Robust Optimisation for Factor Portfolios

Candidate Number 593233
University of Oxford

A thesis submitted for the degree of

MSc in Mathematical Finance

April 10, 2016

Acknowledgements

I am thankful to my supervisor Prof. Raphael Hauser for his outstanding support, his patience throughout the project, and his precious insights into the field of robust optimisation.

I would also like to express my gratitude to Daniele Bortolotti for his invaluable advices and comments on my thesis.

Benoit Mondoloni, Lucien Gaier, Nicolas Joseph, Fadhel Ben Atig, Aurélien Lefèvre and Romy Shioda have always been ready to spare some of their precious time to answer my questions on quantitative finance, mathematics and statistics and I would like to thank them for that.

My friends Kiril Dimitrov and Besi Bezhani proved great critics of this present work and I am grateful for their reviews and feedbacks.

Last but not least, my family and friends have been a precious support and I wish to thank them for their encouragements.

Abstract

We integrate the robust optimisation framework into the construction of factor portfolios. For this purpose, we start by reviewing the concept of risk premium and we describe the salient statistical properties of factors. Further, we consider different possible formulations for an optimal factor portfolio and derive robust equivalent tractable problems. We detail interesting theoretical properties of these portfolios. Lastly, we test this framework on both simulated and market data.

High level results of these tests show that robustness leads to more stable, more diversified portfolios and potentially makes a better use of a turnover constraint.

Contents

List of Figures

List of Tables

Introduction

The recent years in the financial industry have seen a tremendous development of smart-beta strategies. Though this thesis does not claim to be smart, it will nonetheless focus on the portfolio construction underpinning such strategies!

Traditionally, the investment community has been split into two categories, there were the passive managers, believing in market efficiency, and the active managers, rejecting this hypothesis. The seminal work of E. Fama and K. French in 1992 [20] has opened the door to another approach to investment, which has later been called smart-beta, risk-premia, styles or factor investing. Somehow, the smart-beta investor sits between the active and the passive managers. The strategy is not passive as there is no intention to replicate a benchmark. It cannot be classified as active either, because once the strategy has been defined, investment is purely rule-based.

Smart-beta is much of a financial economics concept, and this work attempts to have a mathematically-coherent approach to the trading implementation of that concept. More precisely, focus is given to the portfolio-optimisation part.

A naive approach could be to estimate expected returns from smart betas directly. Then, these estimates could be simply fed into a Markowitz maximum-return problem.

As it would set aside the economics of smart-beta this approach would suffer from many weaknesses. Smart-beta expected returns are similar to the expected returns of other assets in a sense that they are very difficult to estimate. Michaud[37] calls the Markowitz optimisation *error maximization*, because it is highly sensitive to errors in inputs.

Though we remain in the Markowitz framework, special attention will be paid to consistency between errors in inputs and portfolio optimisation. To address the concerns that Markowitz portfolios may go very wrong should the realised parameters slightly differ from expectations, researchers have suggest the use of a technique originating in optimal control theory called robust optimisation.

Unlike previous attempts to make returns estimators less sensitive to estimation errors (e.g. with the use of James-Stein or Black-Litterman shrinkage estimators), robust optimisation is holistic as it directly takes into account errors in the optimisation process. Robust optimisation can be seen as an optimisation of the pessimal situation within a predefined set (the uncertainty set). We will discuss why it is an appropriate tool in our perspective.

In order to put some mathematical structure in the concept of smart-beta, we build on a factor model for returns. In this setting (originating in the Asset Pricing Theory of Ross [45]), returns are modelled as a linear combination of factors and the unexplained part is called alpha.

The modelling of smart-beta into factors also enables to gain a view on their stochastic properties. We will see that such factors have special features that need to be taken into account in the robust optimisation procedure. In particular, the economics of smart-beta mean that excess return gained from exposure to this beta is a necessary compensation for risk (risk being loosely defined as the possibility of losing a significant amount of the initial investment).

In this setting, it can be expected that the probability density function of the returns of a smart-beta type of strategy is far from normal. This should be taken into account into the formulation of a robust problem. Chen *et al.* [15] have introduced deviation measures that fit asymmetrical distributions while keeping a tractable formulation for the robust problem. The use of those measures, called forward and backward deviations, will be discussed in the context of smart-beta portfolios.

