbf{t}_n$ *and scatter* $\mathbf{C}_n$ *are determined by*

$$\mathbf{t}_n = \frac{1}{h}\sum_{j=1}^{h}\mathbf{X}_{i_j}$$

$$\mathbf{C}_n = c\frac{1}{h-1}\sum_{j=1}^{h}(\mathbf{X}_{i_j}-\mathbf{t}_n)(\mathbf{X}_{i_j}-\mathbf{t}_n)'$$

where the set $\mathbf{X}_h = \{\mathbf{X}_{i_1}, ..., \mathbf{X}_{i_h}\}$ *contains the* h *observations whose covariance matrix has the smallest determinant among all possible subsets of size* h *and* c *is a correction factor for consistency (see [5]).*

In [18] several properties of the MCD estimator are established as, for example:

- **Affine Equivariance**: This property follows from the fact that for each set $\mathbf{X}_h$ of size h, the determinant of the covariance matrix $\mathbf{S}$ of the transformed data $\mathbf{AX}_h$, for any nonsingular matrix $\mathbf{A}$, equals

$$|\mathbf{S}(\mathbf{AX}_h)| = |\mathbf{AS}(\mathbf{X}_h)\mathbf{A}'| = |\mathbf{A}|^2|\mathbf{S}(\mathbf{X}_h)|$$

  Hence, the optimal set of points that minimizes $|\mathbf{S}(\mathbf{AX}_h)|$ remains the same, or else, it also minimizes the determinant of the original data covariance matrix, $|\mathbf{S}(\mathbf{X}_h)|$. Similarly, the affine equivariance of the MCD location estimator follows from the equivariance of the mean.

- **Breakdown Point**: It can be proved that the MCD estimates can achieve the highest possible breakdown point value for affine equivariant estimators, when h is chosen as $\lfloor (n+p+1)/2 \rfloor$, where $\lfloor x \rfloor$ indicates the floor of x.

- **Efficiency**: The raw MCD estimator is highly robust but not so efficient. To increase efficiency while retaining robustness, a reweighting process can be applied to the raw estimators, which yields to the estimates:

$$
\begin{aligned}
\mathbf{t}_n^1 &= \frac{\sum_{i=1}^n W(d_i^2)\mathbf{X}_i}{\sum_{i=1}^n W(d_i^2)} \\
\mathbf{C}_n^1 &= c_1 \frac{1}{n} \sum_{i=1}^n W(d_i^2)(\mathbf{X}_i - \mathbf{t}_n^1)(\mathbf{X}_i - \mathbf{t}_n^1)' \quad \text{where} \\
d_i &= \sqrt{(\mathbf{X}_i - \mathbf{t}_n)\mathbf{C}_n^{-1}(\mathbf{X}_i - \mathbf{t}_n)}
\end{aligned}
$$

  is the Mahalanobis distance considering the raw MCD estimates of location and scatter, $W(\cdot)$ is an appropriate weight function and c_1 is again a consistency factor. The reweighted MCD estimates are the default choice in current implementations of most statistical softwares.

The computation of the MCD estimates of location and scatter is far from being trivial. The so-called "naive" algorithm, only feasible for small data sets, would investigate all subsets of size h to find the one with the smallest covariance matrix determinant. To circumvent this obstacle, the FAST-MCD algorithm was proposed in [41]. According to the authors, the key step of this algorithm is the fact that, starting from any approximation to the MCD, it is possible to compute another approximation with an even lower determinant, which they called as "C-Step", where the letter "C" stands for "Concentration". Theorem 1 formalizes the idea behind "C-Step" theorem and the reader can easily find its proof in [41].

Theorem 1. *Let $H_1 \subset \{1, ..., n\}$ with $\#H_1 = h$ and put*

$$
\begin{aligned}
\mathbf{t}_1 &= \frac{1}{h} \sum_{i \in H_1} \mathbf{X}_i \\
\mathbf{C}_1 &= \frac{1}{h} \sum_{i \in H_1} (\mathbf{X}_i - \mathbf{t}_1)(\mathbf{X}_i - \mathbf{t}_1)'
\end{aligned}
$$

If $|\mathbf{C}_1| \neq 0$ define de relative distances

$$
d_1(i) = \sqrt{(\mathbf{X}_i - \mathbf{t}_1)'\mathbf{C}_1^{-1}(\mathbf{X}_i - \mathbf{t}_1)}, \quad i = 1, ..., n
$$

Now take H_2 *such that*

$$\{d_1(i); i \in H_2\} := \{(d_1)_{1:n}, ..., (d_1)_{h:n}\}$$

where $(d_1)_{1:n} \leq (d_1)_{2:n} \leq \cdots \leq (d_1)_{n:n}$ *are the ordered distances, and compute* $\mathbf{t}_2$ *and* $\mathbf{C}_2$ *based on* H_2. *Then*

$$|\mathbf{C}_2| \leq |\mathbf{C}_1|$$

with equality, if and only if, $\mathbf{t}_1 = \mathbf{t}_2$ *and* $\mathbf{C}_1 = \mathbf{C}_2$.

Note that the sequence $(|\mathbf{C}_1|, |\mathbf{C}_2|, |\mathbf{C}_3|, ...)$ is nonnegative and must converge, that is, there must be a index m such that $|\mathbf{C}_m| = 0$ or $|\mathbf{C}_m| = |\mathbf{C}_{m-1}|$.

Finally, a note on how to build the initial subset H_1. According to the authors, instead of randomly drawing any H_1, it is more efficient to focus on a random $(p+1)$-subset J and compute $\mathbf{t}_0 = ave(J)$ and $\mathbf{C}_0 = cov(J)$. If $|\mathbf{C}_0| = 0$, J is extended, or else, more observations are added to J, until $|\mathbf{C}_0| > 0$. Then, for $i = 1, ..., n$ the distances $d_0^2(i) = (\mathbf{X}_i - \mathbf{t}_0)'\mathbf{C}_0^{-1}(\mathbf{X}_i - \mathbf{t}_0)$ are computed and sorted. Then, the initial H_1 subset is the one which comprises the h observations with smallest distance d_0, when one should remember that $h = \lfloor (n+p+1)/2 \rfloor$ yields the highest breakdown value.

2.3.3 Minimum Volume Ellipsoid Estimator

The Minimum Volume Ellipsoid (MVE) estimator was first proposed in [39]. Its rationale is related to the searching of the minimal volume ellipsoid containing, at least, half of the data set points. In this sense, $\mathbf{t}_n$ is taken to be the center of the minimum volume ellipsoid while $\mathbf{C}_n$ is its shape matrix.

More formally, define

$$E(a, S) := \{\mathbf{X}; (\mathbf{X} - \mathbf{a})'\mathbf{S}^{-1}(\mathbf{X} - \mathbf{a}) \leq 1\}$$

as the ellipsoid centered in $\mathbf{a}$ with scatter matrix $\mathbf{S}$ and radius 1.

Definition 14. *The Minimum Volume Ellipsoid (MVE) location estimator* $\mathbf{t}_n$ *and scatter estimator* $\mathbf{C}_n$ *are determined by*

$$\begin{aligned} &min \ \ |\mathbf{C}_n| \\ &subject\ to \ \ \#\{i; (\mathbf{X}_i \in E(\mathbf{t}_n, \mathbf{C}_n)\} \geq \left\lfloor \frac{n+1}{2} \right\rfloor \end{aligned} \tag{2.4}$$

In [48] several properties of the MVE estimator are considered, such as:

- **Affine Equivariance**: This follows from the fact that the affine transformation $\mathbf{X} \to \mathbf{AX} + \mathbf{b}$ transforms an ellipsoid with center $\mathbf{t}_n$ and scatter matrix $\mathbf{C}_n$, containing at least $\lfloor (n+1/2) \rfloor$ points of $\mathbf{X}$ into another ellipsoid with center $\mathbf{At}_n + \mathbf{b}$ and scatter matrix $\mathbf{A}'\mathbf{C_n}\mathbf{A}$, which also contains at least $\lfloor (n+1/2) \rfloor$ points of $\mathbf{AX} + \mathbf{b}$.

- **Breakdown Point**: It can be proved that the MVE estimates of location and scatter achieve the highest possible breakdown point value for the class of equivariant estimators, when the right hand side of (2.4) is replaced by $\lfloor (n+p+1)/2 \rfloor$.

- **Efficiency**: Similarly to the MCD case, a drawback of the MVE estimator also regards to its efficiency. In [6] it is proved that the MVE estimators of location and scatter converge at a rate $n^{-1/3}$ to a non-Gaussian distribution, which implies in a 0% asymptotic efficiency. Also, the finite-sample efficiency of the MVE estimates is low (see e.g. [37]). Therefore, once again the one-step reweighted MVE estimates are proposed to mitigate this efficiency problem. These are given by:

$$
\begin{aligned}
\mathbf{t}_n^1 &= \frac{\sum_{i=1}^n w_i \mathbf{X}_i}{\sum_{i=1}^n w_i} \\
\mathbf{C}_n^1 &= \frac{\sum_{i=1}^n w_i (\mathbf{X}_i - \mathbf{t}_n^1)(\mathbf{X}_i - \mathbf{t}_n^1)'}{\sum_{i=1}^n w_i} \qquad \text{where} \\
w_i &= \begin{cases} 1, & \text{if } \sqrt{(\mathbf{X}_i - \mathbf{t}_n)\mathbf{C}_n^{-1}(\mathbf{X}_i - \mathbf{t}_n)} \leq \sqrt{\mathcal{X}^2_{p,0.975}} \\ 0, & \text{otherwise} \end{cases}
\end{aligned}
$$

  These one-step reweighted MVE estimates are a weighted mean and covariance where regular observations are given weight one, but outliers, according to the initial MVE solution, are given weight zero. The one-step reweighted MVE estimators have the same breakdown value as the initial MVE estimators, as stated in [23], but a much better finite-sample efficiency, as indicated in [37] and [42]. Note that many statistical softwares, such as R, report the one-step reweighted MVE estimates by default.

Regarding the MVE estimates computation, just as like what was presented in the MCD case, the MVE "naive" algorithm would also require an exhaustive search including all combinations of ellipsoids containing h observations, to find the one with the smallest volume. One more time, this combinatorial problem is only feasible for small data sets in low dimensions and, to circumvent this problem, approximate algorithms were proposed.

The standard MVE algorithm limits its search to ellipsoids determined by subsets consisting of $p+1$ observations of $\mathbf{X}$. For each subset of size $p+1$, indexed by $J = \{i_1, ..., i_{p+1}\} \subset \{1, ..., n\}$, its sample mean and sample covariance matrix given by

$$\mathbf{t}_J = \frac{1}{p+1}\sum_{j=1}^{p+1}\mathbf{X}_{i_j}$$

$$\mathbf{C}_J = \frac{1}{p}\sum_{j=1}^{p+1}(\mathbf{X}_{i_j} - \mathbf{t}_J)(\mathbf{X}_{i_j} - \mathbf{t}_J)'$$

are calculated. If a particular $\mathbf{C}_J$ is singular, then observations from $\mathbf{X}$ are added until a subset with nonsingular sample covariance matrix is obtained. Each ellipsoid determined by $\mathbf{t}_J$ and $\mathbf{C}_J$ is then inflated or deflated by the application of a scaling factor until they contain exactly h points. Normally, the scaling factor used is D_J^2/c^2 with $c = \sqrt{\mathcal{X}^2_{p,\alpha}}$ and

$$D_J^2 = ((\mathbf{X}_{i_j} - \mathbf{t}_J)'\mathbf{C}_J^{-1}(\mathbf{X}_{i_j} - \mathbf{t}_J))_{h:n}$$

where $h:n$ indicates the h-th smallest squared distance among the squared distances of the n observations in $\mathbf{X}$.

Finally, the ellipsoid with the smallest volume is then used to obtain the MVE estimates.

2.3.4 Tukey Biweight S-estimator

S-estimators were firstly introduced in [38] and in [22] for regression problems, but there is a direct generalization to the multivariate location and dispersion estimation.

Definition 15. *Let $\rho : \mathbb{R} \to [0, \infty)$ is such that:*

1. *It is a symmetric function;*
2. *It has continuous derivative ψ;*
3. *$\rho(0) = 0$; and*
4. *There exists a finite constant $c_0 > 0$ such that ρ is strictly increasing on $[0, c_0]$ and constant on $[c_0, \infty)$.*

Then, the S estimate of multivariate location and covariance is defined as the solution $(\mathbf{t}_n, \mathbf{C}_n)$ to the problem:

$$\begin{aligned}&\textit{min} \;\; |\mathbf{C}| \\ &\textit{subject to} \;\; \frac{1}{n}\sum_{i=1}^{n}\rho\left(\sqrt{(\mathbf{X}_i - \mathbf{t})'\mathbf{C}^{-1}(\mathbf{X}_i - \mathbf{t})}\right) = b_0\end{aligned}$$

among all $(\mathbf{t}, \mathbf{C}) \in \boldsymbol{\Theta}$, $\boldsymbol{\Theta}$ being the set of all possible pairs $(\mathbf{t}, \mathbf{C})$.

The constant $0 < b_0 < sup\rho$ can be chosen in agreement with an assumed underlying distribution. If an elliptical premise is in place, then b_0 is generally chosen as $\mathbb{E}[\rho(||\mathbf{X}_0||)]$, where $||\cdot||$ stands for the Euclidean norm. In this case, the constant c_0 can be chosen such that:

$$0 < \frac{b_0}{sup\rho} = r \leq \frac{n-p}{2n}$$

which leads to a breakdown point $\varepsilon_n^* = \lceil nr \rceil / n$, where $\lceil x \rceil$ denotes the ceiling function applied to a generic x. For more details on this subject, please refer to [22].

It might be worthwhile to mention that S estimators of location and scatter can be seen as a "robustification" of the least square method. If $b_0 = p$, the problem dimension, then using $\rho(x) = x^2$ in Definition 15 yields the sample mean and the sample covariance as unique solutions. Also, note that the MVE estimator of location and scatter is a special case of S-estimators, using a discontinuous ρ function. However, in [7] the authors demonstrated that choosing a smooth ρ function greatly improves the estimator's asymptotic behavior. More specifically, asymptotical normality can be proved, which means that:

$$\sqrt{n}(\mathbf{t}_n - \boldsymbol{\mu}) \text{ and}$$
$$\sqrt{n}(\mathbf{C}_n - \boldsymbol{\Sigma})$$

both converge in distribution to a multivariate normal distribution with zero mean.

Clearly, the properties of S-estimators depend on the function ρ, which should be chosen considering the existing trade-off between the robustness aspects of an estimator and its efficiency. While there exists a variety of functions proposed by different authors, in this study we employ a popular choice, used in most papers, the ρ function defined as:

$$\rho(u, c_0) = \begin{cases} \frac{u^2}{2} - \frac{u^4}{2c_0^2} + \frac{u^6}{6c_0^4}, & \text{if } |u| \leq c_0 \\ \frac{c_0^2}{6}, & \text{if } |u| > c_0 \end{cases}$$

Its derivative, which is a redescending[3] function, is known as Tukey's biweight function:

$$\Psi(u, c_0) = u\left(1 - \left(\frac{u}{c_0}\right)^2\right)^2 \mathbb{1}_{[-c_0, c_0]}(u)$$

Regarding the properties of S-estimators when a Tukey's biweight function is considered, in [22] there are proofs of their affine equivariance, their bounded influence function and their asymptotic normality, provided some conditions are met. Also the author concluded that S-estimators are able to achieve the asymptotic variances attained by the well known class of M-estimators, first proposed in [25], but in addition they have a breakdown point that becomes considerably higher when the dimension p increases.

[3]A Ψ-function is called redescending if $\Psi(u) = 0$ for all $|u| > c_0$, for $c_0 < \infty$.

There are several proposed algorithms in the related literature with the aim at reducing computational time, such as those proposed in [26] and the famous algorithm proposed in [43] based on an improved resampling technic that reduces the number of times the objective function is evaluated and increases the speed of convergence. With this algorithm, S estimates can be computed in less time than the least median squares (LMS) for regression and the minimum volume ellipsoid (MVE) for location/scatter estimates with the same accuracy.

Another interesting feature demonstrated in [22] relates to the fact that any solution $(\mathbf{t}_n, \mathbf{C}_n)$ to the problem stated in Definition 15 must also be a solution to following system of equations:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\Psi(d_i)}{d_i}(\mathbf{X}_i - \mathbf{t}) = \mathbf{0}$$
$$\frac{1}{n}\sum_{i=1}^{n}\left[p\frac{\Psi(d_i)}{d_i}(\mathbf{X}_i - \mathbf{t})(\mathbf{X}_i - \mathbf{t})' - \Psi(d_i)d_i\mathbf{C}\right] = \mathbf{0}$$

and, accordingly, a numerical method for its solution could be implemented in order to find the optimal solution.

Finally, another interesting study regarding S-estimators can be found in [36]. The authors used a modified ρ function, which they called the translated biweight function (t-biweight) and as a starting point for the numerical algorithm they used the MVE estimates of locations and scatter.

2.3.5 Stahel-Donoho Estimator

The Stahel-Donoho (SD) estimator was obtained independently in [45] and in [9]. Its main idea regards the computation of an "outlyingness measure" for each data point and the application of proper weights to them. The larger the outlyingness measure of an observation, the lower its weight as it is considered to unlikely belongs to the "good" portion of the data. Simplistically, the SD estimators of location and scatter are projection-based estimators in the sense that an outlier or a high leverage point would separate out and away from the bulk of the data when viewed from the right perspective.

Definition 16. *Let μ and σ be shift and scale equivariant univariate location and scale statistics. Then, for any direction $a \in \mathbb{R}^p$, with $||\mathbf{a}|| = 1$ the outlyingness measure r is defined as*

$$r(\mathbf{x}_i, \mathbf{X}) = \sup_{a\in\mathbb{R}^p}\frac{|\mathbf{a}\mathbf{x}_i' - \mu(\mathbf{a}'\mathbf{X})|}{\sigma(\mathbf{a}'\mathbf{X})} \tag{2.5}$$

Definition 17. *The Stahel–Donoho (SD) estimates of location and scatter are a weighted mean and a covariance matrix where the weights imposed to $\mathbf{x}_i$ are the results of the application of a nonincreasing function to their outlyingness measure $r(\mathbf{x}_i, \mathbf{X})$. More precisely, let $W : \mathbb{R}^+ \to \mathbb{R}^+$ be a non-negative weight function. The SD estimators of location and scatter are:*

$$
\begin{aligned}
\mathbf{t}_n &= \frac{\sum_{i=1}^n W(r^2(\mathbf{x}_i, \mathbf{X}))\mathbf{x}_i}{\sum_{i=1}^n W(r^2(\mathbf{x}_i, \mathbf{X}))} \\
\mathbf{C}_n &= \frac{\sum_{i=1}^n W(r^2(\mathbf{x}_i, \mathbf{X}))(\mathbf{x}_i - \mathbf{t}_n)(\mathbf{x}_i - \mathbf{t}_n)'}{\sum_{i=1}^n W(r^2(\mathbf{x}_i, \mathbf{X}))}
\end{aligned}
$$

According to [47], a popular choice for the weight function, which will be employed in this work, is the Huber weight function defined as

$$
W(r(\mathbf{x}_i, \mathbf{X})) = \mathbb{1}_{r(\mathbf{x}_i, \mathbf{X}) \leq c} + \frac{c}{r^2(\mathbf{x}_i, \mathbf{X})} \mathbb{1}_{r(\mathbf{x}_i, \mathbf{X}) > c}
$$

for some threshold c. As one can see, the choice of c is a trade-off between robustness and efficiency. Small values of c quickly start to down weigh observations with increasing outlyingness while larger values of c only down weigh observations with extreme outlyingness value. A suggestion of $c = min(\sqrt{\mathcal{X}^2_{p,0.5}}, 4)$ is provided in [28] for estimation in high dimensions.

Also, define $\mathbf{Y} := \mathbf{AX} + \mathbf{b}$. As μ and σ in Definition 16 were chosen as equivariant univariate location and scale statistics, it can be proved, as it is for example in [52], that $r(\mathbf{y}_i, \mathbf{Y}) = r(\mathbf{x}_i, \mathbf{X})$, for any non-singular $p \times p$ matrix $\mathbf{A}$ and any vector $\mathbf{b} \in \mathbb{R}^p$. Thus, the outlyingness measure is invariant to linear transformations and, as so, the SD estimators are affine equivariant as well.

Actually, regarding μ and σ, it is shown in [46] and in [13] that the SD estimator attains the maximum possible breakdown point for an affine equivariant estimator, when those univariate location statistics are taken to be, respectively, the median (MED) and a modified version for the median absolute deviation (MAD) defined by the authors.

Also, in [27] the authors demonstrated that the SD estimators are $\sqrt{n}$-consistent, they studied their asymptotic bias and illustrated via simulation that they have a high relative efficiency.

Regarding the computation of SD estimators, the practical difficulty lies in computing $r(\mathbf{x}_i, \mathbf{X})$, defined in (2.5). In the light of this equation computing the outlyingness measures seem hopeless since projection to all directions should be considered. To overcome this problem, several approximate procedures have been proposed, such as those based on the subsampling and on the pigeonhole principles, see for example [45] and [40].

For instance, the main ideas surrounding the subsampling procedure are related to the estimates approximation based on a large number of directions. For each subsample $\mathbf{J} = \{\mathbf{X}_{i_1}, ..., \mathbf{X}_{i_p}\}$ of size p, let $\mathbf{a}_J$ be a vector of norm 1 orthogonal to the hyperplane spanned by the subsample. Then, N subsamples $J_1, ..., J_N$ are generated and (2.5) is replaced by:

$$r(\mathbf{x}_i, \mathbf{X}) = \underset{k}{max} \frac{|\mathbf{a}_{J_k} \mathbf{x}_i' - \mu(\mathbf{a}'_{J_k} \mathbf{X})|}{\sigma(\mathbf{a}'_{J_k} \mathbf{X})}$$

CHAPTER 3

Modern Portfolio Theory

In this chapter, we describe the main concepts surrounding the asset allocation problem, extensively investigated in the last decades, mainly after the seminal paper of Harry Markowitz [24]. Our exhibition here is restricted to the relevant ideas for our study and the concepts described, when not explicitly referenced to their sources, may be easily found in [19], [16] and [32], among others.