Robust optimisation is often discarded as it appears over-conservative. As it essentially protects against the worst-case situation, it should be expected that robust optimisation comes at a price. To be very stringent makes sense when designing the optimal structure of a bridge, but it may become an issue when constructing a return-generating strategy. Researchers have addressed this problem by forcing the uncertainty set to keep a special structure. We apply the Net-Zero Alpha Adjustment, a technique developed by Ceria and Stubbs [14] and describe how it avoids shrinkage of the returns estimator.

As mentioned previously, our primary focus is the optimal portfolio construction of a strategy, that builds-up on the concept of smart-beta. Though it should be beneficial over the long-term, smart-beta usually generates thin excess returns. Peculiar care should therefore be given to avoid that is is damaged by trading costs.

In fact trading costs have a paramount role in the success or failure of a smart-beta strategy. Consider for example a value strategy; it buys the stocks that are cheap (*vs.* their peers) and sells the stocks that are expensive (again, *vs.* their peers). To define what is cheap or expensive, a valuation metric is used, it is typically the Price-to-Book ratio or the Price-to-Earnings ratio. As soon as this strategy has some success, it will face turnover in the positions. Indeed, if things go well, the cheap company will see its price rising, it will become expensive and will need to be sold. The same dynamics apply to the short leg of the strategy. This strategy has hence a natural inclination to be affected by trading costs.

According to Renshaw[44], robust optimisation reduces turnover and enables to relax trading costs constraints. Such a reduction in turnover is easily conceivable, as robustness enables more smoothness in the optimal solution and it reduces the sensitivity to inputs. We will see that this hypothesis withstands empirical results only in some cases. Instead of unilaterally reducing turnover, robust optimisation rather seems to make a better use of a turnover budget.

Roadmap

The starting point of this thesis will be a focus on the concept of smart-beta. There is abundant literature on the subject, and we only introduce the cornerstones. We discuss in more detail the economic rationale, as it has important implications on the dynamic behaviour of smart-beta (in term of statistical properties). We also build up a factor model to better formalise the concept of smart beta.

In a second section, we examine the important aspects of a factor portfolio construction, be it with regard to the different possible formulations for the optimisation problem, or to the features it has to take into account. Statistical properties of smart-beta are then studied in detail, with a focus on what needs to be captured in the optimisation process.

The third section examines robust optimisation. First, we establish the necessity of going beyond standard optimisation and review some of the remedies to over-sensitivity that have been proposed by researches. Then we summarizes the main theoretical contributions that have been made to the literature on robust portfolio optimisation over the last 15 years. A few numerical examples are also developed to illustrate the fundamental differences between robust and non-robust optimisation.

The fourth section contains the mathematical substance in order to go from the concept of robustness to problems that are readily solvable by off-the-shelf solvers. All the problems that we focus on can be reformulated as Second-Order Cone Programs (SOCP). The SOCP formulation is derived for the robust mean-variance problem, the net-zero alpha adjustment problem, the robust deviation problem and the robust multi-period mean-variance problem.

Numerical experiments are conducted in a fifth section. Those are of utmost importance, primarily because there is no closed-form solution for the robust problems in most cases. It is thus difficult to gain an understanding of what robust optimisation exactly does. Data are simulated, in the perspective of both extracting the idiosyncratic features of robust optimisation and modelling possible paths for the relevant parameters (in particular the smart-beta exposures and returns). We are cautious about creating artificial bias that may distort conclusions and we will build a framework that enables a fair comparison between the different approaches.

We then review the results of those experiments and discuss their implication for the construction of smart-betas portfolios. A necessary cross-validation with external results from the literature is conveyed.

Concluding remarks summarize the contributions of this present work and introduce avenues for future research.

1 Risk premium and factor modelling

1.1 Introduction and literature review

The modern portfolio theory emerges in 1952 with Markowitz [35] as he derives a portfolio construction methodology coherent with investor's preferences. Expected return is not the sole item to consider, but it should be put in perspective with volatility and correlations. The optimal holding is a diversified portfolio and as all investors are rational, they share the same portfolio, the market portfolio.