Recognizing the fact that individuals' investment decisions under uncertainty are undoubtedly influenced by many considerations, the expected utility hypothesis is a commonly accepted theory for asset choice under uncertainty that provides the underpinnings for the analysis of asset demands. Under this hypothesis, each individual's investment decision is characterized as if he could determine the probabilities of possible asset payoffs, assign an index to each possible outcome, and chose the investment policy to maximize the expected value of that index.

Suppose that an individual at time 0 is facing a problem of how to build a portfolio composed of p different risky assets and one risk free asset, to be held until time 1. If his initial wealth is W_0, the final wealth will be

$$W_1 = \left(W_0 - \sum_{i=1}^{p} x_i\right)(1 + r_f) + \sum_{i=1}^{p} x_i(1 + r_i) = W_0(1 + r_f) + \sum_{i=1}^{p} x_i(r_i - r_f)$$

where x_i, $i = 1, ..., p$ is the amount of money invested in the i-th risk asset, with random rate of return r_i and r_f is the riskless interest rate.

So, at time 1 the individual extracts utility, or satisfaction, from this final wealth, which can be mathematically described by a utility function, $U(\cdot)$, a relationship between wealth and satisfaction of consuming this wealth. Thus, the initial problem at time 0 is related to the proper choice of portfolio weights such that the expected value of the utility function at time 1 is maximized. Mathematically, the portfolio problem is:

$$\underset{x_1,\ldots,x_p}{max}\, \mathbb{E}[U(W_1)] = \mathbb{E}\left[U\left(W_0(1+r_f) + \sum_{i=1}^{p} x_i(r_i - r_f)\right)\right]$$

However, as we are dealing with economies under uncertainty, it is important to characterize an individual's behavior when facing risk.

Definition 18. *A gamble is actuarially fair when its expected payoff is zero,*

$$ph_1 + (1-p)h_2 = 0$$

for a positive return, h_1, with probability p and a negative return, h_2, with probability $(1-p)$.

Definition 19. *An agent is risk–averse if, at any wealth level W, he is unwilling to accept or is indifferent to any actuarially fair gamble, i.e.,*

$$\begin{aligned} &U(W_0) \geq pU(W_0+h_1) + (1-p)U(W_0+h_2) \implies \\ &U(p(W_0+h_1) + (1-p)(W_0+h_2)) \geq pU(W_0+h_1) + (1-p)U(W_0+h_2) \end{aligned} \tag{3.1}$$

By (3.1) one can see that if the possible outcomes in time 1 constitute a fair gamble, then risk aversion implies in a concave utility function.

As with all theoretical models, the expected utility model is not without its limitations. One of them is that it treats uncertainty as objective risk, that is, as a series of coin flips where the probabilities are objectively known. In practice, it is very difficult to assign probabilities to possible outcomes and to not be mistaken, to some degree, on the adopted models and estimation procedures.

3.1 The Mean-Variance Approach

An individual's utility function can be represented as a Taylor series expanded around his expected end of period wealth W_1.

$$U(W_1) = U(\mathbb{E}[W_1]) + U'(\mathbb{E}[W_1])(W_1 - \mathbb{E}[W_1]) + \frac{1}{2}U''(\mathbb{E}[W_1])(W_1 - \mathbb{E}[W_1])^2 + R_3 \tag{3.2}$$

where

$$R_3 = \sum_{n=3}^{\infty} \frac{1}{n!} U^{(n)}(\mathbb{E}[W_1])(W_1 - \mathbb{E}[W_1])^n$$

and $U^{(n)}$ is the n-th derivative of U.

Taking the expectation on both sides of (3.2):

$$\mathbb{E}[U(W_1)] = U(\mathbb{E}[W_1]) + \frac{1}{2}U''(\mathbb{E}[W_1])Var(W_1) + \mathbb{E}[R_3] \tag{3.3}$$

where one can see a preference for expected wealth and an aversion to the variance of wealth for an individual having an increasing and strictly concave utility function. However, it is important to emphasize that (3.3) clearly shows that the expected utility of the end of period wealth, the one that we want to maximize in our asset allocation problem, cannot be defined solely as a function of the expected value and variance of wealth for arbitrary distributions and preferences. The remainder term R_3 which involves higher order moments may not be negligible.

In order to motivate the use of the mean-variance approach, we may distinguish between two cases involving different assumptions of distributions and preferences:

- **Arbitrary Distributions:** In these cases, one should use a quadratic utility function, as the third and higher order derivatives are all equals to zero and, therefore, $\mathbb{E}[R_3] = 0$. For instance, the following utility function is usually cited in most textbooks on this subject:

  $$U(W_1) = W_1 - bW_1^2, \quad b > 0$$

  According to [16], a quadratic utility function provides a viable way to describe the asset allocation optimization problem as being dependent solely on the mean and the variance of the final wealth at period 1. However, it presents drawbacks, such as: satiation and increasing absolute risk aversion.

  The satiation characteristic is due to the parabolic shape of the utility function and implies that an increase in wealth beyond the satiation point decreases utility. In this sense, care must be taken to assure that the outcomes remain in the relevant range of the utility function.

  Regarding absolute risk aversion, the Arrow-Pratt measure $A(W)$ is the most commonly used one, which for a quadratic utility function, is computed as follows:

  $$A(W) = \frac{-U''(W)}{U'(W)} = \frac{2b}{1-2bW} \Longrightarrow A'(W) = \frac{4b^2}{(1-2bW)^2} > 0$$

Thus, it is possible to see that the quadratic utility function exhibits increasing absolute risk aversion, which implies that risky assets are considered as inferior goods. Intuition would suggest that as the level of wealth increases, investors would be more tolerant to risks and not the opposite.

- **Arbitrary Preferences:** In these cases, one should assume that the statistical behavior of rates of return on risky assets can be described by some distribution with third and higher order moments completely expressed as functions of the first two moments. Moreover, it is important that the linear combination of these rates of return, or else, the rate of return of the portfolio itself, also follow the same distribution.

  Although normal distributions fit well to the above-mentioned characteristics and are the usual choice among practitioners, according to [19] they are not the only option. Actually, any distribution function pertaining to the family of elliptical distributions, would be consistent with the mean-variance analysis. In (2.3) we saw that all elements of the class of elliptical distributions are solely defined by two parameters (location and dispersion), are symmetrical and posses the property that linear combinations of respective random variables are completely characterized by mean and variance.

Considering the second option above described, or else, that rates of return follow an elliptical distribution for arbitrary preferences, we develop the mean-variance model. Suppose that the portfolio is composed by p risky assets which are chosen among $N \geq 2$ risky assets with finite variances traded in a frictionless economy where unlimited short selling is allowed. Considering a risk-averse individual who wants to maximize expected return subjected to a given level of risk, or in its dual form to minimize risk subjected to a given level of expected return, it is possible to define the asset allocation problem as:

Definition 20. *A frontier portfolio has the minimum variance among portfolios that have the same expected rate of return. It is a frontier portfolio, if and only if, the vector of portfolio weights* $\mathbf{w}_p$ *is the solution to the quadratic program:*

$$\begin{aligned} & \underset{\mathbf{w}}{min} \frac{1}{2}\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \\ s.t \ & \mathbf{w}^T \mathbf{e} = \mathbb{E}[r_p] \\ & \mathbf{w}^T \mathbf{1} = 1 \\ & \mathbf{w} \in \mathcal{C} \end{aligned} \tag{3.4}$$

where $\boldsymbol{\Sigma}$ *is a positive definite symmetric variance covariance matrix,* $\mathbf{w}$ *is a p-vector of portfolio weights,* $\mathbf{e}$ *denotes the p-vector of expected rates of return of the risky assets,*

$\mathbb{E}[r_p]$ *is the expected rate of return of the portfolio and the set* $\mathcal{C}$ *denotes other constraints that might be included depending on the type of the problem.*

Note that the objective function in (3.4) is quadratic. If the set $\mathcal{C}$ contains only equality or inequality linear constraints, which is a very plausible assumption, then the optimization problem described is know as a quadratic programming problem.

Moreover, regardless of how investment funds are allocated among the securities considered, the covariance matrix $\mathbf{\Sigma}$ should be a positive definite one as it ultimately represents the portfolio variance, a quantity strictly positive. Considering that the objective function is twice differentiable, then one can see that

$$\nabla^2 \left(\frac{1}{2} \mathbf{w}^T \mathbf{\Sigma} \mathbf{w} \right) = \mathbf{\Sigma} > 0$$

which is the second order condition for convexity, meaning that the objective function is convex.

Hence, provided that the set of constraints is a convex set, problem (3.4) falls in the field of convex optimization, where the most important feature is the well-known fact that locally optimal points are globally optimal as well.

In this context, various algorithms are available for the computation of the mean-variance portfolio allocation, such as those based on the active set method, the interior point method and on a trust region.

Actually, if the constraints are all affine, including those contained in the set $\mathcal{C}$, then, according to [32], an analytical solution can be computed. In order to present it, lets ignore, for now, the arbitrary set $\mathcal{C}$ of constraints in (3.4). Computing the Lagrangian of the modified (without the set $\mathcal{C}$) problem, $\mathbf{w}_p$ is also the solution to the following equivalent problem:

$$\underset{\mathbf{w},\lambda,\gamma}{min} L = \frac{1}{2} \mathbf{w}^T \mathbf{\Sigma} \mathbf{w} + \lambda(\mathbb{E}[r_p] - \mathbf{w}^T \mathbf{e}) + \gamma(1 - \mathbf{w}^T \mathbf{1})$$

for λ, γ two positive constants. Then, the first order conditions are:

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{w}} &= \mathbf{\Sigma} \mathbf{w}_p - \lambda \mathbf{e} - \gamma \mathbf{1} = 0 \\ \frac{\partial L}{\partial \lambda} &= \mathbb{E}[r_p] - \mathbf{w}_p^T \mathbf{e} = 0 \\ \frac{\partial L}{\partial \gamma} &= 1 - \mathbf{w}_p^T \mathbf{1} = 0 \end{aligned}$$

which, after some manipulation, results in

$$\mathbf{w}_p = \mathbf{g} + \mathbf{h} \mathbb{E}[r_p]$$

where

$$\mathbf{g} = \frac{1}{D}(B(\mathbf{\Sigma}^{-1}\mathbf{1}) - A(\mathbf{\Sigma}^{-1}\mathbf{e}))$$
$$\mathbf{h} = \frac{1}{D}(C(\mathbf{\Sigma}^{-1}\mathbf{e}) - A(\mathbf{\Sigma}^{-1}\mathbf{1}))$$
$$A = \mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{e}$$
$$B = \mathbf{e}^T\mathbf{\Sigma}^{-1}\mathbf{e}$$
$$C = \mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{1}$$
$$D = BC - A^2$$

So, for each level of expected return on the portfolio $\mathbb{E}[r_p]$, it is possible to obtain the weights for the correspondent frontier portfolios, as in Definition 20. The set of all frontier portfolios is called the portfolio frontier.

As can be seen, $\mathbf{w}_p$ depends directly on the estimation of $\mathbf{e}$ and $\mathbf{\Sigma}$, which are in practice often estimated by their sample counterpart, i.e. by the maximum likelihood estimates under the hypothesis of a multivariate normal model. If this is the case, which rarely is, these estimates are the most efficient ones, but they are not robust, in the sense that their influence functions are unbounded (see [34]). Also, as demonstrated in [42], the effects of atypical points on the ellipsoid associated to an estimate of the covariance structure are at least two: they may inflate its volume; they may tilt its orientation. The first effect is related to inflated scale estimates. The second is the worst one, and may show up as switching the correlations signs.

The above facts should be put together with the known stylized fact that financial variables often deviate from gaussian behavior and a contaminated environment is a much more realistic scenario to be dealt with. In these cases, estimates may be severely impacted by outliers and unless the mean vector and covariance matrix are robustly estimated, the mean-variance optimizer can lead to portfolio compositions heavily influenced as well. According to [29], portfolios constructed based on high breakdown point estimates are meant to be used for long-term objectives, since they capture the dynamics of the majority of the business days. On the other hand, the efficient frontier resulting from the use of classical estimates may reflect neither the usual nor the atypical days.

Before presenting previous studies on robust estimates of location and scatter applied to the mean-variance problem, we briefly introduce an important alternative for portfolio selection, which in some extent can also be understood as a "robustification" of the mean-variance approach.

3.2 The Bayesian Approach

Suppose that the distribution of the generic random vector $\mathbf{X}$ assumed in this chapter depends on a parameter $\boldsymbol{\theta}$ taking values in a parameter space $\boldsymbol{\Theta}$. In Bayesian analysis, the estimate $\hat{\boldsymbol{\theta}}$ is not a single number as it is in classical inferential procedures. Instead it is a random variable, with a given probability density function $f_{pr}(\boldsymbol{\theta})$, called the prior distribution, which intends to reflect our knowledge (if any) of the parameter, before we gather the data itself.

The outcome of the Bayesian approach is the posterior distribution $f_{po}(\boldsymbol{\theta}|\mathbf{x})$, which updates the prior beliefs considering the relevant information in the sample, represented by the likelihood function $f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})$. In summary, the Bayesian approach is composed by:

$$\text{posteior} \propto \text{likelihood} \times \text{prior}$$

When one applies this methodology to a financial market parameter estimation, the main steps that should be followed are

- **Information from the Market:** Considering that the market variables are independent and identically distributed, the joint probability density function of the respective random variables is the product of the probability density functions of the individual variables, also known as the likelihood function:

$$f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta}) = f_{\mathbf{X}_1}(\mathbf{x}_1|\boldsymbol{\theta}) \cdots f_{\mathbf{X}_n}(\mathbf{x}_n|\boldsymbol{\theta})$$

- **Prior Knowledge:** The investor has some prior knowledge on the parameters based, for example, in his experience, and this is modeled by the prior density $f_{pr}(\boldsymbol{\theta})$.

  From the relation between the conditional and the joint probability density functions, the joint distribution of the observations and the market parameters can be expressed as:

$$f_{\mathbf{X},\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta}) = f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta}) f_{pr}(\boldsymbol{\theta})$$

- **Posterior Distribution:** The posterior probability density function is simply the density of the parameters conditional on current information. It follows from the joint density of the observations and the parameters by applying Bayes rule, which in this context reads:

$$f_{po}(\boldsymbol{\theta}|\mathbf{x}) = \frac{f_{\mathbf{X},\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} f_{\mathbf{X},\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta}) d\boldsymbol{\theta}}$$

The choices of distributions that allow analytical results to be obtained are quite limited. Parametric models for the investor's prior and the market variables that give rise to tractable posterior distributions of the market parameters are called conjugate distributions. If analytical results are not available, one has to resort to numerical simulations.

In this context, the Black and Litterman asset allocation model applies the Bayes' rule in order to consider the investor's unique insights and experience in the estimation process. This model was first introduced in [4], expanded in [3] and discussed in great details in several other papers since then, such as in [15] and in [49].

The basic assumption behind the Black-Litterman model is that without privileged or superior information one should consider the information from the market portfolio, assuming that it is at equilibrium. In other words, since market prices are based on individual investors perceptions in the whole market, then the Capital Asset Pricing Model (CAPM) equilibrium estimates, the proxy for the market view, should be adopted by the investor.

This argument reminds the semi-strong form of the market efficiency hypothesis, which states that all public information has already been absorbed into the market prices, at each point in time, and could not be explored to achieve abnormal returns. Without superior insights, the only legitimate forecast should be backed out from the market equilibrium portfolio and it would be optimal to simply use this forecast to construct the portfolio and manage it passively. Starting from this, when one has private beliefs on the assets, the forecasts should be updated based on the Bayes' Rule.

Thus, consider that the multivariate p-random vector $\mathbf{e}$ of risky assets expected returns could be modeled by a multivariate normal distribution, $\mathbf{e} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Also, consider that the market is at equilibrium and all investors hold the market portfolio $\mathbf{w}_{eq}$. Finally, consider that the utility function of the investor can be represented as a function of $\mathbf{w}'\boldsymbol{\Pi} - (\delta/2)\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$, where $\boldsymbol{\Pi}$ is the implied return by any composition $\mathbf{w}$ and δ is a risk aversion parameter, usually specified in advance by the investor according to information on historical data.

The maximization problem of this utility function is known as reverse optimization. Differently from the traditional mean-variance problem, one already knows the portfolio weights ($\mathbf{w}_{eq}$) and wants to know about $\boldsymbol{\Pi}$, the expected return implied by this composition. Practically speaking, $\mathbf{w}_{eq}$ can be obtained from the market capitalization of the securities considered and the matrix $\boldsymbol{\Sigma}$ is usually the respective sample covariance matrix.

Considering the unconstrained optimization problem, the necessary and sufficient first order condition (as in a convex programming) is

$$\frac{dU}{d\mathbf{w}} = \boldsymbol{\Pi} - \delta\boldsymbol{\Sigma}\mathbf{w} = 0 \Longrightarrow \boldsymbol{\Pi} = \delta\boldsymbol{\Sigma}\mathbf{w}_{eq}$$

Therefore, the Bayesian prior distribution $f_{pr}(\boldsymbol{\mu})$ in the Black-Litterman model is considered to be normal, centered at the implied equilibrium returns $\mathbf{\Pi}$, or else

$$\begin{aligned} &\boldsymbol{\mu} = \mathbf{\Pi} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \tau\mathbf{\Sigma}) \Longrightarrow \\ &\boldsymbol{\mu} \sim N(\mathbf{\Pi}, \tau\mathbf{\Sigma}) \Longrightarrow \\ &\mathbf{e} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma}) \end{aligned} \tag{3.5}$$

where τ is a scalar and is a simplifying assumption made by Black and Litterman that the estimate covariance structure is proportional to the covariance of the returns. Its determination is controversy and often is a subjective choice made by investors. It could be used, for example, to account for the uncertainty regarding the prior. In [49] there is some discussion on how to calibrate this parameter.

Second, the investor may insert into the problem his opinion about some securities. These views are modeled as conditional distributions and, by construction, each view is required to be uncorrelated to the other views, which results in a diagonal matrix for their covariance matrix. Also, the views must be fully invested, or else, either the sum of the weights in a view is zero (relative view) or is one (an absolute view). Note that the model does not require a view on any or all assets.

In this context, suppose the investor has opinion about k assets, $k \leq p$. Thus, consider a $k \times p$ matrix $\mathbf{P}$ containing the asset weights for each view, a k-vector $\mathbf{Q}$ for the specific returns for each view and a $k \times k$ diagonal matrix $\mathbf{\Omega}$ for the views covariance structure. As an example of how these matrices should be populated, suppose four assets and two views. First, a relative view in which the investor believes that asset 1 will outperform asset 3 by 2% with confidence ω_1. Second, an absolute view in which the investor believes that asset 2 will return 3% with confidence ω_2. Note that the investor has no view on asset 4, and thus it's return should not be directly adjusted. These views are specified as follows

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \mathbf{Q} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad \mathbf{\Omega} = \begin{bmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{bmatrix}$$

So, the investor views can be expressed as

$$\mathbf{Pr} = \mathbf{Q} + \boldsymbol{\xi} \quad \boldsymbol{\xi} \sim N(\mathbf{0}, \mathbf{\Omega})$$

where $\mathbf{r}$ is simply a random vector representing the expected returns.

As a consequence, the distribution of the views conditioned to the values of expected returned $\boldsymbol{\mu}$, or else, the likelihood function $f_{\mathbf{Pr}|\boldsymbol{\mu}}(\mathbf{Pr}|\boldsymbol{\mu})$ is

$$\mathbf{Pr}|\boldsymbol{\mu} \sim N(\mathbf{Q}, \boldsymbol{\Omega}) \tag{3.6}$$

According to [49], generally it is not possible to convert (3.6) into a useful expression due to the existing mixture of relative and absolute views. So, instead of using (3.6), the author suggests:

$$\mathbf{Pr}|\boldsymbol{\mu} \sim N(\mathbf{P}^{-1}\mathbf{Q}, (\mathbf{P}'\boldsymbol{\Omega}^{-1}\mathbf{P})^{-1})$$

Note that we have not required P to be invertible. If this proves to be a problem, in [33] there is a proposed method of matrix augmentation that makes P invertible and does not change the net results.

Regarding the views covariance matrix, $\boldsymbol{\Omega}$, it is clear that it is inversely related to the investor confidence in his views. However, the basic Black-Litterman model does not provide an intuitive way to quantify this matrix and, as so, several authors have proposed some guidance on this matter. For example, [32] suggests a particularly convenient choice for the uncertainty matrix as

$$\boldsymbol{\Omega} := \left(\frac{1}{c} - 1\right)\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}'$$

where c is a positive scalar that can be used according to the confidence in the investor's predictive skills. Note that $c \to 0$ gives rise to an infinitely disperse views distribution, which ends up having no impact at all, i.e. the investor is not trusted. In the opposite direction, when $c \to 1$, the investor is trusted completely over the prior model. The case $c = 1/2$ corresponds to the situation where the investor is trusted as much as the prior model.

At this point, we can apply Bayes theory to the problem of blending the prior (3.5) and the conditional distribution (3.6) to create the posterior distribution of the relevant parameter conditioned to the views, $f_{po}(\boldsymbol{\mu}|\mathbf{Pr})$.

$$\boldsymbol{\mu}|\mathbf{Pr} \sim N([(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\Pi}+\mathbf{P}'\boldsymbol{\Omega}^{-1}\mathbf{Q}][(\tau\boldsymbol{\Sigma})^{-1}+\mathbf{P}'\boldsymbol{\Omega}^{-1}\mathbf{P}]^{-1}, [\boldsymbol{\Sigma}+[(\tau\boldsymbol{\Sigma})^{-1}+\mathbf{P}'\boldsymbol{\Omega}^{-1}\mathbf{P}]^{-1}]) \tag{3.7}$$

and a complete derivation of (3.7) can be found in [49].