Markowitz's work is extended with the introduction of the Capital Asset Pricing Model (CAPM) throughout the 1960 decade by Treynor [51], Shape [47], Lintner [34] and Mossin [39]. In this framework, there are two types of risk: a systematic risk, coming from the exposure to the market and an idiosyncratic risk, coming from companies specificities and that should be diversified away. At equilibrium, diversification has been achieved and the portfolio's expected return is solely driven by its market component, as measured by the beta.

The CAPM lays the foundations for factor modelling of equity returns, this is indeed a factor model with the market being the sole factor. A natural extension of the CAPM model appears in 1976 with the Arbitrage Pricing Theory (APT) of Ross [45]. He challenges the idea that excess returns would entirely be driven by the beta to the market factor and he instead proposes an equilibrium model with multiple factors. In fact, empirical research has shown that the main CAPM result do not hold (e.g. Fama and French in [22]). The market beta does not have sufficient explanatory power for returns. If a significant relationship were to be found, it would show negative correlation between the beta and returns (this is sometimes referred as the low volatility abnormally).

In his work, Ross does not try to put names on the different factors. The first attempt to identify those factors comes from Fama and French [20, 21], with their so-called Three-Factor Model. They show that more than 90% of a diversified portfolio's returns can be explained by three factors, namely the market factor, the size factor (SMB, Small Minus Big - referring to the company market capitalisation) and the value factor (HML, High Minus Low - referring to the book-to-price accounting measure). Importantly, the regressed coefficients of excess returns on both the SMB and HML factors are found to be positive. This means that small companies tend to outperform big companies and that companies with a high book-to-price multiple tend to outperform companies with a low book-to-price multiple.

1.2 Factor, risk premium and signal

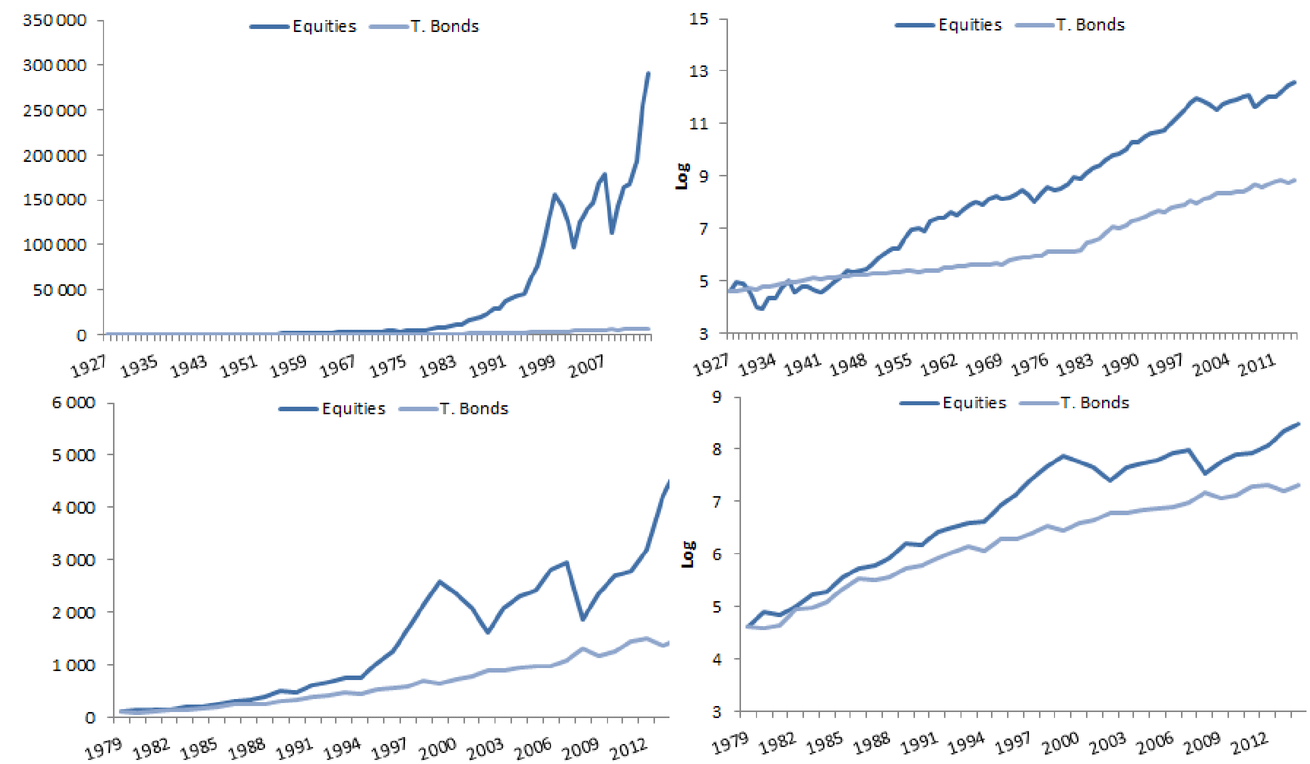
Factor
A factor can loosely be defined as "any characteristic relating a group of securities that is important in explaining their return and risk" (as per the index provider MSCI in 2013 [7]). Since the seminal work of Fama and French, there has been extensive research on identifying pertinent factors, both from the academic side and from the industry side. To qualify as such, a factor needs to have high explanatory power for excess returns, it should be observable over the long-term (though its behaviour may exhibit cyclicity) and it should be observable over a broad range of securities (ideally it should transcend asset-classes, countries, regions and sectors).