Note the intuition behind (3.7). In the one-dimensional case, the posterior updated mean vector is expressed as:

$$\frac{\frac{\pi}{\tau\sigma^2} + \frac{q}{\omega^2}}{\frac{1}{\tau\sigma^2} + \frac{1}{\omega^2}} = z_1\pi + z_2 q$$

where $z_1 = \frac{\frac{1}{\tau\sigma^2}}{\frac{1}{\tau\sigma^2}+\frac{1}{\omega^2}}$ and $z_2 = \frac{\frac{1}{\omega^2}}{\frac{1}{\tau\sigma^2}+\frac{1}{\omega^2}}$ are the confidence weights.

Thus, one can see that (3.7) simply combines the market equilibrium with the investor views through a confidence-weighted average scheme. The more confident the investor is about his view q, the more weight he should put on it and, therefore, the posterior forecast is adjusted more towards q, which reflects in the resulting portfolio weights. Otherwise, the more he should rely on the market portfolio.

Finally, in order to find the recommended portfolio weights, one must plug the optimization problem (3.4) with the assets expected returns and the respective covariance matrix, which are the parameters of the posterior distribution.

3.3 Robust Estimation and Asset Allocation - Bibliographic Review

As far as we know, there is not an extensive list of academic works applying robust estimation to the optimum portfolio selection problem. Most papers that we found investigate one or two robust estimators in this context and, as far as we could verify, there is a narrow controversy regarding their results. In this section, we provide extracts of past studies that, in our opinion, are the most significant to this work.

According to [32], the main reason why estimation risk plays such an important role in financial applications is the extreme sensitivity of the optimization problem to the unknown market distribution parameters. A slightly wrong input can give rise to a very large opportunity cost in terms of suboptimal allocation.

Actually, [8] explains that the sample mean and covariance estimates optimality under the multivariate gaussian framework partly answers why these portfolios are unstable. The efficiency of the MLE-based estimates under normality of returns is highly sensitive to deviations from this assumed (normal) model and, in particular, the respective estimators are not necessarily the most efficient ones when data behavior locates itself in a neighborhood of the assumed model.

In this work, the authors compared the minimum-variance optimum portfolios arising from the Markowitz framework with their robust counterparts, considering M-estimates and S-estimates for the mean and covariance of asset returns. The empirical data set consisted of 11 assets, where 10 assets were portfolios tracking the 10 sectors composing the $S\&P500$ index and the eleventh asset was the U.S. market portfolio represented by the $S\&P500$ index itself. The time span considered went from January 1981 to December 2002 and returns were expressed in excess of the 90-day T-bill. The different methodologies were compared using a "rolling-horizon" procedure which consisted of a fixed 10-year (120 monthly data points) estimation window, updated since a new data point was available.

The authors also assessed results under simulated data. They generated asset return data following a distribution G that slightly deviates from the normal distribution. Concretely, they assumed that G as a mixture of two different distributions, where one of them was the normal distribution and the other a t-student distribution with contamination ratios varying from 0% to 5%.

In terms of results, the authors characterized the influence functions of the M and the S portfolio weights estimators. These influence functions demonstrated that the robust weights estimators were less sensitive to normality deviations of the asset-return distribution when compared with the influence functions arising from the traditional minimum-variance policy. Moreover, numerical results confirmed that the proposed policies were indeed more stable.

According to the authors, the stability of the proposed portfolios makes them a credible alternative to the traditional portfolios since investors are generally reticent to follow policies whose recommended portfolio weights change drastically over time. Overall, the proposed M and S portfolios improved the stability properties of the traditional minimum-variance portfolios while preserving (or slightly improving) their good out-of-sample Sharpe ratios. Finally, the explanation for the good behavior of the proposed robust portfolios regards the fact that they are based on robust estimation techniques and, consequently, they are much less sensitive to deviations of the asset-return distribution from normality than the traditional portfolios.

In [29], the authors used the MCD location estimate both with the MCD and with the classical covariance estimates in order to build optimal portfolios. The empirical dataset consisted of 5 assets possessing higher than average volatility, in a global context: the Brazilian stock index IBOVESPA; a Brazilian fixed income (CDI) benchmark for money market yields; the American stock index S&P500; the MSCI EAFE index to represent global stocks traded in the rest of the world; and the J. P. Morgan Latin American EMBI to represent US dollar emerging market bonds (Brady Bonds). They used 1413 daily returns from January 1996 to May 2001, a period embedding several extreme points due to local and global crisis periods.

Several aspects of the out-of-sample performance of the robust and classical portfolios were investigated. Firstly, the authors found that robust portfolios typically yield higher accumulated returns. Also, for any given type of portfolio in the efficient frontier, the robust portfolios showed a more concentrated distribution with higher expected returns. Secondly, they concluded that the robust estimates were able to reduce the instability of the optimization process, mainly due to their definition and statistical properties. Finally, they found that this stability property carried over to the weights associated to the robust portfolios.

In [1], a quite interesting and somewhat different result was obtained. The author constructed rolling minimum variance portfolios with no short-selling constraints using a data set comprised of 5 assets from the DAX index: Adidas, Allianz, Bayer, Beiersdorf, BMW. Daily logarithmic returns from the period between March 2003 and February 2012 were collected and an estimation window length of 120 days was set. To determine the robust portfolio the author used the MCD estimator with 5% breakdown point, while to determine the classic portfolio the sample mean and the sample covariance matrix were employed.

The author characterized two important properties of robust portfolios. The first property, a desirable one, regards the fact that extreme returns have a significantly less influence on robust portfolio weights than what is verified in classic portfolios. However, this fact also revealed an undesirable property of robust portfolios: non-extreme returns have the potential to cause greater changes in weights for robust portfolios than for classic ones.

According to the author's opinion, decreasing sensitivity to outliers increases sensitivity to lesser observations (especially in a small sample). In practice, the share of outliers is minor, thus for a rolling portfolio, most observations cause greater changes in weights for robust portfolios than for classic ones, whereas only for a small number of observations (outlying ones), robust portfolios are less sensitive than classic portfolios. Hence, as the breakdown point for the given estimator increases, its sensitivity for lesser observations grows, while its sensitivity to observations which are more distant from the bulk of data decreases.

In [34], the authors studied at a theoretical level and by means of real market data and simulation, the behavior of the optimal portfolio weights estimator when computed using the sample mean and covariance as input (classical estimation). Then, robust estimators were proposed and their remarkable behavior in the presence of outlying observations were presented. More specifically, the authors analytically demonstrated that the necessary and sufficient condition for the mean-variance portfolio optimizer to be robust to local nonparametric departures from multivariate normality is that the estimators of the model's parameters be robust with bounded influence functions. Also, the authors developed a diagnostic tool for detecting the outlying data in the sample that can potentially have an abnormally large influence on the optimization process.

In this work, the Rocke's translated biweight S-estimator was employed and several types of contamination, based on a mixture model, were considered in the simulation exercise. The authors found that the classically estimated mean-variance efficient frontier model suffers from model risk when data underlying its computation are not exactly generated by a multivariate normal distribution, and that model risk dominates estimation risk. They also expect that robust portfolio would outperform the classical one if future returns were generated under the same process. This is simply because robustly estimated parameters would be closer to the true parameter values of the underlying generating process than their classical counterparts.

In a related work, the authors in [50] employed robust estimators resulting from FAST-MCD, Iterated Bivariate Winsorization, and Fast 2-D Winsorization procedures, in order to investigate the possible effects on stability of portfolios and asset turnover. They collected observed daily returns on 51 MSCI US industry sector indexes, from March 1995 to July 2005. They found these methods to be valuable tools in improving risk-adjusted portfolio performance and in reducing asset turnover. However, the results were achieved at the cost of significantly higher computation complexity.

Finally, in [31], the authors robustly estimated the multivariate distribution associated to the data collected from emerging markets, more specifically, six-dimensional contemporaneous daily log-returns from the most traded Brazilian stocks. By using pair-copulas, they obtained the inputs that define the robust efficient frontier and they analyzed the trajectory of a target return and the minimum risk portfolios during a 2-years period (out-of-sample investigation). Then, the robust portfolios were compared to their classical version based on the sample empirical estimates.

The authors found that despite the investment type, the robust methodology always outperform the classical version. Also, they determined the best rule for restoring the portfolio to its original balance and keep the allocations optimal. In their opinion, the best strategy depends on the investor risk profile, and that pair-copulas based robust minimum risk portfolios monitored by a manager which checks its composition twice a year provides the best long run investment.

Making the connection from past studies to our work, from the computational point of view, we follow the ideas proposed in [34] and in [29] known as the two-step approach, in the sense that first we compute the robust mean and covariance estimates. Second, we find the optimum portfolio by solving the classical minimum-variance problem, but replacing the sample mean and covariance matrix by their robust counterparts.

We try to innovate, though, by departing from the ε-contaminated model traditionally used in past works in simulation exercises. In this regard, we propose a resampling technic in order to formally address hypotheses tests on the performance and stability of portfolios. Also, we try to distinguish from previous works by using different robust estimators, to the greatest extent possible. A sensitivity analysis is also provided with the purpose to cover a more varied range of investors and their respective investment policies and objectives.

CHAPTER 4

Empirical Analysis and Simulation

In this chapter we analyze both the performance and stability of robust portfolios vis-a-vis the traditional asset allocation methodology based on the mean-variance framework. The following sections present the adopted notation, the problem modeling, the chosen metrics for portfolios comparison, the specifics of each practical implementation and the respective results, including a sensitivity analysis.

4.1 Problem Modeling and Metrics for Comparison

Let $S := \{S_1, S_2\}$ be the set of strategies contemplated in this work, S_1 meaning a buy-and-hold strategy and S_2 meaning a rebalancing strategy[4]. These are strategies often pursued by investors and we investigate the potential benefits generated by robust estimators in both cases. Let $E := \{E_1, E_2\}$ be the set of "environments" considered, where E_1 relates to the real dataset for selected stocks and E_2 means that the data selected in E_1 is replaced for simulated random values. Lastly, consider that $P := \{P_1, P_2, P_3, P_4, P_5\}$ is the set of optimum portfolios investigated in this work, where P_1 is the resulting portfolio from the so-called classical mean-variance framework, or else, when the sample mean and sample covariance matrix of past returns are used as measures of return and risk in (3.4), respectively. Additionally, P_k, $k = 2, ..., 5$, regard the resulting portfolios when the risk and return estimates used in P_1 are replaced by those derived from robust estimators of location and scatter, more specifically those described in Chapter 2: the MCD, MVE, S

[4]In later subsections we give the motivations behind each investment strategy that make them popular among investors.

and SD estimators, respectively.

Let the set $R := E \times S \times P = \{(E_i, S_j, P_k) | E_i \in E, S_j \in S, P_k \in P\}$, $i = 1, 2$, $j = 1, 2$, $k = 1, ..., 5$ be the Cartesian product between the sets E, S and P. This is the set that contains the final results segregated by type of strategy, type of environment and type of estimator used in (3.4).

In the portfolio selection problem we allow for short selling, which means that negative asset weights are possible. We consider that the investment policy pursued by investors dictates a preference to keep the maximum Sharpe ratio portfolio[5] (MSR) at each estimated efficient frontier. Note that in Section 4.5 both premises (short selling and MSR portfolios) are relaxed to enlarge the set of investors' profiles covered by this work.

The risk and return estimates, inputs of the optimization problem, are based on a 7-year window of past returns (estimation window), either real or simulated, depending on the environment considered. To compare the performance and stability of robust and non-robust (classical) portfolios we use an 8-year window, denoted in this work as the backtesting window. According to the notation used in this section, we say that the whole multivariate time series (15 years) goes from t_0 to T and the optimum portfolios are selected in t_w, $t_0 < t_w < T$, in such a way that the interval $[t_0, t_w]$ is a 7-year time horizon and $[t_w, T]$ is an 8-year time horizon.

Finally, the robust estimates of location and scatter are computed based on dynamically determined breakdown points, calibrated according to the theory and methods described in Section 4.2.

In this chapter, results are compiled through a set of selected performance and stability measures, presented in terms of the excess achieved by robust portfolios with respect to the classical one. More specifically, we consider cumulative excess return, excess risk, excess Sharpe ratio and excess transaction costs defined through (4.1) to (4.6). The general idea embedded in this chapter is to use the real stock returns (environment E_1) to generate estimates of these metrics and hypotheses about the respective unknown parameters. Then, in the environment E_2 we simulate random variables resembling the real data collected, using a semi-parametric bootstrapping technique, in order to get the desired understanding on the sampling distributions of the estimates evaluated in E_1. With this in hand, hypothesis tests are performed.

In this context, let $\{E_i S_j P_k R_t\}_{t \in [t_w, T]}$ be the sequence of the P_k optimum portfolio

[5]The Sharpe Ratio is a widely used tool to examine the performance of an investment by adjusting for its risk. The ex-post ratio measures the excess return (or risk premium) per unit of deviation, or else

$$S_T = \frac{\mathbb{E}[R_T] - r_f}{SD_T}$$

where S_T, R_T, SD_T and r_f are, respectively, the Sharpe ratio, the cumulative return of the investment at time T, its standard deviation and the risk-free interest rate.

returns over the backtesting window, considering the j-th strategy and the i-th environment. In the buy-and-hold strategy S_1, portfolios are not rebalanced. So, the excess returns (ER_t)[6] of robust portfolios with respect to the classical one and the respective cumulative excess returns (CER_t), both at an specific time t, are defined as:

$$\begin{aligned} &E_iS_1P_kER_t = E_iS_1P_kR_t - E_iS_1P_1R_t \\ &E_iS_1P_kCER_t = \prod_{l=t_w}^{t}(1 + E_iS_1P_kER_l) - 1 \\ &i = 1,2; \quad k = 2,...,5; \quad t = t_w + 1,...,T \end{aligned} \tag{4.1}$$

However, in the rebalancing strategy S_2 one should properly account for transaction costs when evaluating cumulative returns, as portfolios' compositions are frequently changed over the backtesting window. In this sense, consider that $T^* := \{t_w, t_w + \Delta t, t_w + 2\Delta t, ..., T - \Delta t\}$ is the set that contains the dates when portfolios are evaluated, according to a frequency Δt. Let $\{E_iS_2P_kW_t\}_{t \in T^*}$ be the sequence of the P_k optimum portfolio weights considering the i-th environment and the S_2 strategy, evaluated at each $t \in T^*$. Define:

$$\begin{aligned} &E_iS_2P_kNW_t = ||E_iS_2P_kW_t - E_iS_2P_kW_{t-\Delta t}||_1 \\ &i = 1,2; \quad k = 1,...,5; \quad t \in T^{**} := \{t_w + \Delta t, t_w + 2\Delta t, ..., T\} \end{aligned}$$

as the 1-norm of the difference between P_k portfolio weights after and before a rebalancing day.

We assume that at each time $t \in T^{**}$, the value of $E_iS_2P_kNW_t$ impacts the return achieved by the P_k portfolio through a fixed factor c, $c > 0$. That is, at each time $t \in T^{**}$, securities are bought and/or sold and these operations consume financial resources. We are accepting the premise that this consumption, at any currency unit, converts to financial return through the constant c and, therefore, that transaction costs are proportional to the change in the portfolio weights[7] and not to the number of buying or selling transactions. If this is true, then, by construction, it is clear that the larger the value of $E_iS_2P_kNW_t$, the higher the related transaction cost.

So, in strategy S_2 we redefine ER and CER as:

[6] It might be confusing that the capital letter E is used in different contexts in this work, but please note that there are remarkable differences in its usage. When we present expected values, we use the $\mathbb{E}[\cdot]$ notation instead of the simple E. When we reference the types of environment studied in this work, the capital letter E is always associated with the subscript i, or their assumed values, $1, 2$ (e.g. E_1, E_2). Finally, when the capital letter E appears followed by another capital letter (e.g. ER, ED) then it means the excess of some performance or stability metric.

[7] In later subsections we present the reasons that led to this assumption.

$$
\begin{aligned}
&E_i S_2 P_k ER_t = (E_i S_2 P_k R_t - cE_i S_2 P_k NW_t) - (E_i S_2 P_1 R_t - cE_i S_2 P_1 NW_t) \\
&E_i S_2 P_k CER_t = \prod_{t \in T^{**}} (1 + E_i S_2 P_k ER_t) - 1 \\
&i = 1, 2; \quad k = 2, ..., 5; \quad t \in T^{**}
\end{aligned}
\tag{4.2}
$$

We also define

$$
\begin{aligned}
&E_i S_2 P_k TNW_T = \sum_{t \in T^{**}} E_i S_2 P_k NW_t \\
&i = 1, 2; \quad k = 1, ..., 5
\end{aligned}
\tag{4.3}
$$

as the total norm weight (TNW) over the whole backtesting window, which can be understood as a proxy for the total transaction cost resulting from all rebalancing activities, clearly a function of the P_k portfolio stability. In the same line with the metrics above described, we define the excess total norm weights (ETNW) of robust portfolios with respect to the classical one, as:

$$
\begin{aligned}
&E_i S_2 P_k ETNW_T = E_i S_2 P_k TNW_T - E_i S_2 P_1 TNW_T \\
&i = 1, 2; \quad k = 2, ..., 5
\end{aligned}
\tag{4.4}
$$

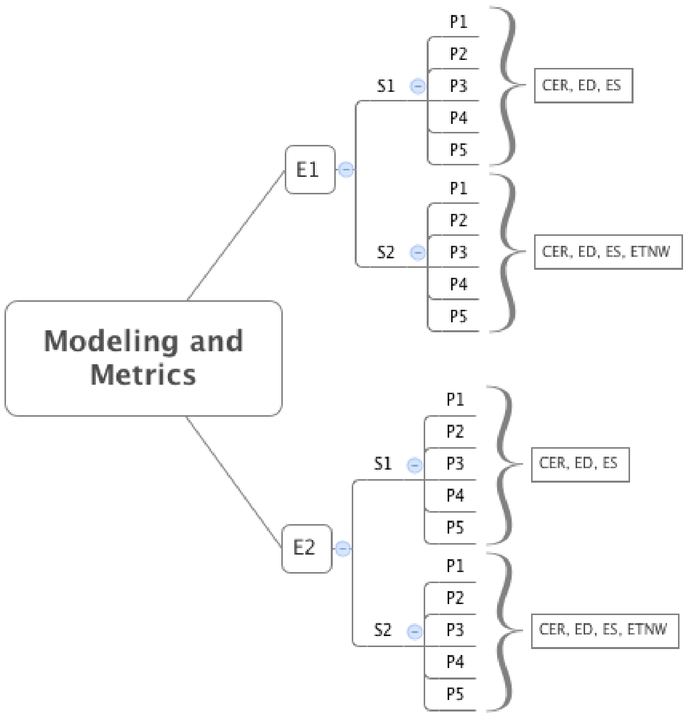

Figure 4.1: Summary of the proposed modeling and metrics for comparison among portfolios.

Regarding risk comparison we use the maximum drawdown measure which, roughly speaking identifies the highest decline in a portfolio sequence of returns from a historical peak. Let $E_iS_jP_kD_T$ be the maximum drawdown of the P_k optimum portfolio over the whole backtesting window $[t_w, T]$ considering the j-th strategy and the i-th environment. The excess risk or excess maximum drawdown (ED_T) at time T is defined as:

$$\begin{aligned} & E_iS_jP_kED_T = E_iS_jP_kD_T - E_iS_jP_1D_T \\ & i = 1,2; \quad j = 1,2; \quad k = 2, ..., 5 \end{aligned} \tag{4.5}$$

Finally, let $E_iS_jP_kS_T$ be the Sharpe Ratio of the P_k optimum portfolio considering its performance on the backtesting window $[t_w, T]$[8]. Define the excess Sharpe ratio of robust portfolios with respect to the classical one as:

$$\begin{aligned} & E_iS_jP_kES_T = E_iS_jP_kS_T - E_iS_jP_1S_T \\ & i = 1,2; \quad j = 1,2; \quad k = 2, ..., 5 \end{aligned} \tag{4.6}$$

Figure 4.1 summarizes the proposed modeling and metrics for this work. The following sections give more details of each environment and strategy implementation.

4.2 Breakdown Point Calibration and Outlier Detection

As explained in Chapter 2, the breakdown point plays a vital role in the well-known trade-off between robustness and efficiency levels of an estimator. Its value may privilege one of these two properties, as unnecessary high breakdown points generate estimators that are more robust than what is required by data contamination and, this fact ends up deteriorating the related efficiency.

The exercise presented in this section employs the concepts and methodologies described in [37], in order to identify the proportion of existing outliers in a particular sample. Knowing the percentage of "bad" data, the breakdown point could be easily set at a value that the correspondent robustness would be consistent to what is required.

However, outliers are much harder to identify in multivariate data clouds than in the univariate case. This happens mainly because one can not rely on visual inspection anymore and, in general, it is not sufficient to look at each variable separately or even

[8]The Sharpe Ratio equation needs a proper value for the risk-free interest rate. Aside from the fact that there is some controversy on which rate should be considered as the risk-free, note that this rate is not an essential parameter in this study. Actually, any choice would equally impact all portfolios measures of performance and results would not be relatively changed depending on the chosen value for r_f. In this sense, we choose to simplify in this work by setting $r_f = 0$.

at all plots of pairs of variables. Indeed, the author in [37] provides examples when clear outliers in the multidimensional context are not identified when single or pair of variables are analyzed. In most cases, such outliers do stick out in certain projections, but this does not make them easy to find. Most projections do not reveal anything.

The classical approach to outliers detection in the multivariate context considers the Mahalanobis distance (MD), introduced in Chapter 2, using the arithmetic mean and the sample covariance matrix as estimates of location and scatter. This distance should inform how far a particular point is from the center of the data cloud, taking into account its shape as well. However, it is well documented in the literature that this procedure suffers from the masking effect, by which multiple outliers do not necessarily present large values of MD. This is due to the fact that the used estimates of location and scatter are not robust and, therefore, a small cluster of outliers could potentially attract the location estimate, while inflating the scatter one.

To overcome this weakness, the author in [37] suggests the usage of robust distances (RD) instead of MD. RD are evaluated for each data point in the same way as the MD, with the exception that the arithmetic mean and the sample covariance matrix are replaced by robust estimators of location and scatter. Then, each distance is compared to a chi-squared distribution critical value, $\chi^2_{p,\alpha}$, considering a specific level of significance α and p being the problem dimension. When $RD_i^2 > \chi^2_{p,\alpha}$, the correspondent p-dimensional point $\mathbf{x}_i$ is considered an outlier.

Following the suggestion given in [37], we apply this methodology to our real dataset, exposed in Section 4.3, using the MVE estimates of location and scatter. Figure 4.2 presents the percentage of outliers identified ($RD_i^2 > \chi^2_{15,0.975}$) in the single estimation window available in the E_1S_1 strategy. Different breakdown point values, ε_n^*, in the MVE estimation process were considered in the interval $[0.01, 0.5]$, which includes the maximum breakdown possible to be attained by the class of affine equivariant estimators.

The results plotted in Figure 4.2 clearly demonstrate that there is, indeed, some level of contamination in the E_1S_1 estimation window. Also, it is possible to see that an assumed breakdown point smaller than 20% or higher than 25% is pointless. In the first case, the estimator should be more robust than it is, but in the second case, there are not so many identified outliers justifying such high values of breakdown without impairing the estimator's efficiency. In this sense, 22% seems to be a good choice for the breakdown point, which precisely identifies 22.46% of outliers in this estimation window. Figure 4.3 presents the evaluated squared robust distances for points pertaining to this window, comparatively to the threshold $\chi^2_{15,0.975} = 27.488$.

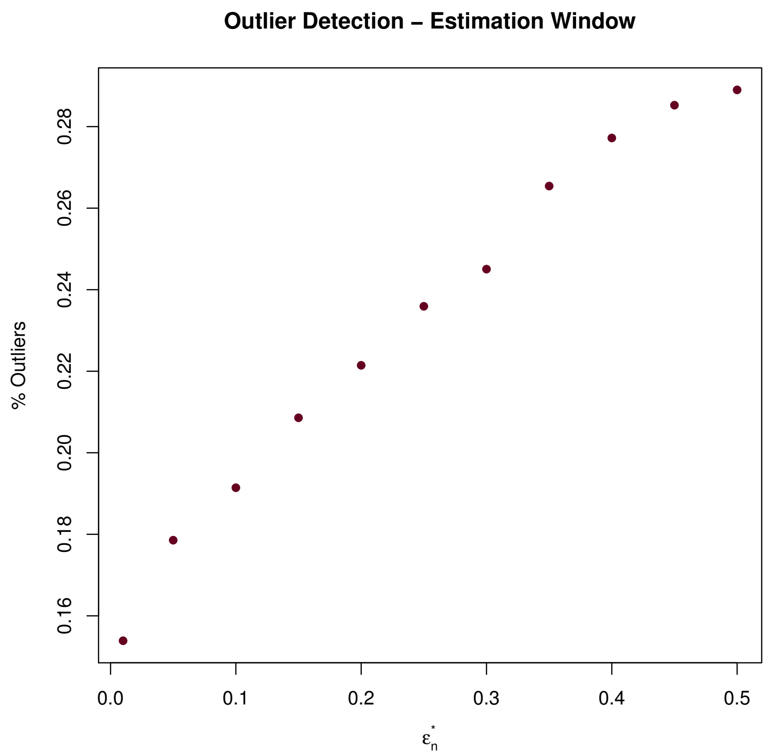

Figure 4.2: Percentage of outliers identified ($RD_i^2 > \chi^2_{15,0.975}$) in the single estimation window available in the E_1S_1 strategy. Different breakdown point values, ε_n^* were assumed in the MVE estimation process.

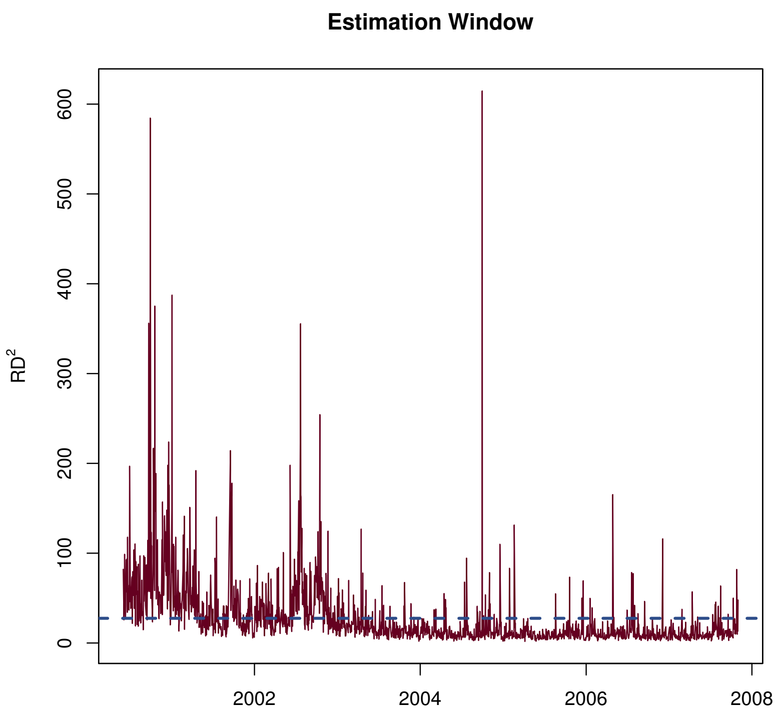

Figure 4.3: Squared robust distances evaluated in the E_1S_1 estimation window, considering a breakdown value of 22%. The horizontal dashed line at 27.488 represents the critical value $\chi^2_{15,0.975}$.

Finally, as it will be better explained in Section 4.3.2, the rebalancing strategy S_2 uses a rolling estimation window. Data set used to estimate portfolios changes frequently, which implies that, although the 22% breakdown value proved to be adequate for the first estimation window, it might be erroneous for the subsequent windows. Thus, following the same rationale described in this section, Figure 4.4 presents the most proper breakdown values for each estimation window in the E_1S_2 strategy. Aside from two estimation windows, where 21% and 23% were more adequate values, one can see that, indeed, 22% is a level of contamination that represents the majority of the data.

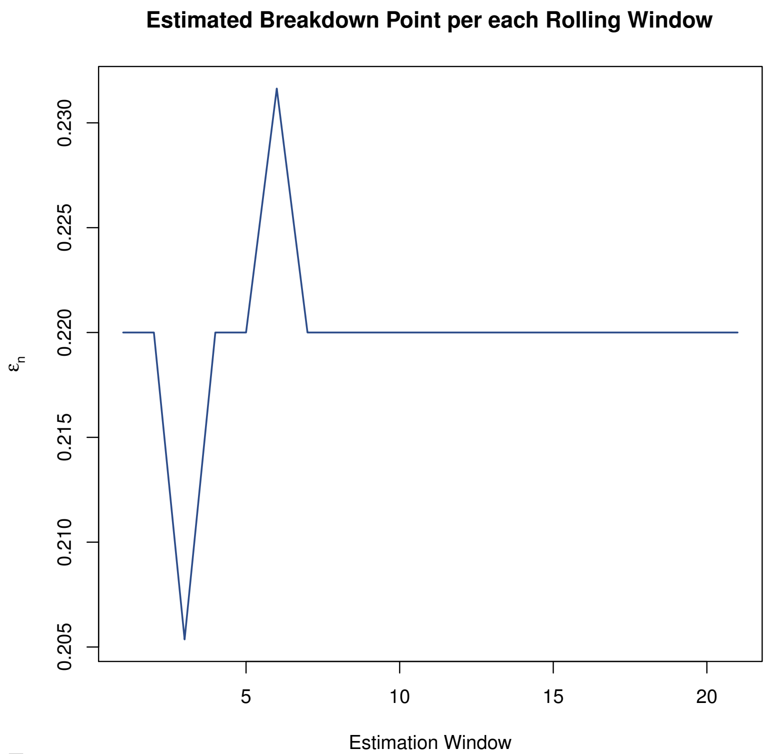

Figure 4.4: Estimated Breakdown Point per each window in the E_1S_2 strategy.

4.3 Environment $\mathbf{E_1}$ - Real Dataset

In environment E_1 we have at our disposal observed returns for collected stocks, or else, one single realization of the subjacent stochastic process. In this sense, the only procedure possible to be done here is to estimate statistics derived from the metrics defined in Section 4.1. That is, we compute the estimates $E_1S_jP_k\widehat{CER}_T$, $E_1S_2P_k\widehat{ETNW}_T$, $E_1S_jP_k\widehat{ED}_T$ and $E_1S_jP_k\widehat{ES}_T$, conditioning on our sample, the specific sequence of stock returns $\{E_1S_jP_kR_t\}_{t\in[t_0,t_w]}$ observed in the market.

To accomplish this, we selected 15 stocks negotiated either at New York Stock Exchange (NYSE) or NASDAQ Stock Market as exhibited in Table 4.1. The rationale behind this stock screening guarantees that the securities collected enjoy desirable features such as: high liquidity, long track record and differentiation, to the greatest extent possible, regarding the respective industries and sectors.

The chosen time horizon goes from 2000/06/02 to 2015/07/01 (estimation plus back-testing windows) that contains several market turbulences and, at least, two important financial crises such as those known as the ".com" and the "subprime" crises. Stock prices were collected on a daily basis.

Table 4.2 presents descriptive statistics for the returns of the stocks listed in Table 4.1. The most important point behind these results regards the fact that there is clear evidence suggesting that these stock returns do not follow a Gaussian distribution. Actually, one can see that for most selected stock returns, the values obtained for excess kurtosis with respect to the Normal reference indicate the existence of fat tails in these distributions.

Table 4.1: Selected stocks for this work and their corresponding economic sectors and industries.

Stock	**Sector**	**Industry**
Alcoa	Basic Materials	Aluminum
Apple	Consumer Goods	Electronic Equipment
AT&T	Technology	Telecom Services
Citigroup	Financial	Money Center Banks
The Coca-Cola Company	Consumer Goods	Beverages - Soft Drinks
The Dow Chemical Company	Basic Materials	Chemicals - Major Diversified
Ford Motor	Consumer Goods	Auto Manufacturers - Major
General Electric Company	Industrial Goods	Diversified Machinery
Intel Corporation	Technology	Semiconductor - Broad Line
J. C. Penney Company	Services	Department Stores
JPMorgan Chase	Financial	Money Center Banks
Merck	Healthcare	Drug Manufacturers - Major
Microsoft Corporation	Technology	Business Software & Services
Pfizer	Healthcare	Drug Manufacturers - Major
The Procter & Gamble Company	Consumer Goods	Personal Products

Source:*Bloomberg*

In order to complement this first impression we performed the well known Jarque-Bera test,[9] whose results are also presented in Table 4.2. It can be seen that in all cases it is

[9] In general terms, it is a goodness-of-fit test of whether the sample data have the skewness and kurtosis

possible to reject the Jarque-Bera null hypothesis that the sample in analysis comes from a Gaussian distribution. Although this is a well-known fact concerning stock returns, this result is particularly important for this work as it undermines the use of the sample mean of returns and the sample covariance matrix as estimates of return and risk, respectively, in the portfolio optimization problem. We have already seen in Chapter 2 that these estimates are optimal when returns follow a multivariate Gaussian distribution, which appears not to be case, but can be heavily distorted in contaminated environments or under other multivariate distributions.

Table 4.2: Descriptive statistics for the stock returns listed in Table 4.1 and the corresponding Jarque Bera test p-values. Exc. Kurt. means excess kurtosis values with respect to the Normal reference.

Stock	Mean	Min.	Max.	Std. Dev.	Skewness	Exc. Kurt.	JB p-val
Alcoa	0.0005	−0.1101	0.1300	0.0223	0.2142	2.3204	≈ 0
Apple	0.0017	−0.5187	0.1316	0.0316	−2.2817	40.6138	≈ 0
AT&T	0.0003	−0.1266	0.0911	0.0182	−0.1192	3.8077	≈ 0
Citigroup	0.0002	−0.1573	0.1264	0.0183	−0.0907	7.4569	≈ 0
Coca-Cola	0.0002	−0.1006	0.0781	0.0134	−0.2469	5.8545	≈ 0
Dow Chemical	0.0005	−0.1058	0.1139	0.0195	0.1376	3.8800	≈ 0
Ford Motor	−0.0002	−0.1469	0.1562	0.0245	0.3615	4.2651	≈ 0
G. E.	0.0001	−0.1067	0.1246	0.0175	0.2719	5.7646	≈ 0
Intel	$< 10^{-4}$	−0.2203	0.2012	0.0283	−0.2271	6.9782	≈ 0
J. C. Penney	0.0010	−0.1217	0.1761	0.0262	0.6067	4.6265	≈ 0
JPMorgan	0.0003	−0.1811	0.1604	0.0213	0.4305	8.8133	≈ 0
Merck	0.0002	−0.2678	0.1303	0.0178	−1.8303	31.5940	≈ 0
Microsoft	0.0004	−0.1182	0.1957	0.0203	0.5195	9.0629	≈ 0
Pfizer	−0.0001	−0.1115	0.0771	0.0171	−0.3569	4.5016	≈ 0
P & G	0.0005	−0.0798	0.0840	0.0125	0.0179	5.8827	≈ 0

A final aspect that must be checked before the core exercises of this chapter are presented, regards whether stock returns are stationary or not. In general terms, in time

matching a normal distribution. The test statistic JB is defined as

$$JB = \frac{n}{6}\left(S^2 + \frac{1}{4}(C-3)^2\right)$$

where n is the number of observations, S is the sample skewness and C is the sample kurtosis. If the data come from a normal distribution, the JB statistic asymptotically follows a chi-squared distribution, and it can be used to test the hypothesis that the data comes from a normal distribution.

series analysis, sample estimates make sense only if the quantities used in this process are market invariants, i.e. if they display the same statistical behavior independently across different periods, allowing the effective learning from the past.

Thus, after collecting the securities raw prices, one should extract the invariance intrinsic to this dataset, which for stocks are often embedded in their returns. More specifically, we are looking for time homogenous invariants, which are defined as those quantities whose distributions do not depend on the reference time.

Table 4.3: Specification and results for the ADF test, including the best autoregressive order p according to the AIC and the test statistics with the corresponding significancy level.

Stock	$\{\alpha, \delta\}$	p	Test Stat.
Alcoa	$\alpha = \beta = 0$	0	$-42.7827^{(**)}$
Apple	$\alpha \neq 0, \beta = 0$	0	$-44.7773^{(**)}$
AT&T	$\alpha = \beta = 0$	0	$-42.7223^{(**)}$
Citigroup	$\alpha = \beta = 0$	0	$-43.8458^{(**)}$
The Coca-Cola Company	$\alpha = \beta = 0$	0	$-42.1991^{(**)}$
The Dow Chemical Company	$\alpha = \beta = 0$	0	$-44.2906^{(**)}$
Ford Motor	$\alpha = \beta = 0$	0	$-47.1242^{(**)}$
General Electric Company	$\alpha = \beta = 0$	0	$-44.3227^{(**)}$
Intel Corporation	$\alpha = \beta = 0$	1	$-32.3664^{(**)}$
J. C. Penney Company	$\alpha = \beta = 0$	1	$-33.1793^{(**)}$
JPMorgan Chase	$\alpha = \beta = 0$	0	$-44.6576^{(**)}$
Merck	$\alpha = \beta = 0$	0	$-42.1349^{(**)}$
Microsoft Corporation	$\alpha = \beta = 0$	0	$-43.2860^{(**)}$
Pfizer	$\alpha = \beta = 0$	1	$-34.5363^{(**)}$
The Procter & Gamble Company	$\alpha \neq 0, \beta = 0$	0	$-45.7420^{(**)}$

(**) Significant at 1% level

We perform stationarity analysis on the stock returns using the *Augmented Dickey Fuller* test (ADF)[10], and the respective results are presented in Table 4.3. It is clear from the test statistics exhibited that, in all cases, it is possible to reject the null hypothesis

[10] In order to briefly describe this test, let $\{Y_t\}$ be an observed time series. The functional form of the test can assume three different formats, but the complete version is

$$\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum_{j=1}^{p} (\delta_j \Delta Y_{t-j}) + \epsilon_t$$

of existence of a unit root, at significance levels lower than 1%. In this way, hereafter we consider the selected stock returns as homogeneous market invariants.

4.3.1 E_1S_1 - Buy-and-Hold Strategy on Observed Returns

The buy-and-hold strategy S_1 combined to the environment E_1 (E_1S_1) in this subsection means that we select optimum portfolios based on the sequence of observed stock returns between t_0 and t_w and that they remain unchanged until T. Consequently, there are no portfolios stability analyses related to rebalancing strategies in this subsection.

One could point out that this strategy is too naive to be implemented, as portfolios would be exposed too much time to different types of market turbulences without rebalancing. However, there are several arguments both in the financial theory and in the related practice supporting this strategy. One of them relates to the efficient-market hypothesis which, in general terms, states that if every security is fairly valued at all times, then there is really no point to trade at all. In the extreme, supporters of this strategy may advocate that a security, once bought, should never be sold, unless the investor really needs the money. This viewpoint holds that market timing, i.e. the concept that one can enter the market on the lows and sell on the highs, does not work and attempting to do this would give negative results, at least for small or unsophisticated investors. It would be better for them to simply buy and hold.

Another argument favorable to the buy-and-hold strategy is related to cost-based grounds. Costs such as brokerage fees and bid/ask spreads are incurred in all transactions, and the buy-and-hold strategy implies in the fewest number of transactions for a given amount invested in the market, all other things being equal. Taxation laws also have some effect. Tax for long-term capital gains may be lower in most jurisdictions, and tax may be due only when the asset is sold (or never if the person dies).

Having said that, consider that $t_0 = 2000/06/02$, $t_w = 2007/11/01$ and $T = 2015/07/01$, which are roughly speaking, 7 years of past information of observed returns, $[t_0, t_w]$, and 8 year for performance comparison, $[t_w, T]$.

Considering the collected data set, Figures 4.5 and 4.6 illustrate some differences between robust and classical portfolio estimation.

Using a breakdown point of 22%, contextualized in Section 4.2, Figure 4.5 compares

where α is a constant called drift, β is the coefficient on a time trend, γ is the coefficient presenting process root, p is the lag order of the first-difference autoregressive process and ϵ_t is an independent identically distributed error term. The focus of the test is whether the coefficient γ equals to zero, meaning that the original time series process has a unit root.

Several authors have suggested different ways of specifying the deterministic structure, $\{\alpha, \delta\}$ and the lags p for the autoregressive component. In this work, we determine the first by joint significance analysis of estimated coefficients using F-type tests and the second based on the Akaike Information Criteria (AIC).

the estimated efficient frontiers for the portfolios $E_1S_1P_k$, $k = 1, ...3$. One should notice that there is a remarkable difference between classical and robust efficient frontiers, as the last one surrounds the former one. This fact suggests that it would be possible to allocate assets in a way that optimum robust portfolios would enjoy less risk for the same level of expected return or, conversely, attain higher levels of return for a given level of risk, comparatively to the classical portfolio. Unfortunately, for market practitioners, this evidence would be extremely important only in the unrealistic occasion that future returns resemble past returns. This would be the case where robust portfolios would guarantee better performance and risk profiles. Also, please note that although Figure 4.5 compares P_1 solely with P_2 and P_3, the same pattern can be observed with P_4 and P_5.

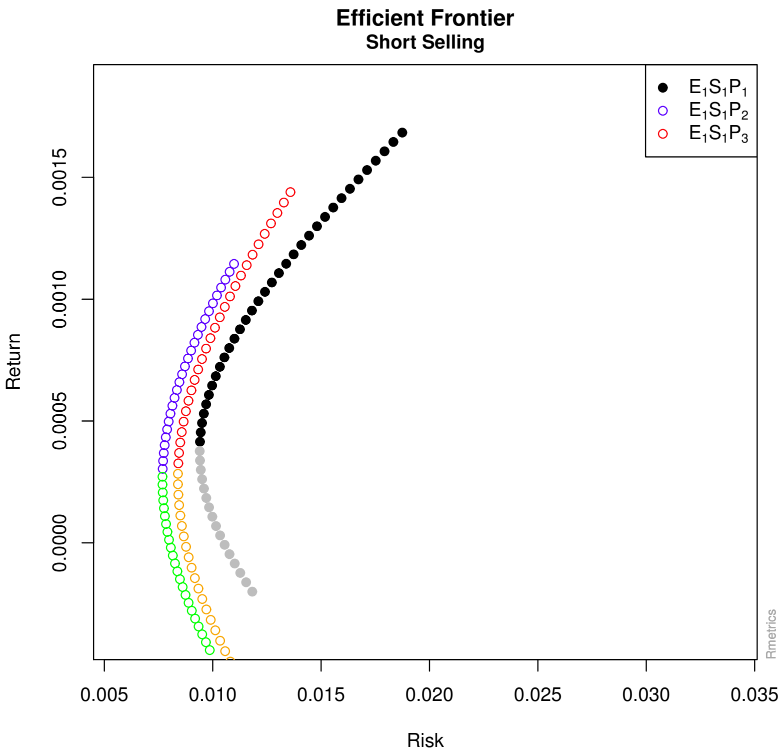

Figure 4.5: Comparison between $E_1S_1P_1$, $E_1S_1P_2$ and $E_1S_1P_3$ efficient frontiers built upon 50 portfolios in each case. Robust portfolios were estimated based on a breakdown point of 22%. Short Selling is allowed.

Figure 4.6 introduces the comparison among $E_1S_1P_k$ MSR portfolios weights, $k = 1, ..., 5$, when estimated at t_w. As expected, it is possible to see high similarities among the robust portfolios weights while they exhibit relevant differences when compared to the classical one.

Keeping these five portfolios unchanged through the backtesting window $[t_w, T]$, Figures 4.7 and 4.8 present estimates of the cumulative excess return ($\widehat{CER_t}$), excess risk ($\widehat{ED_T}$) and excess Sharpe ratio ($\widehat{ES_T}$), defined in (4.1), (4.5) and (4.6), respectively.