Factor modelling proves to be useful in a wide range of situations. For example, a investment manager may be interested in understanding what happens to her portfolio should the Federal Reserve decides to raise its deposit and lending rates. Alternatively, she might want to know which assets would be the most affected by a sudden increase in oil prices. Calculating the sensitivity of the portfolio to the independent variables oil or Fed rates (or to changes in those variables) can be considered as the premises of factor modelling. In this case, the factors would be macro-economic factors, and the perspective is to allocate ex-post returns or to manage ex-ante risk. This modelling does not aim at predicting the portfolio performance. If the portfolio manager want to generate excess return, she needs to either augment the model with a short-term view on the direction of the factor (e.g. the Fed will definitely raise rates in the next 3 months) or she need to be convinced that the factor over-performs over the long-term (e.g. oil-producing companies will appreciate because the transition from polluting energy sources to green energy sources will be slower than the increase in oil production).

Risk premium
The most consensual risk premium can be found by comparing the performance of a bonds portfolio vs. an equities portfolio over the long-term.

Figure 1: Cumulative performance of $100 in equities and bonds - Total return

The plots show the performance of $100 invested in equities (S&P500) and $100 invested in bond (US Treasury Bonds). On the top, the investments start in 1927, on the bottom, the investments start in 1980. The left plots show the cumulative performance, while the right plots show the logarithm of the cumulative performance. It is assumed that reinvestment (on the cash paid in the form of coupons or dividends) is done on a yearly basis.



In the above plot, it can clearly be seen that over the long-term, equities outperform bonds. The annualised difference between the two returns is called Equity-Risk Premium (ERP - there are multiple alternative methods to quantify this premium).

This empirical evidence is consistent with quadratic utility for investors. Put simply, for an equal amount of return, investors prefer lower risk. Equivalently, for a equal amount of risk, investors prefer higher return. Equities are riskier than bonds, hence they should yield a higher return. The long-term historical ERP has been found to be around 2.4% by Arnott and Bernstein [1]. This number should be taken cautiously, as it much depends on the chosen market and time span (Arnott and Bernstein analysed two centuries of data for the US market). It represents an *average* premium and may be significantly different for a shorter time-horizon (the ERP changes drastically throughout the business cycle and may well turn negative). It is however indisputable that the ERP has been positive at most times and for most markets.

Building on this introductory example, a risk premium can be defined as: any characteristics relating a group of securities making them risky and because of which they are bound to yield long-term excess return (on top of the risk-free asset). For example, the equity can to be considered as a contingent claim on a firm's value, and this particular place in the capital structure underpins the existence of the ERP.

Signal
Lastly, a signal may be considered as: any characteristics relating a group of securities that have strong predicting power for their future return. Importantly, this definition does not take into account any risk dimension. Finding good signals belongs to the world of alpha generation.

A simple example of a signal may be found in the literature on earnings revisions. This signal builds on two empirical evidences: earnings revisions are positively correlated with prices changes and earnings revisions typically exhibit autocorrelation. The natural conclusion is that one can go long stocks with positive earnings revisions and short stocks with negative earnings revisions. To the extent that efficient portfolio construction is achieved (and that this abnormality pertain in the future), one would gain excess return. This strategy is in no regard bound to bear excess risk (compared to a similar portfolio that does not focus on earnings revisions).