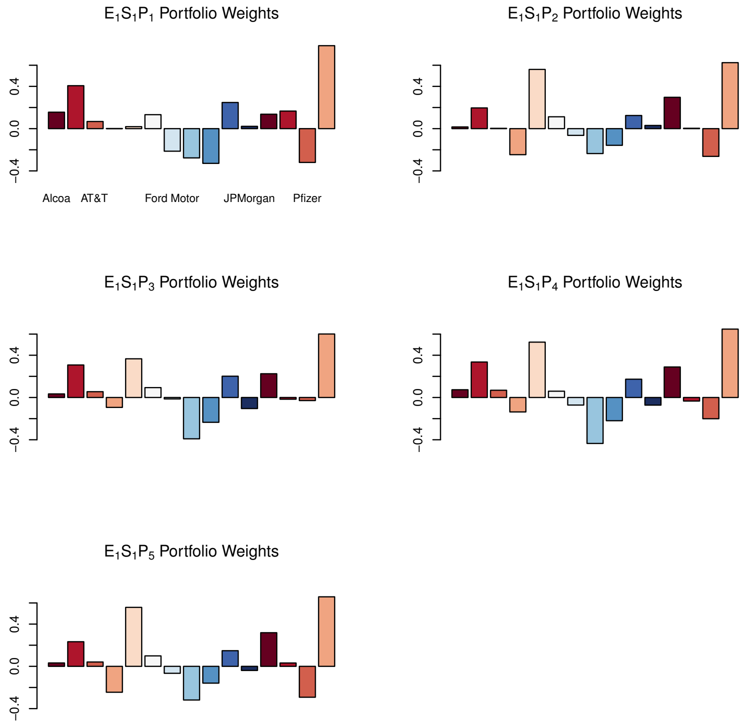

Figure 4.6: Comparison among $E_1S_1P_k$, $k = 1, ..., 5$ portfolio weights. Robust portfolios were estimated based on a breakdown point of 22%. Short Selling is allowed.

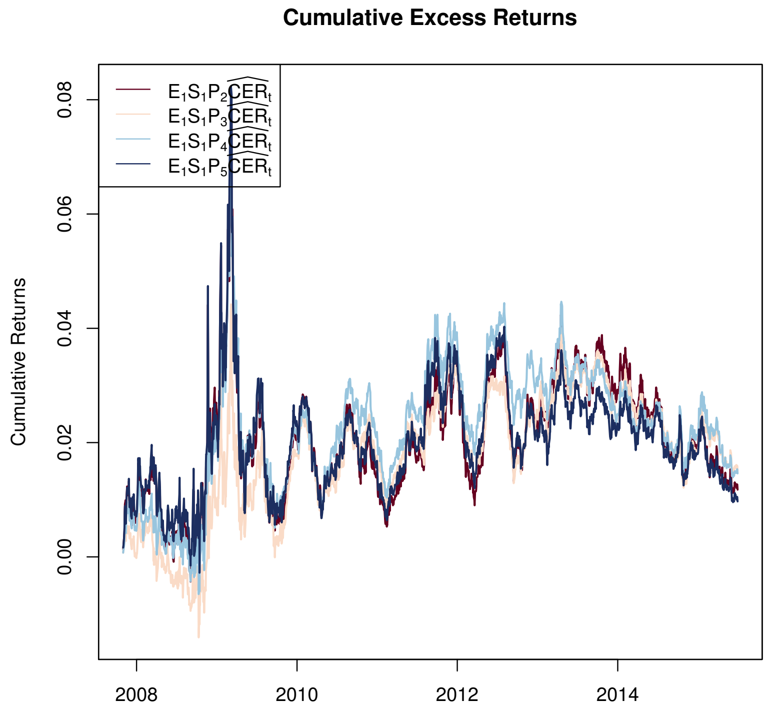

Figure 4.7: Comparison among estimates of Cumulative Excess Returns of MSR robust portfolios ($E_1S_1P_k\widehat{CER}_t$, $k = 2, ..., 5$, $t \in [t_{w+1}, T]$).

Considering the time frame and the dataset chosen, one can see in Figure 4.7 and in Figure 4.8 that all robust portfolios exhibit positive excess cumulative returns, negative excess risk and positive excess Sharpe ratio, with respect to the classical one. Or else, robust portfolios presented better return and risk profiles, but with more pronounced drawdowns than the classical portfolio.

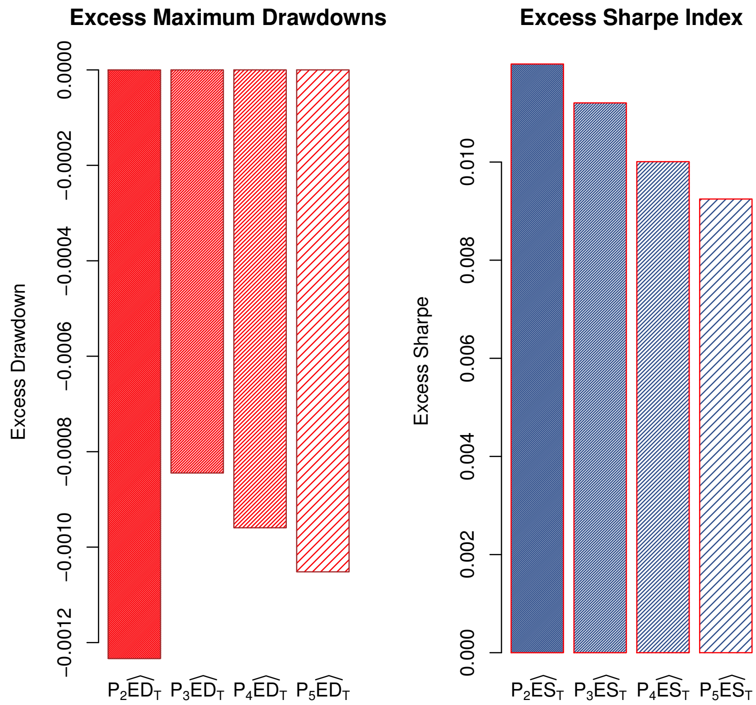

Figure 4.8: (left part) Comparison among estimates of Excess Maximum Drawdowns ($E_1 S_1 P_k \widehat{ED}_T$, $k = 2, ..., 5$) of MSR robust portfolios. (right part) Comparison among estimates of Excess Sharpe Index ($E_1 S_1 P_k \widehat{ES}_T$, $k = 2, ..., 5$) of MSR robust portfolios.

At this point of the work, we acknowledge this apparent superior performance but we do not have proper foundation to assert that this is a fact, regardless of this particular past history of returns. Only by means of simulation, to be performed in environment E_2, we will be able to reach more solid conclusions.

Thus, we formulate the first group of research question in the form of hypothesis to be tested in the simulation exercise.

Hypothesis 1. ***At a given confidence level, can we reject the hypotheses that the real parameters for Excess Sharpe Index of MSR robust portfolios with respect to the classical one are those that were presented in Figure 4.8, when***

considering a buy-and-hold strategy?

$$\begin{aligned}
&\mathbf{H}_{12}: E_1S_1P_2ES_T = 0.0120\\
&\mathbf{H}_{13}: E_1S_1P_3ES_T = 0.0112\\
&\mathbf{H}_{14}: E_1S_1P_4ES_T = 0.0100\\
&\mathbf{H}_{15}: E_1S_1P_5ES_T = 0.0092
\end{aligned} \tag{4.7}$$

4.3.2 $\mathbf{E_1S_2}$ - Rebalancing Strategy on Observed Returns

In this section, MSR optimum portfolios are estimated and rebalanced (reevaluated) at each $t \in T^* = \{t_w, t_w + \Delta t, t_w + 2\Delta t, ..., T - \Delta t\}$, Δt being a specified time length. In this process, we use a rolling-type estimation window, in the sense that for each $t \in T^*|t > t_w$, the oldest Δt values of the sample are deleted while new Δt values arrive. Robust portfolios are estimated considering the breakdown values presented in Figure 4.4.

In this work we chose $\Delta t = 90$ days (quarterly rebalancing). Actually, we tried to shorten Δt, but we faced the challenge of keeping computational effort at manageable levels. One should notice that in the simulated environment E_2, portfolios are also estimated at each $t \in T^*$, but this process is repeated many times over a wide range of simulated past returns. Robust portfolios algorithms rely mostly on resampling procedures and, according to our measurement, estimating $P_1, ..., P_5$ portfolio weights lasts approximately 40 seconds to 1 minute. So, if rebalancing activities were performed more frequently and/or if the number of past simulated returns were higher, then environment E_2 would be an important bottleneck for this study in terms of computational effort.

Figure 4.9 presents estimates of excess total norm weights ($\widehat{ETNW_T}$), as defined in (4.4). All estimates are negative values, which means that robust optimum portfolios are apparently more stable, implying in relatively less transaction costs, considering the particularities of the sample analyzed.

The stability and transaction costs relationship emerges from the norm weights definition itself. Note that, generically speaking, the estimated portfolio weights vector $\widehat{W_t}$ contains the percentage invested in each asset. Aside from a rare coincidence, $\widehat{W_t}$ is different from $\widehat{W_{t-\Delta t}}$, implying that assets were bought and/or sold in the rebalancing process. Evaluating the norm weight $||\widehat{W_t} - \widehat{W_{t-\Delta t}}||_1$, we account for all percentage differences in absolute values, acknowledging the fact that it does not matter if an specific security percentage increased or decreased in the portfolio, by means of buying or selling activities, all else being equal. All of them would be charged, therefore impacting transaction costs.

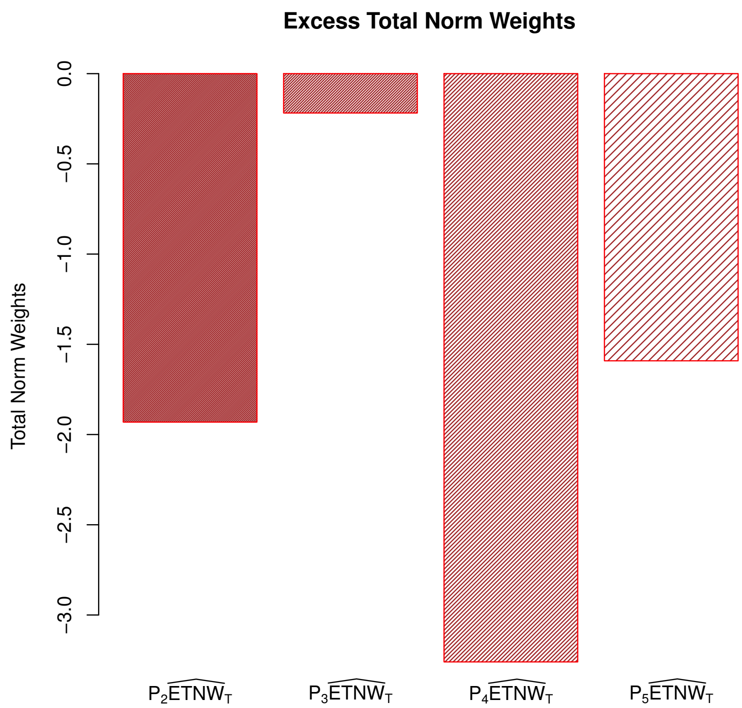

Figure 4.9: Comparison among estimates of Excess Total Norm Weights ($E_1S_2P_k\widehat{ETNW}_T$, $k = 1, ..., 5$) of MSR robust portfolios.

This evidence leads to our second research question formulated in hypothesis 2 which will be tested in environment E_2.

Hypothesis 2. ***At a given confidence level, can we reject the hypotheses that the real parameters for Excess Total Norm Weights of MSR robust portfolios with respect to the classical one are those that were presented in Figure 4.9, when considering a rebalancing strategy?***

$$\begin{aligned} \mathbf{H}_{22} &: E_1S_2P_kETNW_T = -1.9310 \\ \mathbf{H}_{23} &: E_1S_2P_kETNW_T = -0.2181 \\ \mathbf{H}_{24} &: E_1S_2P_kETNW_T = -3.2600 \\ \mathbf{H}_{25} &: E_1S_2P_kETNW_T = -1.5913 \end{aligned} \tag{4.8}$$

Regarding cumulative excess returns ($\widehat{CER}_t$, $t \in T^{**}$), as defined in (4.2), recall that we are assuming that portfolio norm weights are good proxies for transaction costs and, as such, they negatively impact portfolios returns through a proportionality constant c.

Actually, there is a long and widespread debate on the role of transaction costs in the functioning of financial markets. Such costs can be imputed to different reasons, like: brokerage commission fees, bid/ask spreads, time involved in acquiring knowledge and record keeping, transaction taxes and so on. They can be charged as a fixed value per posted buying or selling order, they can be charged as a function of the actual amount bought or sold, they might depend on geographical regions and their development levels, or a combination of different criteria. Therefore, there is a wide range of transaction costs currently being charged in financial markets, while some of its components are even unknown to others than the investor himself.

While we do not have privileged information that supports the choice for an adequate transaction costs "formula", we also claim that this subject is not fundamental for the rest of the work. Unfortunately, the chosen value of c has the potential to change relative performances. The left part of Figure 4.10 presents estimates of $\widehat{CER}_T$ considering $c = 0.001$, while its right part presents the same estimates for $c = 0.1$. It can be seen that, although the rebalancing activities removed the relative superiority of robust portfolios, as measured by cumulative returns, the higher the transaction cost the better the relative performance of these portfolios.

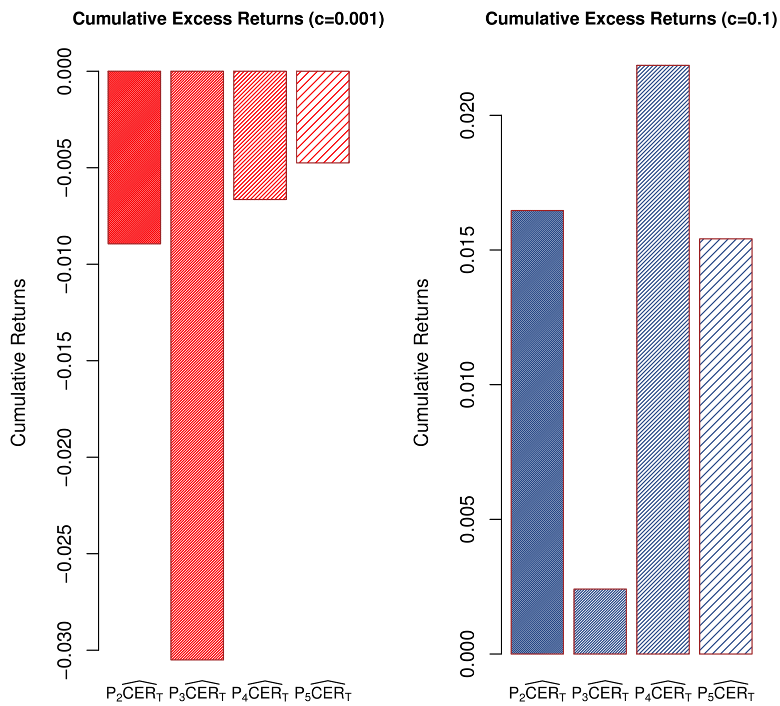

Figure 4.10: (left part) Comparison among estimates of Cumulative Excess Returns ($E_1 S_2 P_k \widehat{CER}_T$, $k = 2, ..., 5$) of MSR robust portfolios, considering $c = 0.001$. (right part) The same estimation, for $c = 0.1$.

In this context, we claim that the core of the portfolios stability lives in the results presented in Figure 4.9, regarding $\widehat{ETNW}_T$, while its consequent impact on performance depends on the chosen formula for transaction costs, even the value for the constant c in our simplistic model. In this work, in order to provide a more complete picture on this subject, we present performance results considering two values of c (0.001 and 0.1).

Finally, note that the growing values of transaction costs might be associated with different financial markets around the world. In an interesting work, the authors in [2] explicitly connected an economy's capital markets efficiency to its level of transactions cost through an inverse relation. They concluded that as the later falls, the general effect is that agents tend to make longer-term, transaction-intensive investments. Consequently, a higher rate of return on savings is possible, as well as a change in its composition. In this sense, the lower value of c presented in this work could be associated to a more efficient financial market, possibly belonging to a developed economy, while the opposite is true concerning the higher value of c.

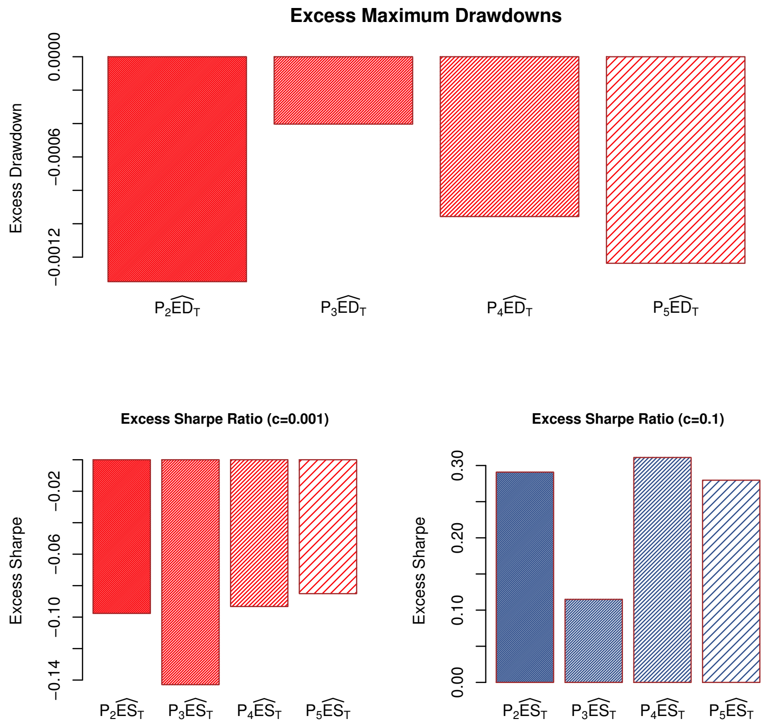

Figure 4.11: (top) Comparison among estimates of Excess Maximum Drawdowns ($E_1S_2P_k\widehat{ED}_T$, $k = 2, ..., 5$) of MSR robust portfolios. (bottom left) Comparison among estimates of Excess Sharpe Ratios ($E_1S_2P_k\widehat{ES}_T$, $k = 2, ..., 5$) of MSR robust portfolios, considering $c = 0.001$. (bottom right) The same estimation, for $c = 0.1$.

The top of the Figure 4.11 presents evidence suggesting that MSR robust portfolios present lower risk profile, as measured by the maximum drawdowns, while the bottom part of this figure complements the performance analysis showing that the ex-post Sharpe Ratio after transaction costs is also severely impacted by the chosen value of c. Again, the higher the cost, the better is the relative performance of robust portfolios. Note that the results plotted at the bottom part of Figure 4.11 are not inconsistent with those presented in Figure 4.10. In the latter, cumulative returns were evaluated at the specific point in time T, while in the former, the Sharpe ratio accounts for the expected return over the whole period.

Finally, we formulate our third research question in hypothesis 3

Hypothesis 3. ***At a given confidence level, can we reject the hypotheses that the real parameters for Excess Sharpe Index after transaction costs (**$c = 0.1$**) of MSR robust portfolios with respect to the classical one are those that were presented in Figure 4.11, when considering a rebalancing strategy?***

$$
\begin{aligned}
\mathbf{H}_{32} &: E_1 S_2 P_k ES_T = 0.2909 \\
\mathbf{H}_{33} &: E_1 S_2 P_k ES_T = 0.1148 \\
\mathbf{H}_{34} &: E_1 S_2 P_k ES_T = 0.3111 \\
\mathbf{H}_{35} &: E_1 S_2 P_k ES_T = 0.2797
\end{aligned}
\tag{4.9}
$$

To sum up the results achieved so far, we can say that MSR robust portfolios exhibited relatively better performance in the buy-and-hold strategy, while in the rebalancing strategy, performance values are impacted and conditioned to transaction costs. Regarding stability, we saw that robust portfolios seem to be more stable than the classical one, notably the S portfolio which seems to present the best stability profile. But of course, in this section we have dealt with one single realization of the subjacent stochastic process that governs this set of stocks behavior and, consequently, the presented results may have occurred simply by chance. Maybe performance is not so good in the buy-and-hold strategy and the same rationale applies to the stability results. Overall, this section served us mainly to build our three research questions that will be answered in the next section by means of simulations.

4.4 Environment E_2 - Simulated Dataset

In this section we implement the simulated environment E_2 with the purpose of testing, under different past trajectories, the three research questions formulated in hypotheses 1, 2 and 3. To that end, we generate several realizations for 15 random variables mimicking the observed behavior of the 15 stock returns collected in environment E_1. All characteristics adopted in E_1 will be kept in environment E_2, in the sense that the length of each simulated time series is the same as the observed in the real data set, as they also are the segregation between estimation and backtesting windows and the rebalancing frequency. Aside from the fact that now we are working on a wide range of different past returns, the same modeling, metrics and premises are used in this section, allowing a smooth and coherent transition between environments E_1 and E_2.

It remains to be answered how we generate random variables realizations resembling the real data. To accomplish this task, we use the semi-parametric bootstrapping procedure adopted in [30].

The bootstrapping method introduced in [11] is a very general resampling procedure for estimating the distributions of statistics based on independent observations. With the purpose to briefly describe the general characteristics of this method, consider that the sample $X_1, X_2, ..., X_n$ consists of realizations of independent random variables following a common distribution function F and that there is an estimate $\hat{\theta}$ of our interest. We would like to know its sampling distribution, but we often face a common problem: we do not know F for sure. The main idea of the parametric bootstrap regards the assumption that F is unknown up to an unknown parameter η, i.e. $F(x|\eta)$. So we simulate data from $F(x|\hat{\eta})$ where $\hat{\eta}$ should be a good estimate of η. When F is completely unknown, one generally resorts to the non-parametric bootstrapping and resampling is performed directly in the empirical c.d.f. F_n.

To be faithful to the data at hand we should be as close as possible to the true multivariate distribution generating the data. In this sense, the authors in [30] propose a semi-parametric bootstrapping procedure that parametrically takes care of all marginal characteristics of the returns data, but also considers the existing dependence structure. Given that we observe n observations of p asset returns in the real data, to obtain their semi-parametric replications we follow the steps below:

- Each series of returns is modeled by an ARMA-GARCH process. Each of the resulting uncorrelated residuals series is unconditionally characterized by an appropriate distribution function.
- From the $n \times p$ matrix containing the ARMA-GARCH residuals, we generate, column by column, a $n \times p$ matrix $\mathbf{R}$ of their respective ranks.
- From the unconditional distributions fitted to each series of residuals we generate $nsim$ random realizations, where $nsim$ is the chosen number of simulations. In other words, $nsim$ $n \times p$ matrices of simulated residuals are generated.
- Each $n \times p$ matrices of simulated residuals leads to a corresponding $n \times p$ matrix of ranks $\mathbf{R}_m$, $m = 1, ..., nsim$.
- The values of each matrix $\mathbf{R}_m$ are rearranged in order to preserve the pairing observed in the original matrix $\mathbf{R}$.
- To each rearranged matrix $\mathbb{R}^*_m$ there is an unique $n \times p$ matrix of simulated residuals that, in last instance, is used to build a new (simulated) time series using the ARMA-GARCH parameters previously estimated.

The first step of the algorithm previously described concerns the ARMA-GARCH fitting. There are many surveys covering the mathematical and statistical properties of ARMA-GARCH models and our intention here is to just briefly review the model and its practical implementation. First of all, note that ARMA-GARCH means that the conditional mean is modeled by an ARMA equation and the conditional variance by a GARCH equation. For an univariate generic time series $\{X_t\}_{t\geq 0}$, these equations are:

$$\begin{aligned} X_t &= \mu + \sum_{i=1}^{m} a_i X_{t-i} + \sum_{j=1}^{n} b_j \varepsilon_{t-j} + \varepsilon_t \\ \varepsilon_t &= z_t \sigma_t \\ \sigma_t^2 &= \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \end{aligned}$$

where, μ is an intercept for the conditional mean equation, a_i, b_j, m and n are the autoregressive and moving average coefficients and orders, respectively, $\{\varepsilon_t\}_{t\geq 0}$ are the innovations of the time series process, $\{z_t\}_{t\geq 0}$ is an i.i.d process with zero mean and unit variance, ω is an intercept for the conditional variance equation and α_i, β_j, p and q are the ARCH and GARCH coefficients and orders, respectively.

Given an observed univariate return series, the maximum log-likelihood method is often used to estimate the parameter set for a specific ARMA-GARCH. In this process, generally $\{z_t\}_{t\geq 0}$ is assumed to follow a standard normal probability function. Other choices frequently adopted in the literature are the standard student-t distribution and the generalized error distribution, a symmetrical unimodal member of the exponential family, such as the double exponential and the Laplace distributions.

Having said that, one should specify the orders m, n, p, q of the ARMA-GARCH model in order to estimate its parameters. Several model selection procedures are cited in the relevant literature and one can say that all of them present benefits and drawbacks. In this work, we adopted the so-called specific-to-general approach[11], conditional to the simultaneous attendance to the following criteria:

- Models should be parsimonious;
- Residuals (difference between observed and fitted values) should not be autocorrelated to an specified number of lags and considering a specific significance level, as judged by the Ljung Box test[12]; and

[11]In general terms, the specific-to-general approach implies that the researcher tries to explain a variable of interest with a simple model, in the first place. The model estimation and consequent diagnostic tests will give indications if the assumed simple model is sufficient or if a more refined model, possibly with more explanatory variables, would be necessary.

[12]In order to briefly describe the Ljung Box autocorrelation test, let $\{\hat{e}_t\}_{0\leq t\leq T}$ be the standardized

- Estimated coefficients should be statistically different from zero up to a chosen significance level.

Table 4.4: Best ARMA(m,n)-GARCH(p,q) specification for the stock returns listed in Table 4.1. In this table, we present the best set $\{m, n, p, q\}$ of autoregressive orders, the degrees of freedom of the standard-t distribution assumed to represent $\{z_t\}_{t\geq 0}$ behavior and the obtained p-values for the Ljung-Box (LB) autocorrelation test, considering 5 lags.

Stock	m	n	p	q	ν	LB p-value
Alcoa	0	0	1	1	6.6425	0.3313
Apple	0	0	1	1	5.1752	0.1169
AT&T	1	1	1	1	7.0672	0.7218
Citigroup	0	0	1	1	6.1120	0.5151
Coca-Cola	0	0	1	1	5.2929	0.9836
Dow Chemical	1	1	1	1	6.8063	0.9091
Ford Motor	0	0	1	1	5.4557	0.2590
General Electric	0	0	1	1	5.8215	0.8801
Intel	0	0	1	1	6.0929	0.7904
J. C. Penney	1	1	1	1	4.6453	0.3531
JPMorgan	1	0	1	1	6.2201	0.5048
Merck	0	0	1	1	4.3754	0.2509
Microsoft	1	0	1	1	4.5331	0.5490
Pfizer	1	1	1	1	6.0251	0.3322
P&G	1	1	1	1	4.8078	0.9052

residuals from fitting a generic time series regression model, and let

$$\hat{r}_k = \frac{\sum_{t=k+1}^{T} \hat{e}_t \hat{e}_{t-k}}{\sum_{t=1}^{T} \hat{e}_t^2} \quad k = 1, 2, ...$$

be their autocorrelations. The M lag Ljung-Box statistic is defined as:

$$Q_M = T(T+2) \sum_{k=1}^{M} \frac{\hat{r}_k^2}{T-k}$$

If there is sufficient lack of autocorrelation, Q_M asymptotically follows a $\chi^2_{(M-d)}$, where d amounts for the number of parameters estimated. The null hypothesis of test assumes that the first M autocorrelations are all equal to zero. In this sense, results obtained in this test are generally accepted as important assistances in the process of concluding about residuals independence.

Table 4.4 presents selected models specification and diagnostics for each of the univariate time series $[t_0, T]$ collected in E_1. More specifically, the following findings are provided: the best autoregressive lag orders (m, n, p, q), the degrees of freedom of the standard-t distribution assumed to represent $\{z_t\}_{t\geq 0}$ behavior (there was no case where the Gaussian distribution provided a better fit than the t distribution) and the obtained p-values for the Ljung-Box autocorrelation test, considering 5 lags in the residuals. As one should notice, these p-values indicate that in all case we can not reject, at comfortable significance levels, the Ljung-Box null hypothesis of lack of autocorrelation in the residuals. We consider that these results offer enough subsidies for the approximately i.i.d residuals assumption required by the bootstrapping methodology.

Table 4.5 presents the parameter estimates for the models specified in Table 4.4. Even though we do not indicate in Table 4.5, for convenience matters, all estimates are statistically different from zero at a 10% significance level.

Table 4.5: Parameters estimates for the best ARMA(m,n)-GARCH(p,q) specifications presented in Table 4.4, or else, the intercepts (μ, ω) of the mean and the volatility equations, respectively, and the estimates for the autoregressive (a_1), moving average (b_1), ARCH (α_1) and GARCH (β_1, β_2) coefficients. All coefficients are statistically significant at 10% significance level.

Stock	(μ, ω)	(a_1, b_1)	(α_1, β_1)
Alcoa	$(0, < 10^{-4})$	$(N/A, N/A)$	$(0.0467, 0.9493)$
Apple	$(0.0016, < 10^{-4})$	$(N/A, N/A)$	$(0.0374, 0.9594)$
AT&T	$(< 10^{-4}, < 10^{-4})$	$(0.8517, -0.8724)$	$(0.0556, 0.9400)$
Citigroup	$(0, < 10^{-4})$	$(N/A, N/A)$	$(0.0869, 0.9156)$
Coca-Cola	$(0.0004, < 10^{-4})$	$(N/A, N/A)$	$(0.0564, 0.9368)$
Dow Chem.	$(0.0013, < 10^{-4})$	$(-0.8818, 0.8656)$	$(0.0713, 0.9212)$
Ford Motor	$(0, < 10^{-4})$	$(N/A, N/A)$	$(0.0659, 0.9308)$
G. E.	$(0, < 10^{-4})$	$(N/A, N/A)$	$(0.0633, 0.9324)$
Intel	$(0.0005, < 10^{-4})$	$(N/A, N/A)$	$(0.0382, 0.9590)$
JC Penney	$(0, < 10^{-4})$	$(-0.5529, 0.5658)$	$(0.0339, 0.9642)$
JPMorgan	$(0.0007, < 10^{-4})$	$(-0.0414, N/A)$	$(0.0710, 0.9276)$
Merck	$(0.0004, < 10^{-4})$	$(N/A, N/A)$	$(0.0782, 0.8995)$
Microsoft	$(0, < 10^{-4})$	$(-0.0352, N/A)$	$(0.0498, 0.9468)$
Pfizer	$(0, < 10^{-4})$	$(0.3991, -0.4266)$	$(0.0732, 0.9148)$
P&G	$(0, < 10^{-4})$	$(0.4901, -0.5386)$	$(0.0507, 0.9391)$

With the above-mentioned results in hand, we move forward to the next steps of the algorithm previously presented, completing the bootstrapping procedure. The resulting multivariate simulated time series preserves the dependence structure observed in the real data and is treated in this chapter as a multivariate return series that did not occurred in the past but plausibly could have occurred.

In the same way we proceeded before, short selling is also allowed in the portfolio selection optimization problem and the assumed investment policy is that investors want to keep MSR portfolios each time they are estimated. Regarding the simulation itself, in both strategies S_1 and S_2, 500 different time series are generated and partitioned into estimation and backtesting windows. As already mentioned before, this choice keeps computational costs at manageable levels.

Finally, although simulated returns are generated from fixed ARMA-GARCH models, presented in Tables 4.4 and 4.5, there is no reason to accept that the contamination levels among different estimation windows are constant. On the contrary, from the simulation purpose itself, it is expected that each simulated series to materially differ from the previous one. In this sense, using a fixed breakdown point regardless of the characteristics presented in each simulated series can be potentially misleading. To keep an adequate balance between robustness and efficiency, we employed in this section the same rationale presented in Section 4.2, in the sense that robust location and scatter estimates are evaluated considering different breakdown points per estimation window. These values are presented in due time in the next subsections.

4.4.1 $\mathbf{E_2S_1}$ - Buy-and-Hold Strategy on Simulated Data

In this subsection we inspect and get a better understanding on the sampling distributions of the statistics estimated in environment E_1S_1 to conclude about the location and dispersion of the respective parameters themselves. The results obtained with 500 different simulated past realizations for 15 random variables are characterized in terms of their descriptive statistics, while for the specific metric questioned in hypothesis 1, confidence intervals are provided and a more intimate analysis is performed.

Following the procedure described in Section 4.2, Figure 4.12 presents the most adequate breakdown point to be used in each of the 500 simulated estimation window. The interval $[13\%, 17\%]$ encompasses the estimated contamination levels and the corresponding adopted breakdown values for the simulated windows.

Table 4.6 presents descriptive statistics for the Cumulative Excess Returns ($P_k\widehat{CER}_T$, $k = 2, ..., 5$) and for the Excess Risk ($P_k\widehat{ED}_T$, $k = 2, ..., 5$) sampling distributions. In the first case, it is not possible to reject at comfortable significance levels, that the sampling distributions follow a Gaussian law. Besides that, one can verify that these distributions are centered at small negative values, indicating that the real parameters $P_k CER_T$, $k = 2, ..., 5$, locate themselves in a symmetrical region around values quite close to zero.

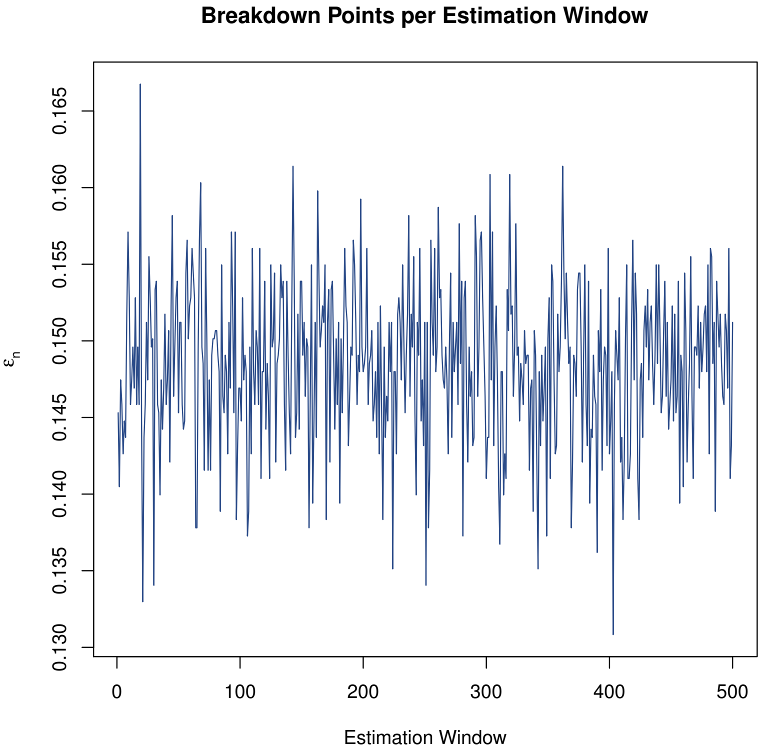

Figure 4.12: Breakdown Points used in each of the 500 simulated estimation window for the strategy E_2S_1. We followed the procedure described in Section 4.2.

Table 4.6: Descriptive statistics for the Cumulative Excess Returns ($E_1S_1P_k\widehat{CER}_T$, $k = 2, ..., 5$) and for the Excess Risk ($E_1S_1P_k\widehat{ED}_T$, $k = 2, ..., 5$) sampling distributions. 500 different simulated past realizations for 15 random variables were generated according to the models specified in Tables 4.4 and 4.5.

Estimator	Mean	Min.	Max.	Std. Dev.	Skewness	Exc. Kurt.	JB p-val
$E_1S_1P_2\widehat{CER}_T$	-0.0221	-0.1191	0.0720	0.0336	0.1301	-0.0315	0.4897
$E_1S_1P_3\widehat{CER}_T$	-0.0037	-0.0717	0.0788	0.0258	-0.0308	-0.3649	0.2598
$E_1S_1P_4\widehat{CER}_T$	-0.0140	-0.0821	0.0654	0.0260	0.1808	-0.0974	0.2353
$E_1S_1P_5\widehat{CER}_T$	-0.0188	0.1070	0.0775	0.0322	0.2836	0.0479	0.0331
$E_1S_1P_2\widehat{ED}_T$	0.0001	-0.0015	0.0046	0.0005	2.0181	17.9864	≈ 0
$E_1S_1P_3\widehat{ED}_T$	0.0001	-0.0009	0.0030	0.0004	1.6887	14.4682	≈ 0
$E_1S_1P_4\widehat{ED}_T$	0.0001	-0.0012	0.0031	0.0004	1.7121	11.8140	≈ 0
$E_1S_1P_5\widehat{ED}_T$	0.0001	-0.0011	0.0037	0.0004	1.5497	9.1264	≈ 0

Regarding Excess Risk, sampling distributions are clearly non-gaussian and present non-negligible right skewness that can potentially have inflated the mean values. In summary, these results do not indicate clear performance gains due to the usage of robust estimators.

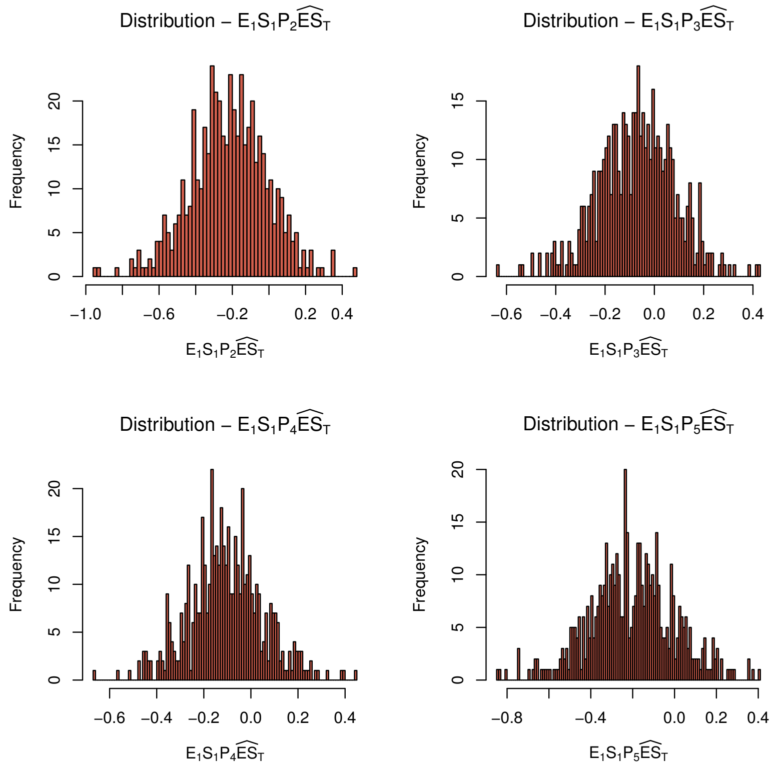

Figure 4.13: Sampling distributions for Excess Sharpe Ratio ($E_1S_1P_k\widehat{ES}_T$, $k = 2, ..., 5$) of MSR robust portfolios, considering 500 past simulated trajectories and the models specified in Tables 4.4 and 4.5.

Table 4.7: Descriptive statistics for the Excess Sharpe ($E_1S_1P_k\widehat{ES}_T$, $k = 2, ..., 5$) sampling distributions. 500 different simulated past realizations for 15 random variables were generated according to the models specified in Tables 4.4 and 4.5.

Estimator	Mean	Min.	Max.	Std. Dev.	Skewness	Exc. Kurt.	JB p-val
$E_1S_1P_2\widehat{ES}_T$	-0.2183	-0.9488	0.4744	0.2103	-0.1156	0.3612	0.1323
$E_1S_1P_3\widehat{ES}_T$	-0.0717	-0.6306	0.4283	0.1571	-0.1285	0.4344	0.0618
$E_1S_1P_4\widehat{ES}_T$	-0.1045	-0.6675	0.4435	0.1584	0.0805	0.4933	0.0522
$E_1S_1P_5\widehat{ES}_T$	-0.2035	-0.8495	0.4043	0.2051	-0.0435	0.2978	0.3372

Regarding the Excess Sharpe Ratio ($P_k\widehat{ES}_T$, $k = 2, ..., 5$) sampling distributions, Table 4.7 and Figure 4.13 present their main features. It is possible to observe that robust portfolios are not rewarding in a buy-and-hold strategy, comparatively to the classical one, as the Excess Sharpe sampling distributions are fairly symmetrical around small negative values. The real unknown parameters do not locate themselves in "profitable" regions to the point that justify the use of a more sophisticated estimation methodology.

In (4.10) we present the 95% confidence intervals built on the $\widehat{ES}_T$ sample distributions. Regarding hypothesis 1 formulated in environment E_1, conclusions are provided in (4.11).

$$\begin{aligned}
E_1S_1P_2\widehat{ES}_T &\in [-0.6453, 0.1772] \\
E_1S_1P_3\widehat{ES}_T &\in [-0.4023, 0.2108] \\
E_1S_1P_4\widehat{ES}_T &\in [-0.4317, 0.2137] \\
E_1S_1P_5\widehat{ES}_T &\in [-0.6362, 0.1981]
\end{aligned} \tag{4.10}$$

$$\begin{aligned}
\mathbf{H}_{12} &: E_1S_1P_2ES_T = 0.0120 && \text{Not Rejected} \\
\mathbf{H}_{13} &: E_1S_1P_3ES_T = 0.0112 && \text{Not Rejected} \\
\mathbf{H}_{14} &: E_1S_1P_4ES_T = 0.0100 && \text{Not Rejected} \\
\mathbf{H}_{15} &: E_1S_1P_5ES_T = 0.0092 && \text{Not Rejected}
\end{aligned} \tag{4.11}$$

Note that it is not possible to reject, at 5% significance level, the higher performance achieved by all robust portfolios, comparatively with P_1. However, the key result presented in this subsection surrounds the symmetrical shape of $\widehat{ES}_T$ sampling distributions around small negative values, implying that the real ES_T parameters locate themselves in a region that positive or negative values occur, with fairly equal probabilities. Note that there is evidence of robust estimation performance superiority in the specific history told by the past returns collected in the environment E_1. However, in a different context, an adverse number of past returns for the same set of stocks could lead to negative excess Sharpe index for robust portfolios and we would not be able to reject that. In conclusion, results are not striking convincing favoring robust estimation in a buy-and-hold strategy.

In some sense, the inconclusiveness embedded in this section's results corroborates what was said about Figure 4.5. In that occasion we claimed that robust efficient frontiers surround the classical one and that it would be possible to allocate assets in a way that optimum robust portfolios would enjoy less risk for the same level of expected return or, conversely, attain higher levels of return for a given level of risk, comparatively to the classical portfolio. In other words, positive excess Sharpe ratios. But we also stated that this evidence would be extremely important only if the unrealistic premise that future returns resemble past returns holds. This would be the case where robust portfolios would guarantee better performance and risk profiles.

Figure 4.5 was built based on past returns and in the buy and hold strategy, portfolios are not rebalanced when future returns deviate form the investor perspectives. This strategy does not explore the potential of robust estimation to its full extent, as it is done only one time, at t_w, in a period of 8 years. Next section presents the reverse occasion, when new information is relevant for the portfolios composition.

4.4.2 $\mathbf{E_2S_2}$ - Rebalancing Strategy on Simulated Data

In this subsection we investigate the sampling distributions of the metrics estimated in environment E_1S_2 (rebalancing strategy), to conclude about the respective parameters. Results are presented following the same criteria used before, or else, descriptive statistics are provided for general metrics, while for those specifically questioned in hypotheses 2 and 3, confidence intervals are developed and a deeper analysis is performed.