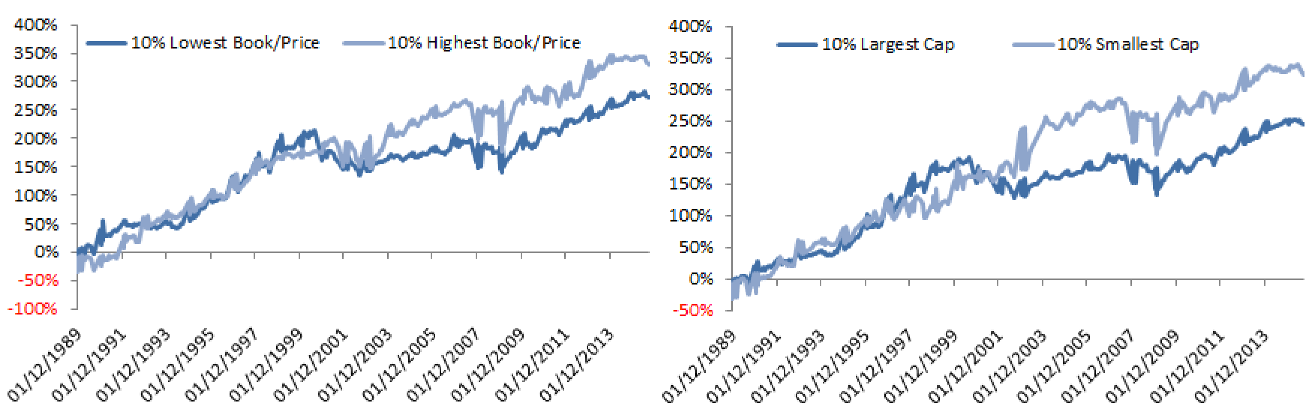
1.3 Economics of common equity risk premia

The work of Fama and French [20, 21] can be seen as an attempt to break down the ERP into subcomponents that would be more easily identifiable. In order to tie-up the Three-Factor Model with the concept of risk premium, a closer look at the SMB and the HML factors is required.

Stocks with small capitalisation tend to be less liquidly traded, it is therefore more difficult (and costly) for an investor to enter or exit a position on a small-cap name. Additionally, small caps tend to be less robust than large caps in an economic downturn scenario (for example because they have more difficult access to the capital markets or because they focus on very specialised products). Hence, they are more sensitive to the macro environment.

Similarly, value companies are generally highly levered and therefore very sensitive to tightening of financial conditions. Such companies are also common in declining industries, and they are the first to be hit by an economic downturn. Thus, the positive relationship between excess returns and SMB/HML indicates that those factors are in fact risk premia. A portfolio of small caps needs to outperform a portfolio of large caps over the long term because it is inherently more risky.

Figure 2: Fama-French value and size portfolios
Total return for the value and size portfolios from the Fama-French database [24].
Long-only, equally-weighted US portfolios.
Select the top and bottom 10% in term of Book/Price ratio (left), and market capitalisation (right).



Following the work of Fama and French, researchers have found numerous factors that can be considered as risk premia (this research has predominantly taken place in the equities world). Jegadeesh and Titman show the standard random walk assumption for log-prices does not hold. Past winners significantly outperform past losers, this is the momentum effect [30]. Another such anomaly is reported by Pastor and Stambaugh in 2003, less liquid stocks outperform more liquid stocks [42]. More recently, Fama and French introduced a five-factor model [23]. They augment their three-factor model with a profitability factor as well as a factor capturing the differences in investment styles.

The risk explanation as to why those factors keep outperforming over time is just one of the numerous approaches that authors have suggested. Other common explanations come from the behavioural finance side or from potential asymmetries in the markets (investment constraints affecting only some actors and not others).

Below are briefly described the most widely accepted equity risk premia. We do not try to be extensive (be it with regard to providing a full list, all the possible proxies or a complete survey about the justifications of their existence), but rather present them to give a clearer picture of what a risk premium may represent.