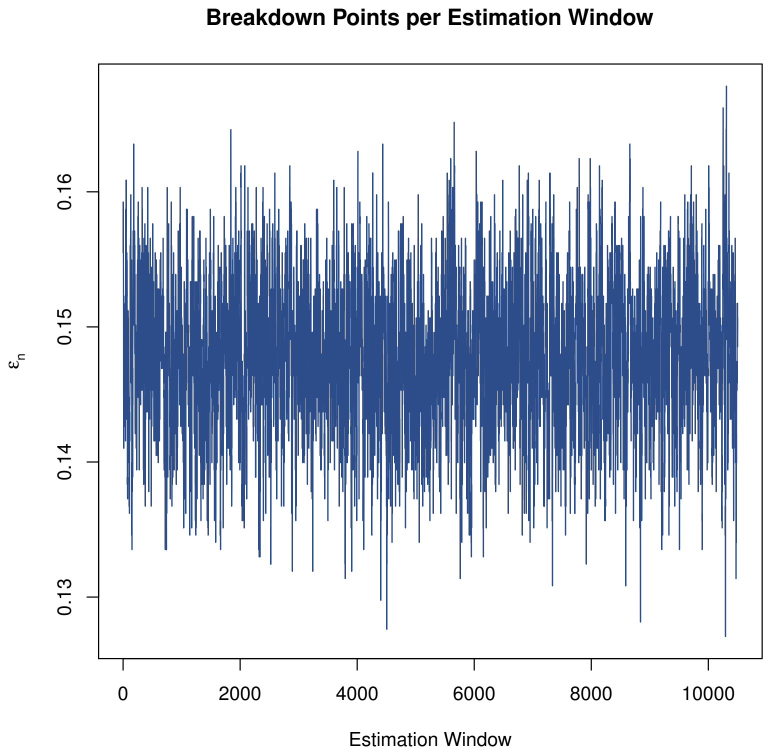

Figure 4.14: Breakdown Points used in each of the 10500 simulated estimation window for the strategy E_2S_2. We followed the procedure described in Section 4.2.

In the same way we proceeded before, Figure 4.14 presents the most adequate breakdown point to be used in each of the 10500 simulated estimation window (21 estimation windows due to rebalancing activities repeated 500 times due to different simulated trajectories). The interval $[13\%, 17\%]$ contains the estimated contamination levels for the simulated windows and the corresponding adopted breakdown values.

Figure 4.15 presents sampling distributions for Excess Total Norm Weights ($P_k\widehat{ETNW}_T$, $k = 2, ..., 5$) of MSR robust portfolios. By simple inspection, one can see that these distributions concentrate mass at the negative portion of the graphics, where negative excess total norm weights translate themselves to more stable portfolios.

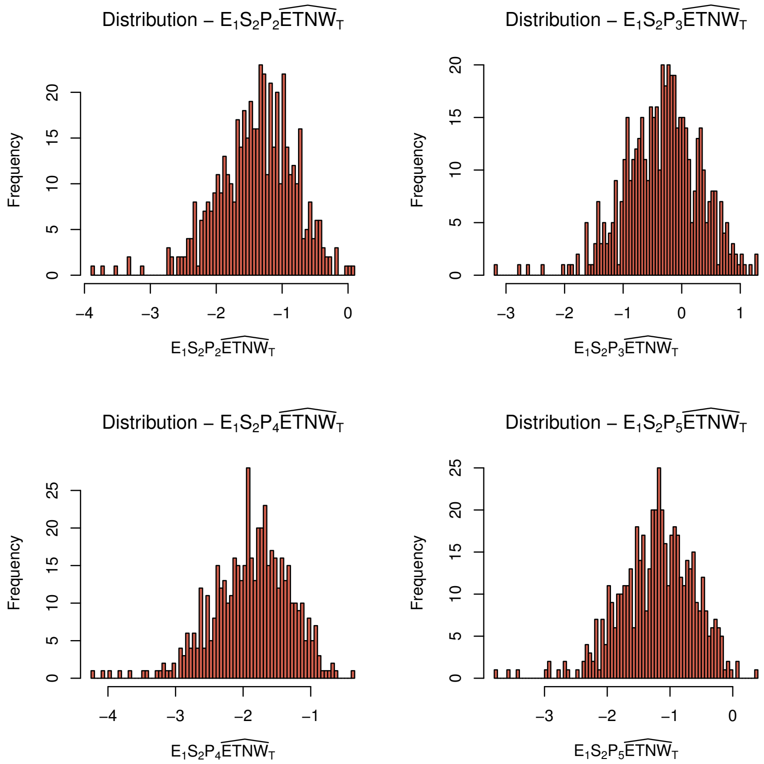

Figure 4.15: Sampling distributions for Excess Total Norm Weights ($E_1S_2P_k\widehat{ETNW}_T$, $k = 2, ..., 5$) of robust portfolios, considering 500 past simulated trajectories and the models specified in Tables 4.4 and 4.5.

Table 4.8: Descriptive statistics for the Excess Total Norm Weights ($E_1S_2P_k\widehat{ETNW}_T$, $k = 2, ..., 5$) sampling distributions. 500 different simulated past realizations for 15 random variables were generated according to the models specified in Tables 4.4 and 4.5.

Estimator	Mean	Min.	Max.	Std. Dev.	Skewness	Exc. Kurt.	JB p-val
$E_1S_2P_2\widehat{ETNW}_T$	−1.3918	−3.8812	0.0532	0.5813	−0.5683	1.0973	≈ 0
$E_1S_2P_3\widehat{ETNW}_T$	−0.3267	−3.1575	1.2807	0.6391	−0.4330	0.9075	≈ 0
$E_1S_2P_4\widehat{ETNW}_T$	−1.8761	−4.2071	−0.3649	0.5698	−0.5949	0.9361	≈ 0
$E_1S_2P_5\widehat{ETNW}_T$	−1.2088	−3.7501	0.3765	0.5910	−0.5611	0.9821	≈ 0

Table 4.8 provides descriptive statistics for these sampling distributions. All distributions are fairly symmetrical around negative mean values, which corroborates the above-mentioned stability. Note how the vast majority of simulated $P_k\widehat{ETNW}_T$ are, indeed, negative values. The percentage of negative simulated $P_k\widehat{ETNW}_T$ in each of the four sampling distribution is, respectively: 99.6%, 70.2%, 100% and 99.4%. Except the

P_3 portfolio, the other robust portfolios present more than 99% of negative simulated excess total norm weight. In the P_4 portfolio, all of the 500 simulations produced more stable portfolios comparatively to the classical one. In this sense, we consider that the S robust portfolio presents the more interesting stability profile, as the corresponding 95% confidence interval for the real parameter $ETNW_T$ is negatively more pronounced than those obtained from the other robust estimations.

The more stability presented by robust portfolios is undeniably the most important result achieved in this work. In fact, during the course of the several estimations performed, It was possible to verify the perverse effect caused by data contamination in the attraction and inflation of the classical estimates. This ends up translating itself in higher portfolio weights instability and, consequently, higher transaction costs. It is clear that the explanation to the more stability fact resides in the high breakdown properties of robust estimators. Robust portfolios weight vectors do not greatly change at rebalancing days, because new transitory turbulent information is generally not taken into account, preserving the bulk of the data to the greatest extent possible. Only persistent outlying behavior, or else, permanent regime changes, has the potential to shift location and scatter estimates to completely new different values, which by its turn, would impact the norm weights.

Regarding the second research question proposed in this work, the 95% confidence intervals built on the $\widehat{ETNW}_T$ sample distributions are presented in (4.12), while the respective conclusion are provided in (4.13).

$$\begin{aligned} E_1S_2P_2\widehat{ETNW}_T &\in [-2.5553, -0.4081] \\ E_1S_2P_3\widehat{ETNW}_T &\in [-1.6069, 0.8289] \\ E_1S_2P_4\widehat{ETNW}_T &\in [-3.0688, -0.9293] \\ E_1S_2P_5\widehat{ETNW}_T &\in [-2.3698, -0.2062] \end{aligned} \tag{4.12}$$

$$\begin{aligned} &\mathbf{H}_{22} : E_1S_2P_2ETNW_T = -1.9310 && \text{Not Rejected} \\ &\mathbf{H}_{23} : E_1S_2P_3ETNW_T = -0.2181 && \text{Not Rejected} \\ &\mathbf{H}_{24} : E_1S_2P_4ETNW_T = -3.2600 && \text{Rejected} \\ &\mathbf{H}_{25} : E_1S_2P_5ETNW_T = -1.5913 && \text{Not Rejected} \end{aligned} \tag{4.13}$$

At 5% significance level, $\mathbf{H}_{24}$ is the only rejected hypothesis. However this does not imply that P_4 robust portfolios are less stable than the classical one. On the contrary, we have already presented that this portfolio has the best stability profile among other robust portfolios. The value of -3.2600 estimated considering the real data collected was just an atypical one, according to the simulation performed.

Table 4.9 presents descriptive statistics for the $E_1S_2P_k\widehat{ED}_T$, $k = 2, ..., 5$, sampling distributions. Note that, even though Figure 4.11 presented negative values for $P_k\widehat{ED}_T$ estimates performed in environment E_1, the simulation results exhibited in Table 4.9 do not corroborate that robust portfolios are always riskier than the classical one. Actually, $P_k\widehat{ED}_T$ sampling distributions can be assumed to be symmetrical around zero and, as so, there is little that can be said about the risk characteristics of robust portfolios in a rebalancing strategy.

Table 4.9: Descriptive statistics for the Excess Risk ($E_1S_2P_k\widehat{ED}_T$, $k = 2, ..., 5$) sampling distributions. 500 different simulated past realizations for 15 random variables were generated according to the models specified in Tables 4.4 and 4.5.

Estimator	Mean	Min.	Max.	Std. Dev.	Skewness	Exc. Kurt.	JB p-val
$E_1S_2P_2\widehat{ED}_T$	$< 10^{-4}$	-0.0015	0.0013	0.0003	-0.0187	3.6434	≈ 0
$E_1S_2P_3\widehat{ED}_T$	$< 10^{-4}$	-0.0013	0.0009	0.0002	-0.2920	4.2440	≈ 0
$E_1S_2P_4\widehat{ED}_T$	$< 10^{-4}$	-0.0011	0.0007	0.0002	-0.5854	3.1824	≈ 0
$E_1S_2P_5\widehat{ED}_T$	$< 10^{-4}$	-0.0013	0.0011	0.0003	-0.0849	3.1758	≈ 0

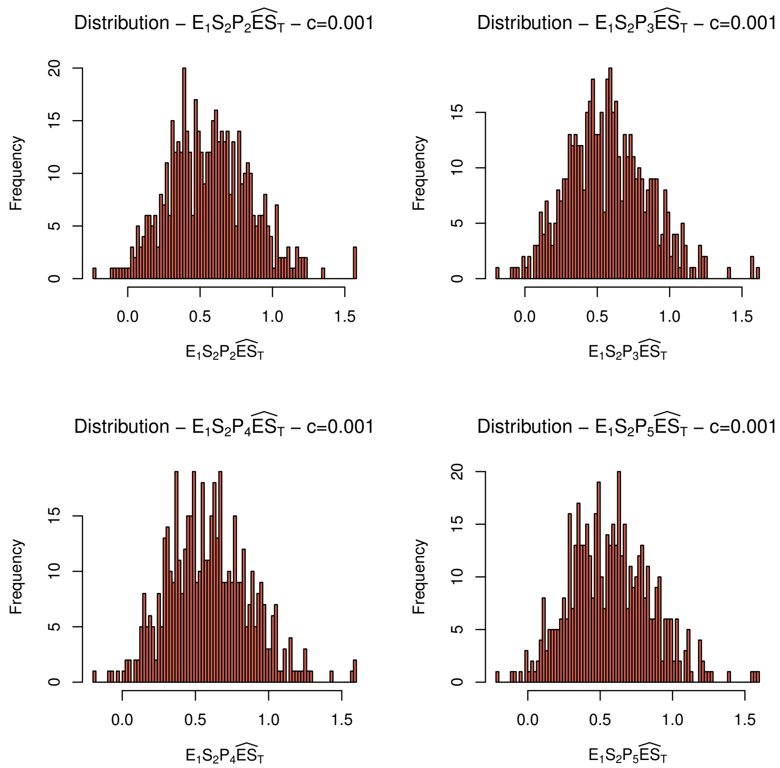

Figure 4.16: Sampling distributions for Excess Sharpe Ratio ($E_1S_2P_k\widehat{ES}_T$, $k = 2, ..., 5$) of MSR robust portfolios, considering 500 past simulated trajectories, $c = 0.001$ and the models specified in Tables 4.4 and 4.5.

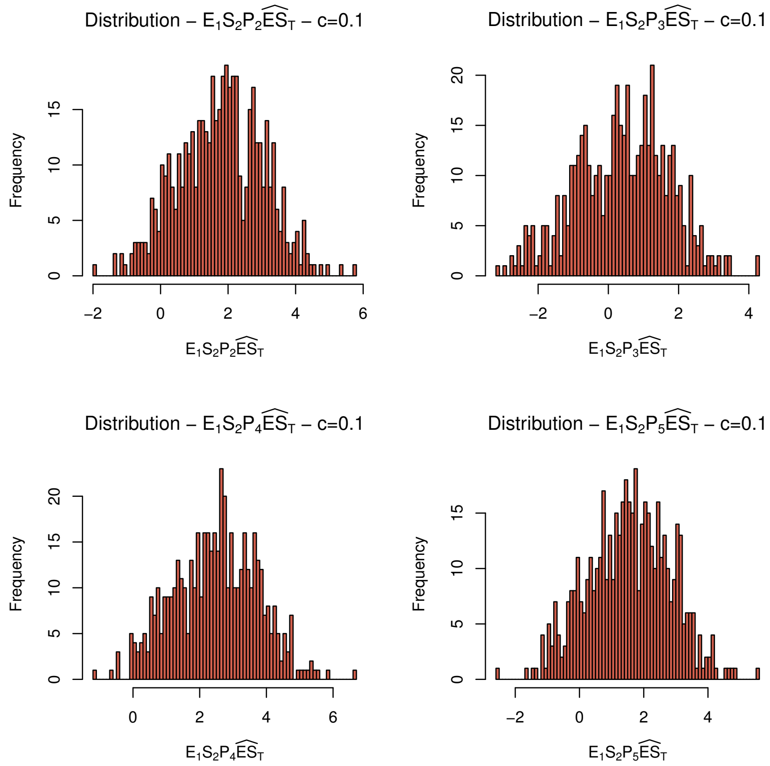

Figure 4.17: Sampling distributions for Excess Sharpe Ratio ($E_1S_2P_k\widehat{ES}_T$, $k = 2, ..., 5$) of MSR robust portfolios, considering 500 past simulated trajectories, $c = 0.1$ and the models specified in Tables 4.4 and 4.5.

Figures 4.16 and 4.17 present the Excess of After Cost Sharpe Ratio sampling distributions, $E_1S_2P_k\widehat{ES}_T$, $k = 2, ..., 5$, considering $c = 0.001$ and $c = 0.1$, respectively. Comparing both figures, its is clear that in the second case, most distributions exhibit performance improvement. Indeed, for $c = 0.1$, more weight is applied to transaction costs and, due to the more stability presented by robust portfolios, there is a greater chance that the real $\widehat{ES}_T$ parameter to be even more rewarding, than there is when $c = 0.001$.

Table 4.10: Descriptive statistics for the Excess After Cost Sharpe Ratio ($E_1S_2P_k\widehat{ES}_T$, $k = 2, ..., 5$) sampling distributions, considering $c = 0.1$. 500 different simulated past realizations for 15 random variables were generated according to the models specified in Tables 4.4 and 4.5.

Estimator	Mean	Min.	Max.	Std. Dev.	Skewness	Exc. Kurt.	JB p-val
$E_1S_2P_2\widehat{ES}_T$	1.7812	−1.9451	5.7765	1.2542	−0.0273	−0.2415	0.5574
$E_1S_2P_3\widehat{ES}_T$	0.3916	−3.1410	4.2584	1.3287	−0.1254	−0.3156	0.1965
$E_1S_2P_4\widehat{ES}_T$	2.4700	−1.1293	6.6300	1.2597	−0.0208	−0.3022	0.4056
$E_1S_2P_5\widehat{ES}_T$	1.5439	−2.5087	5.5848	1.2812	−0.0548	−0.2877	0.3968

Finally, recall that hypothesis 3 was formulated solely for $c = 0.1$. In this respect, Table

4.10 presents descriptive statistics for the $E_1S_2P_k\widehat{ES}_T$, $k = 2, ..., 5$ sampling distributions and, in (4.14) and (4.15), we provide the respective 95% confidence intervals for the real parameters and the hypothesis tests results.

$$\begin{aligned} E_1S_2P_2\widehat{ES}_T &\in [-0.6404, 4.1555] \\ E_1S_2P_3\widehat{ES}_T &\in [-2.3162, 2.7728] \\ E_1S_2P_4\widehat{ES}_T &\in [0.0303, 4.7243] \\ E_1S_2P_5\widehat{ES}_T &\in [-0.9154, 3.9647] \end{aligned} \tag{4.14}$$

$$\begin{aligned} &\mathbf{H}_{32} : E_1S_2P_2ES_T = 0.2909 && \text{Not Rejected} \\ &\mathbf{H}_{33} : E_1S_2P_3ES_T = 0.1148 && \text{Not Rejected} \\ &\mathbf{H}_{34} : E_1S_2P_4ES_T = 0.3111 && \text{Not Rejected} \\ &\mathbf{H}_{35} : E_1S_2P_5ES_T = 0.2797 && \text{Not Rejected} \end{aligned} \tag{4.15}$$

First of all, one can see in Table 4.10 that all sampling distributions are fairly symmetrical around positive values, which indicates "rewarding" regions for the location of the unknown ES_T parameters. Indeed, except the P_3 portfolio, the 95% confidence intervals exhibited in (4.14) have lower bounds much more smaller, in absolute values, than the higher bounds. This exception is consistent with the results presented in Figure 4.15, where it is possible to verify that P_3 portfolio exhibited the worst stability profile.

This indicates that there is evidence suggesting that the higher the transaction costs the better the relative performance of most MSR robust portfolios comparatively with the classical one. As we have already mentioned, the authors in [2] related high transaction costs to less efficient financial markets, possibly in less developed economies. In these cases, robust estimation might be an important tool to be used.

4.5 Sensitivity Analysis

In this chapter, several decisions have been taken through the various estimation processes that potentially have impacted the achieved results. The adopted rebalancing frequency, the number of simulations performed, the chosen value for the breakdown point, the type of efficient portfolio analyzed and the portfolio optimization constraints are, among others, important variables that should be included in a sensitivity exercise in order to get a complete performance and stability picture of robust portfolios.

We have already presented a sensitivity analysis for the portfolios performance, considering two different values for the transaction costs c. Also, the chosen values for breakdown point were justified based on the procedure described in Section 4.2. In this section we extend this analysis for two important variables in order to "robustify" our conclusions and to enlarge the range of investors and policies covered by this work. They are:

- **The type of efficient portfolio analyzed:** During this chapter, our analysis focused on the Maximum Sharpe Ratio efficient portfolio (MSR). Intuitively, this portfolio seems to be a good option for investors, as the Sharpe ratio consolidates into one single number the portfolio risk adjusted return information. However, the set of strategies, philosophies and policies followed by investors in different financial markets is so wide that it would be virtually impossible to contemplate most of them in one single work. Nevertheless, there is another efficient portfolio that is very popular among practitioners and stands out from the others: the global minimum variance portfolio (GMV). As its name says, this portfolio has the lowest possible volatility among all efficient portfolios. In this sense, the next paragraphs present the robust GMV portfolios performance and stability comparatively to the GMV traditional mean-variance portfolio.

- **Short selling constraint:** In all portfolio optimization problems performed in this chapter short selling was allowed. Indeed, short selling transactions are frequent in financial markets and may be motivated by a variety of objectives. Among others, speculators may sell short in the hope of realizing a profit on an instrument that appears to be overvalued and traders or fund managers may hedge a long position or a portfolio through one or more short positions.

  However, short selling adds risks that do not exist in traditional buying and selling positions. First of all, "long's" losses are limited, because the stock price can not be negative, but gains are not, as there is no limit, in theory, on how high the price can go. The reverse occurs to a short selling position, as gains are limited, whereas losses can be unlimited, in theory. Besides that, a given stock may become "hard to borrow" based on lack of availability. In these cases, a broker may charge a "hard to borrow" fee daily, without notice. These and other risks are well documented in the literature and mainly because of them, regulators around the world often impose restrictions to the short selling activity. In this sense, relaxing this assumption may enhance the target audience of this work.

Finally, note that the S_2 rebalancing strategy provides a more complete picture in terms of stability and performance than the S_1 strategy (only performance). Also, we have established in previous sections that relative performance of robust portfolios with respect to the classical one is highly sensitive to transaction costs. In this sense, we choose to solely present in this section the stability results for GMV portfolios and for MSR portfolios built with no permission for short-selling, under the S_2 strategy.

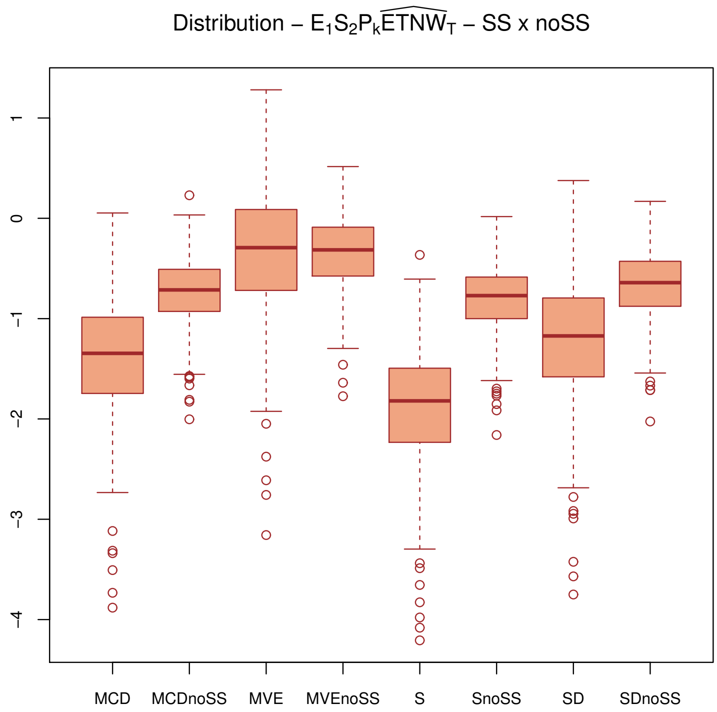

Figure 4.18: Box plots for the sampling distributions of Excess Total Norm Weights ($P_k\widehat{ETNW}_T$, $k = 2, ..., 5$), considering MSR portfolios with and without short selling.