Table 1: Equity risk premia

Name	Side	Proxies	Reasons for existence
Size	Small-cap outperforms large-cap	• Market Cap	• Sensitivity to macro environment • Information uncertainty • Access to capital markets
Value	Cheap outperforms expensive	• B/P • P/E • EV/EBITDA	• Sensitivity to macro environment • Investors overreaction
Momentum	Positive momentum outperforms negative momentum	• Past returns (e.g. 12-month trailing return)	• Compensation for tail risk • Herding behaviour • Slow diffusion of information
Volatility	Low volatility outperforms high volatility	• Returns volatility • Market beta • Earnings volatility	• Structural constraints (e.g. leverage and short selling constraints)
Quality	High quality outperforms low quality	• FCF/GCI • ROE • Piotroski score	• No risk-based grounds • Higher total return than price return
Liquidity	Low liquidity outperforms high liquidity	• ADV • MDV • Free float	• Structural constraints (e.g. minimum liquidity constraint) • Trading costs • Sensitivity to macro environment

The below abbreviations have been used:

- B/P: Book to Price ratio
- P/E: Price to Earnings ratio
- ROE: Return On Equity
- ADV: Average Daily Volume
- MDV: Median Daily Volume
- FCF/GCI: Free Cash Flow to Gross Cash Invested
- EV/EBITDA: Enterprise Value to Earnings Before Interest, Taxes, Depreciation and Amortisation

1.4 Breaking-down the returns with a factor model

In this section, we introduce a factor model and comment on its relations with the construction of a a risk-premium portfolio. This factor model will not be directly used in our robust optimisation approach. However, it gives an understanding as to why we may focus on factor exposures: there is a linear mapping between factor exposures and expected returns, hence, with the assumption that a factor return is positive, maximising the expected return is equivalent to maximising the factor exposure.

Generic factor model

The starting equation is:

$$r = Bf + \epsilon \tag{1}$$

r is a vector of asset returns
B is a matrix of factor loadings (equivalently called factor exposures or factor betas)
f is a vector of factor returns
ϵ is a vector of residual (unexplained) returns

These definitions are voluntarily ambiguous, as there are important modelling choices that need to be made to better specify the equation.

There are a least three possibilities to specify equation (1):

1. Cross-sectional approach.
2. Time-series approach.
3. Principal-Components Analysis.

In our perspective, the natural choice is the cross-sectional approach. We now introduce what is often called a fundamental factor model (other standard factor models are statistical models and macro-economic models). A fundamental model focuses on each assets specific characteristics and most of the inputs can found by considering the asset itself (e.g. its country of incorporation, its financial statements, its sector, its historical price pattern, its daily traded volume...).

Fundamental factor model

For a fixed time t:
r is a vector of the returns of n different assets
B is a $n \times m$ matrix of loadings. A row corresponds to the different factor loadings of one asset.
f is a vector of m factor returns
ϵ is a vector of n assets residual returns

We also assume that $n \geq m$, i.e. there are more assets than factors.

In this setting, the inputs are r and B and the outputs are f and ϵ. It should be noted that this is not the approach that Fama and French took in their original model[20]. Instead, they input estimates for f by computing the returns of a relevant portfolio. For example, the SMB factor return is the difference between a small-caps portfolio vs. a large-caps portfolio returns. Further, the output of their regression is the matrix B. Fixing the loadings matrix B and estimating the factor returns f, is the current standard in the industry, and this is what the main factor model providers do (e.g. MSCI Barra [36] and Axioma).

The latter approach may seem counter-intuitive, but a by-product of regressed factor returns is their orthogonality. This proves very useful to do performance and risk attribution: *e.g.* it allows to allocate the total portfolio returns to distinct factors and simply sum the different contributions. On the other hand, estimating B directly is useful to understand the portfolio's response to a univariate shock: if the portfolio β to oil returns is 2, then for every 1%

increase in oil prices, the portfolio gains 2%. β has a much more limited use when it comes to allocate the ex-post returns to different factors.

Therefore, in a fundamental factor model, the matrix B is fixed and the factor returns are estimated. This does not mean however that the matrix B is set in stone. It is only set for the period over which the returns are estimated and may be updated between two estimations to take new information into account.

In order to estimate (1) via least-squares, residuals need to be homoscedastic. Taking the residuals as they are, this assumption would be violated. Small-cap stocks tend to have much higher residuals that large-cap stocks as they typically have higher idiosyncratic risk. Therefore, we need to weight the regression accordingly. Ideally, weighting should be done according to the inverse of the residuals, but the prior estimation of residuals is impractical (and unstable).

As the standard deviation of residuals decreases with the size of a company (with decreasing marginal decrease), practitioners usually take a concave function of capitalisation to get appropriate weighting. A logarithmic weighting or a square-root weighting could for example be used.

The heteroskedasticity-corrected version of equation (1) is then a weighted-least square problem :

$$\Pi^{\frac{1}{2}} r = \Pi^{\frac{1}{2}} B f + \Pi^{\frac{1}{2}} \epsilon \tag{2}$$

Π is a diagonal $n \times n$ matrix of weights such that (the weights do not need to add up to 1):

$$\Pi^{\frac{1}{2}} = \begin{bmatrix} \frac{1}{\log(a_1)} & & 0 \\ & ... & \\ 0 & & \frac{1}{\log(a_n)} \end{bmatrix} \quad \text{or} \quad \Pi^{\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{a_1}} & & 0 \\ & ... & \\ 0 & & \frac{1}{\sqrt{a_n}} \end{bmatrix}$$

a is a vector of size n, with component i being the market capitalisation for stock i.

The problem is then:

$$\min_{f \in \mathbb{R}^{\kappa}} \|\Pi^{\frac{1}{2}}(r - Bf)\| \tag{3}$$

The $\| \|$ norms refers to the Euclidean norm (this convention pertains in the remaining of the dissertation).

Assuming that the B matrix has full column rank, the problem is well defined (i.e. there is a unique solution $\bar{f}$ to problem (3)).
The First-Order Conditions enable to solve equation (3):

$$\bar{f} = (B^T \Pi B)^{-1} B^T \Pi r \tag{4}$$

Factors specifications

Unfortunately, we cannot assume that B has full column rank. In order for the factor model to be useful in the first place, the matrix B needs to contain most of the factors that have explanatory power for assets returns (keeping in mind that redundancies should be avoided).

Below are the common factors that have been found to have explanatory power:

- The market. All assets have unit exposure to this factor
- Countries or regions. Dummy variables can be used. For example, an asset listed and traded in Switzerland has an exposure of one to the Switzerland factor and an exposure of zero to any other countries.
- Industries or sectors. Dummy variables can be used.

- Styles factors (also called risk factors). The exposures need not to be discrete, but they should ideally be unitless and with comparable order of magnitude.

The concept of risk premium as discussed in section (1.3) can now be tied-up with factor modelling. What were previously called "risk premia" usually correspond to style factors in a factor-model context. Of course, style factors do not necessarily need to earn a risk premium over the long term. However, researchers have found that the the common risk premia like size, value or momentum link well with the market, countries and industries within a factor model to explain risk and returns.

This particular structure for B introduces collinearity. This issue arises because of the conjunction of unit exposure to the market factor, and dummy variables for industries and countries.

For example, take a situation where there are 8 assets, 1 market, 2 industries, 2 countries and 1 style.

Assume that the B matrix is:

$$B = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & -0.34 \\ 1 & 0 & 1 & 0 & 1 & 2.01 \\ 1 & 0 & 1 & 1 & 0 & -1.26 \\ 1 & 1 & 0 & 1 & 0 & -0.05 \\ 1 & 0 & 1 & 0 & 1 & 2.24 \\ 1 & 1 & 0 & 1 & 0 & 1.49 \\ 1 & 1 & 0 & 1 & 0 & 0.27 \\ 1 & 0 & 1 & 1 & 0 & -0.78 \end{bmatrix}$$

The first column corresponds to the market exposure, the two next are the industry exposures, the fourth and fifth are for countries and the last one is for the style exposure.

It can immediately be seen that the matrix is rank deficient (there are only 4 independent columns, because $C1 = C2 + C3 = C4 + C5$).

In order to have a well-defined regression output, we need to further specify the problem. In theory, any additional constraints, reducing the dimensionality by two can be chosen. Some constraints are more judicious than others however.

A reasonable choice is to force the market portfolio to have a net-zero return coming from both sectors and countries. The market portfolio is defined as the portfolio where each stock is held in proportion to its weight in the total market capitalisation. A portfolio manager will find country and sector contributions to her portfolio performance only if she takes *active* positions (*i.e.* different from the market).