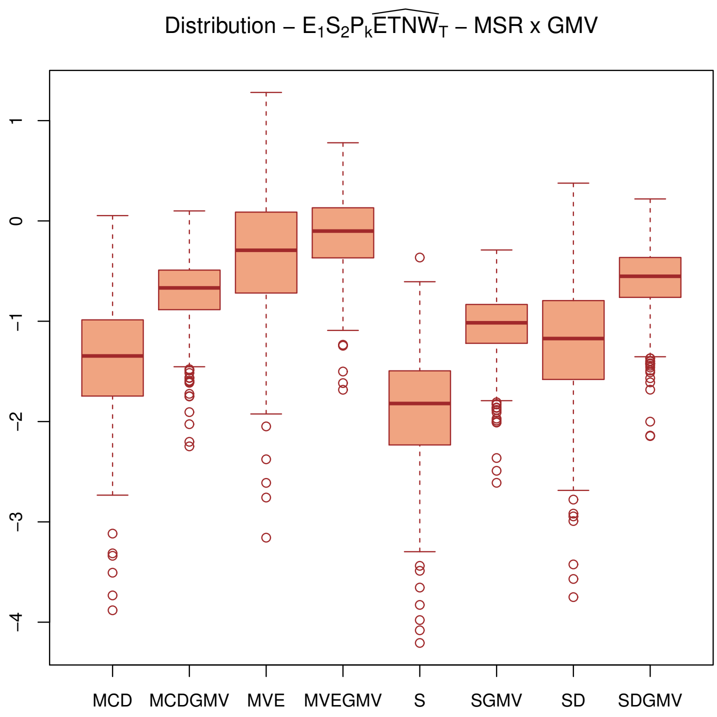

Figure 4.19: Box plots for the sampling distributions of Excess Total Norm Weights ($P_k\widehat{ETNW}_T$, $k = 2, ..., 5$), considering MSR and GMV portfolios with short selling.

Table 4.11: Descriptive statistics for the Excess Total Norm Weights ($E_1S_2P_k\widehat{ETNW}_T$, $k = 2, ..., 5$) sampling distributions of MSR and GMV portfolios built with short selling (MSR-SS and GMV-SS) and MSR portfolios without short selling (MSR-noSS). 500 different simulated past realizations for 15 random variables were generated according to the models specified in Tables 4.4 and 4.5.

Estimator	Mean	Min.	Max.	Std. Dev.	Skew.	Kurt.	JB p-val
$P_2\widehat{ETNW}_T$ (MSR-SS)	−1.3918	−3.8812	0.0532	0.5813	−0.5683	1.0973	≈ 0
$P_3\widehat{ETNW}_T$ (MSR-SS)	−0.3267	−3.1575	1.2807	0.6391	−0.4330	0.9075	≈ 0
$P_4\widehat{ETNW}_T$ (MSR-SS)	−1.8761	−4.2071	−0.3649	0.5698	−0.5949	0.9361	≈ 0
$P_5\widehat{ETNW}_T$ (MSR-SS)	−1.2088	−3.7501	0.3765	0.5910	−0.5611	0.9821	≈ 0
$P_2\widehat{ETNW}_T$ (MSR-noSS)	−0.7337	−2.0037	0.2296	0.3355	−0.4011	0.3831	0.0002
$P_3\widehat{ETNW}_T$ (MSR-noSS)	−0.3477	−1.7735	0.5158	0.3593	−0.4419	0.2701	0.0001
$P_4\widehat{ETNW}_T$ (MSR-noSS)	−0.8107	−2.1601	0.0173	0.3238	−0.5821	0.5980	$< 10^{-4}$
$P_5\widehat{ETNW}_T$ (MSR-noSS)	−0.6658	−2.0253	0.1687	0.3409	−0.4817	0.3227	$< 10^{-4}$
$P_2\widehat{ETNW}_T$ (GMV-SS)	−0.7158	−2.2472	0.0998	0.3495	−0.8856	1.6369	≈ 0
$P_3\widehat{ETNW}_T$ (GMV-SS)	−0.1437	−1.6832	0.7792	0.3767	−0.6638	1.0227	≈ 0
$P_4\widehat{ETNW}_T$ (GMV-SS)	−1.0595	−2.6107	−0.2898	0.3369	−0.9010	1.5882	≈ 0
$P_5\widehat{ETNW}_T$ (GMV-SS)	−0.5992	−2.1450	0.2192	0.3466	−0.9034	1.7564	≈ 0

Results are compared with the main case studied in this work, MSR portfolios with allowance for short selling. The associated performance results, although omitted, can be easily inferred by the stability results and by the previous argument. The chosen breakdown values per estimation window are the same as those presented in Figure 4.14, as all simulations were jointly performed.

Having said that, Figure 4.18 presents the box plots for the sampling distributions of Excess Total Norm Weights ($P_k\widehat{ETNW}_T$, $k = 2, ..., 5$), considering MSR portfolios with and without short selling. Figure 4.19 presents the same comparison but for sampling distributions obtained from MSR and GMV portfolios, both with short selling. In other words, in the first comparison we assume that MSR portfolios are imposed by investors' preference but we analyze the impact that short selling, a restriction in the optimization problem, could have on the results. In the second comparison, short selling is allowed, and we investigate the effect of keeping a complete different portfolio. Table 4.11 consolidates the descriptive statistics for all of these distributions.

A few remarks are worth noting regarding these results:

1. It is clear that in all policy investments robust portfolios prove to be more stable

than the classical one, which implies in less transaction costs irrespective of the studied cases. The stability property seems to be robust to variations of the main case studied in this work.

2. GMV portfolios excess transaction costs seem to be more predictable, as the real Excess Total Norm Weights parameter locates in a narrower region than those verified in the other policy investments. This fact is a combination of the robustness properties intrinsic to the used robust estimators and the stability property of the GMV portfolio itself. It is well-known the fact that GMV portfolios evaluation does not need the estimated expected returns and, as so, it is less sensitive to estimation risk.

3. In an environment with barriers to the short selling activities, investors who choose MSR portfolios built upon this consideration incur in more predictable transaction costs than their counterparts who engage in short selling. This can be regarded to the lower degrees of freedom when portfolio weights are restricted to be positive. The perverse inflation effect that contamination can have on classical estimates is more severe when portfolio weights are free to vary in the real line.

4. If predictability is not an important issue, than the leading case investigated in this work, MSR portfolios with short selling, can be an interesting alternative. Note in Table 4.11 that the maximum value obtained in all sampling distributions do no greatly differ, but the MSR-SS minimum values are the most rewarding one. Given that we do not know where the real parameter $ETNW_T$ locates and the purpose of sampling distributions is to give an idea of this location, the MSR-SS sampling distributions are more stretched to the left, yielding more desirable possible regions, in the sense of even less transaction costs.

5. In the same line to what was concluded in previous sections of this chapter, we consider that the S robust portfolios (P_4) present the more interesting stability profile. Confidence intervals for the real parameter $ETNW_T$ built from S estimators are negatively more pronounced than those obtained from the other robust estimations. This fact is not conditioned to the above-mentioned predictability concerns and, even in the GMV case (more predictability), S portfolios seem to be a better choice.

CHAPTER 5

Final Remarks

In this work, we investigated the potential benefits derived from the application of robust estimation to the asset allocation problem. More specifically, we replaced the sample estimates of location and scatter, the conventional mean-variance portfolio problem inputs, by robust alternatives, namely the high breakdown point, affine equivariant Minimum Volume Ellipsoid (MVE), Minimum Covariance Determinant (MCD), S and Stahel-Donoho (SD) multivariate estimators.

In this sense, Chapters 2 and 3 presented theoretical aspects of robust estimation and modern portfolio theory, respectively. Using those concepts, in Chapter 4 we estimated robust and classical portfolios based both on observed and simulated data, considering both buy-and-hold and rebalancing strategies. The assumed investor's policy investment dictates a preference for the maximum Sharpe ratio efficient portfolio (MSR) and that there are no barriers to the short selling activity. These two premises were relaxed in a sensitivity analysis exercise. In all cases, portfolios performance and stability were assessed through a set of defined measures.

In our opinion, there are not many papers approaching the optimum portfolio problem under the robustness perspective. Most of them investigate one or two robust estimators and it is possible to verify a narrow controversy regarding their results. We consider that this work innovates and aggregates to the academic discussion mainly in the following matters:

- Breakdown points for the robust estimators were dynamically determined, in the sense that they were not set as fixed values regardless of the corresponding data sample used for estimation purposes. Mainly in rebalancing activities, estimation windows change as new information becomes available. Allowing breakdown points to freely vary attempts to increase the estimator's level of efficiency, without sacrificing the level of robustness required by the existing contamination in each window;

- The traditional ε-contaminated model extensively used in previous works, regarding simulation exercises, was replaced by a resampling technic, based on the semi-parametric bootstrapping procedure adopted in [30]. In this way, hypotheses tests on the portfolios performance and stability unknown parameters were formally addressed and a thorough framework is provided for investors who want to check the usefulness of robustness under different data generating processes;

- Most past works investigates the properties of robust estimation using one or two estimators at the same time. This work dealt with four robust and one non-robust estimator, allowing a more complete panorama of the performance and stability measures; and

- A sensitivity analysis was provided, extending the methods and procedures used in the course of this work to two important alternatives: the change of the type of efficient portfolio analyzed - MSR portfolios were replaced by Global Minimum Variance portfolios (GMV) - and the restriction to short selling activities. The purpose here was to "robustify" our conclusions and to enlarge the range of investors and policies covered by this work.

To accomplish the above-mentioned goals, we collected 15 daily stock prices negotiated either at New York Stock Exchange (NYSE) or NASDAQ Stock Market, from 2000/06/02 to 2015/07/01. This was the used sample for the estimation of robust and classical portfolios and their respective metrics of performance and stability. The results in this step lead us to formulate three research questions structured in the form of three hypotheses tests to be examined in the simulation exercise.

Then, ARMA-GARCH models were fitted to each time series and the semi-parametric bootstrapping procedure suggested in [30] was employed with the purpose to generate new (simulated) time series under the same dependence structure observed in the real data. These are sequences of multivariate returns that did not occur in the past, but plausibly could have occurred. 500 simulated past histories were used to generate sampling distributions for each statistic estimated using the observed data.

In general terms, considering the collected data, its subjacent multivariate stochastic process and the statistical behavior reproduced in the simulated time series, the conclusions obtained in this work are:

1. It is not possible to reject, at 5% significance level, the higher performance achieved by all robust portfolios in the buy-and-hold strategy, as measured by the Excess Sharpe Ratio, $\widehat{ES}_T$. However, a key result for this strategy surrounds the symmetrical shape of the $\widehat{ES}_T$ sampling distributions around small negative values. This implies that the unknown ES_T parameters locate themselves in regions where positive or negative values occur with fairly equal probabilities. In a different context, an adverse number of past returns for the same set of stocks could lead to negative excess Sharpe ratios for robust portfolios and we would not be able to reject them as well.

   In other worlds, results are not striking convincing favoring robust estimation in the buy-and-hold strategy and, as so, the usage of a more sophisticated estimation methodology, although robust to deviations of assumed models, might not be rewarding. Generally speaking, this strategy does not explore the potential of robust estimation to its full extent, as portfolio weights evaluation are performed only once, in a period of 8 years.

2. In the rebalancing strategy, we could not reject, at 5% significance level, that robust portfolios presented higher stability than the classical one, as measured by the Excess Total Norm Weights ($\widehat{ETNW}_T$). We verified in the various processes of estimation in the course of this work the perverse effect of data contamination in the inflation of the classical estimates. This ends up translating itself in higher portfolio weights instability and, consequently, higher transaction costs.

   It is clear that the explanation for the stability property resides in the high breakdown properties of robust estimators. Robust portfolios weight vectors do not greatly change at rebalancing days, because new transitory turbulent information is generally not taken into account, preserving the bulk of the data to the greatest extent possible. Only persistent outlying behavior, or else, permanent regime changes, has the potential to shift location and scatter estimates to completely new different values. The more stability presented by robust portfolios is undeniably the most important result achieved in this work and is compliant with the findings reported in [29] and [8], among others.

3. There is evidence suggesting that the higher the transaction costs incurred in a rebalancing strategy, the better the relative performance of MSR robust portfolios comparatively with the traditional mean-variance one. As we have already mentioned, the authors in [2] related high transaction costs to less efficient financial markets, possibly in less developed economies. In these cases, robust estimation might be an important tool to be used, possibly leading to positive excess Sharpe

ratios.

4. The above-mentioned stability results are robust both to the change of the analyzed portfolio (e.g. GMV portfolios) and to the inclusion of the restriction to short selling activities. In this regard, GMV portfolios excess transaction costs seem to be more predictable, as the real Excess Total Norm Weights parameter locates in a narrower region than those verified in the other policy investments. For MSR portfolios, restricting short selling activities could generate more predictable transaction costs. If predictability is not an important issue, than the leading case investigated in this work, MSR portfolios with short selling, can be an interesting alternative, as the respective sampling distributions are more stretched to the left, yielding more desirable possible regions for the $ETNW_T$ real parameter.

5. Based on the results achieved, we consider that the S robust portfolios (P_4) present the more interesting stability profile, regardless of the chosen policy investment.

5.1 Suggestions for Further Research

Although the sensitivity analysis performed in Chapter 4 expanded the applicability of this work to different investment policies alternatives, we believe that further research would provide valuable complements, mainly in the following unanswered questions:

1. What is the impact of different rebalancing frequencies on robust portfolios stability? Would it be true that more frequent rebalancing activities could lead to even more relative stability and transaction costs saving?

2. Given that we assumed a simplistic transaction cost model in this work, represented by a proportionality constant, what would be the effect of a more sophisticated transaction cost modeling on the relative performance of robust portfolios?

3. Given that high liquid stocks with long track record, issued by global mature companies, might be affected by similar risk factors and might present similar statistical behavior, could we weakly extend the results obtained in this work to other sets of similar stocks? Results would change with different dependence structure among assets?

4. If investors want to hold not only risky assets, but also the risk free asset in their portfolios, what would be the changes, if any, in the stability and performance results?

Bibliography

[1] Kaszuba Bartosz, *Applications of robust statistics in the portfolio theory*, Mathematical Economics (2012), no. 8 (15), 63–82.

[2] Valerie R Bencivenga, Bruce D Smith, and Ross M Starr, *Equity markets, transactions costs, and capital accumulation: an illustration*, The World Bank Economic Review **10** (1996), no. 2, 241–265.

[3] Fischer Black and Robert Litterman, *Global asset allocation with equities, bonds, and currencies*, Fixed Income Research **2** (1991), 15–28.

[4] Fischer Black and Robert B Litterman, *Asset allocation: Combining investor views with market equilibrium*, The Journal of Fixed Income **1** (1991), no. 2, 7–18.

[5] Christophe Croux and Gentiane Haesbroeck, *Influence function and efficiency of the minimum covariance determinant scatter matrix estimator*, Journal of Multivariate Analysis **71** (1999), no. 2, 161–190.

[6] Laurie Davies, *The asymptotics of rousseeuw's minimum volume ellipsoid estimator*, The Annals of Statistics (1992), 1828–1843.

[7] P Laurie Davies, *Asymptotic behaviour of s-estimates of multivariate location parameters and dispersion matrices*, The Annals of Statistics (1987), 1269–1292.

[8] Victor DeMiguel and Francisco J Nogales, *Portfolio selection with robust estimation*, Operations Research **57** (2009), no. 3, 560–577.

[9] David L Donoho, *Breakdown properties of multivariate location estimators*, Ph.D. thesis, Harvard University, 1982.

[10] David L Donoho and Miriam Gasko, *Breakdown properties of location estimates based on halfspace depth and projected outlyingness*, The Annals of Statistics (1992), 1803–1827.

[11] B. Efron, *Bootstrap methods: Another look at the jackknife*, The Annals of Statistics **7** (1979), no. 1, 1–26.

[12] Gabriel Frahm, Markus Junker, and Alexander Szimayer, *Elliptical copulas: Applicability and limitations*, Statistics & Probability Letters **63** (2003), no. 3, 275–286.

[13] Ursula Gather and Torsten Hilker, *A note on tyler's modification of the mad for the stahel-donoho estimator*, The Annals of Statistics (1997), 2024–2026.

[14] Frank R. Hampel, Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel, *Robust statistics - the approach based on influence functions*, Wiley, 1986.

[15] Guangliang He and Robert Litterman, *The intuition behind black-litterman model portfolios*, Available at SSRN 334304 (2002).

[16] C. Huang and R.H. Litzenberger, *Foundations for financial economics*, Prentice Hall, 1988.

[17] P. J. Huber, *Robust statistics*, Wiley New York, 1981.

[18] Mia Hubert and Michiel Debruyne, *Minimum covariance determinant*, Wiley interdisciplinary reviews: Computational statistics **2** (2010), no. 1, 36–43.

[19] J.E. Ingersoll, *Theory of financial decision making*, Rowman & Littlefield studies in financial economics, Rowman & Littlefield, 1987.

[20] Markus Junker and Angelika May, *Measurement of aggregate risk with copulas*, The Econometrics Journal **8** (2005), no. 3, 428–454.

[21] Jana Jureckova and Jan Picek, *Robust statistical methods with r*, Chapman - Hall/CRC, 2006.

[22] Hendrik P Lopuhaa, *On the relation between s-estimators and m-estimators of multivariate location and covariance*, The Annals of Statistics (1989), 1662–1683.

[23] Hendrik P Lopuhaa and Peter J Rousseeuw, *Breakdown points of affine equivariant estimators of multivariate location and covariance matrices*, The Annals of Statistics (1991), 229–248.

[24] Harry Markowitz, *Portfolio selection*, The journal of finance **7** (1952), no. 1, 77–91.

[25] R.A. Maronna, *Robust m-estimates of multivariate location and scatter*, The Annals of Statistics (1976), 51–67.

[26] R.A. Maronna, D.R. Martin, and V.J. Yohai, *Robust statistics: Theory and methods*, Wiley Series in Probability and Statistics, Wiley, 2006.

[27] Ricardo A Maronna and Victor J Yohai, *The behavior of the stahel-donoho robust multivariate estimator*, Journal of the American Statistical Association **90** (1995), no. 429, 330–341.

[28] Ricardo A Maronna and Ruben H Zamar, *Robust estimates of location and dispersion for high-dimensional datasets*, Technometrics **44** (2002), no. 4.

[29] Beatriz V. M. Mendes and Ricardo P. C. Leal, *Robust modeling of multivariate financial data*, (2003).

[30] ———, *Portfolio management with semi-parametric bootstrapping*, Journal of Risk Management in Financial Institutions **3** (2010), no. 2, 174–183.

[31] Beatriz V. M. Mendes and Daniel S. Marques, *Choosing an optimal investment strategy: The role of robust pair-copulas based portfolios*, Emerging Markets Review **13** (2012), no. 4, 449–464.

[32] A. Meucci, *Risk and asset allocation*, Springer Finance, Springer, 2007.

[33] Attilio Meucci, *Beyond black-litterman in practice: A five-step recipe to input views on non-normal markets*, Available at SSRN 872577 (2006).

[34] Cédric Perret-Gentil and Maria-Pia Victoria-Feser, *Robust mean-variance portfolio selection*, Available at SSRN 721509 (2005).

[35] Svetlozar T Rachev, Young Shin Kim, Michele L Bianchi, and Frank J Fabozzi, *Financial models with lévy processes and volatility clustering*, vol. 187, John Wiley & Sons, 2011.

[36] David M Rocke, *Robustness properties of s-estimators of multivariate location and shape in high dimension*, The Annals of statistics (1996), 1327–1345.

[37] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*, John Wiley & Sons, Inc., 1987.

[38] Peter Rousseeuw and Victor Yohai, *Robust regression by means of s-estimators*, Robust and nonlinear time series analysis, Springer, 1984, pp. 256–272.

[39] Peter J Rousseeuw, *Multivariate estimation with high breakdown point*, Mathematical statistics and applications **8** (1985), 283–297.

[40] ______, *A resampling design for computing high-breakdown regression*, Statistics & probability letters **18** (1993), no. 2, 125–128.

[41] Peter J Rousseeuw and Katrien Van Driessen, *A fast algorithm for the minimum covariance determinant estimator*, Technometrics **41** (1999), no. 3, 212–223.

[42] Peter J Rousseeuw and Bert C Van Zomeren, *Unmasking multivariate outliers and leverage points*, Journal of the American Statistical Association **85** (1990), no. 411, 633–639.

[43] David Ruppert, *Computing s-estimators for regression and multivariate location/dispersion*, Journal of Computational and Graphical Statistics **1** (1992), no. 3, 253–270.

[44] C.G. Small, *Expansions and asymptotics for statistics*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, CRC Press, 2010.

[45] W. A. Stahel, *Breakdown of covariance estimators*, Research Report 31, Fachgruppe fur Statistik, E.T.H. Zurich, Switzerland, 1981.

[46] David E Tyler, *Finite sample breakdown points of projection based multivariate location and scatter statistics*, The Annals of Statistics (1994), 1024–1044.

[47] Stefan Van Aelst, *Stahel–donoho estimation for high-dimensional data*, International Journal of Computer Mathematics (2014), no. ahead-of-print, 1–12.

[48] Stefan Van Aelst and Peter Rousseeuw, *Minimum volume ellipsoid*, Wiley Interdisciplinary Reviews: Computational Statistics **1** (2009), no. 1, 71–82.

[49] Jay Walters, *The Black-Litterman Model in Detail*, Social Science Research Network Working Paper Series (2009).

[50] Roy E Welsch and Xinfeng Zhou, *Application of robust statistics to asset allocation models*, REVSTAT–Statistical Journal **5** (2007), no. 1, 97–114.

[51] Rand R Wilcox, *Introduction to robust estimation and hypothesis testing*, Academic Press, 2012.

[52] Yijun Zuo, *Projection-based affine equivariant multivariate location estimators with the best possible finite sample breakdown point*, Statistica Sinica (2004), 1199–1208.