This choice leads to the two constraints:

$$w_S^T f_S = 0 \quad \text{and} \quad w_C^T f_C = 0 \tag{5}$$

w_S and w_C are vectors for resp. the sectors and the countries market capitalisation weights.
f_S and f_C are sub-vectors from f gathering resp. the sectors and the countries returns.

As a result, the problem (3) becomes:

$$\begin{aligned} \hat{f} = \arg \min_{f \in \mathcal{R}^k} \quad & \| \quad \Pi^{\frac{1}{2}}(r - Bf)\|^2 \\ s.t \quad & w_S^T f_S = 0 \\ & w_C^T f_C = 0 \end{aligned} \tag{6}$$

A closed-form solution may be obtained for (6) using the QR decomposition. In annexes, section A.2, we derive the solution for this problem.

Risk Model

The risk model is:

$$\Sigma_r = B \Sigma_f B^T + \Delta \tag{7}$$

Σ_r is the $n \times n$ asset-level returns covariance matrix.
B is the $n \times m$ matrix of factor exposures as introduced in (1).
Σ_f is the $m \times m$ factor-level returns covariance matrix.
Δ is the $n \times n$ matrix of asset-specific variances.

In order to get equation (7) from (1), it should be assumed that the residual returns are uncorrelated (both with themselves and with the factor returns).

Ω can be estimated from the observations of factor returns:

$$\bar{\Sigma}_{f\,ij} = \frac{1}{T-1} \sum_{t=1}^{T} (\bar{f}_{i,t} - \mathrm{E}(f_i))((\bar{f}_{j,t} - \mathrm{E}(f_j))$$

2 Hurdles in efficient portfolios construction

In section 1, several risk premia have been identified. A question naturally appears: how should they be traded?

We start by reviewing different portfolio construction techniques, first from a univariate perspective (single-factor portfolio), then from a multivariate perspective (multi-factor portfolio). Further, we comment on the trade-off between the frequency at which a portfolio should be rebalanced and the associated costs. Lastly, the statistical properties of factors behaviours are investigated. In particular, we focus on alpha-decay, skewness of expected returns and the intertemporal behaviour of exposures for different factors.

2.1 Factor-mimicking portfolios and maximum-exposure portfolios

When introducing the concept of risk premium, we did not really specify their how they materialise. Indeed, no instrument named *SMB Factor* is listed on the New York Stock Exchange.
Here, we discuss the tradable implementation of such a factor.

Going back to the framework introduced previously to model returns, we saw that:

$$\Pi^{\frac{1}{2}}r = \Pi^{\frac{1}{2}}Bf + \Pi^{\frac{1}{2}}\epsilon$$

Define a portfolio as a vector $x = (x_1, x_2, ..., x_n)$ where x_i represents the weight invested in asset i.

2.1.1 Full replication

Though we only temporarily keep this assumption, it is worth looking at the factor-mimicking portfolio in the case where B has full column rank.
The portfolio $x^{j,FMP}$ mimicking factor j should respect m constraints:

$$B^T x^{j,FMP} = e_j \tag{8}$$

$e_j = (0, , ..., 0, 1, 0..., 0)$ is a size m vector of zeros, with 1 in position j only. The factor-mimicking portfolio has unit exposure to factor j and no exposure to any other factor.

We previous saw that the least-square estimator for the factor return $\bar{f}$ is:

$$\bar{f} = (B^T \Pi B)^{-1} B^T \Pi r$$

In this setting, getting a well-specified factor-mimicking portfolio is straightforward.
Indeed, we have:

$$B^T((B^T \Pi B)^{-1} B^T \Pi)^T e_j = B^T \Pi B (B^T \Pi B)^{-1} e_j = e_j$$

Hence

$$x^{j,FMP} = \Pi B (B^T \Pi B)^{-1} e_j$$

The factor-mimicking portfolio is simply the j^{th}-row of the $(B^T \Pi B)^{-1} B^T \Pi$ matrix. By construction, this portfolio has unit exposure to factor j and no exposure to other factors, it fully-replicates the performance of factor j.

Anecdotally, we note that the Gram-Schmidt decomposition could provide a robust alternative to avoid inverting $B^T \Pi B$. Indeed, when B is close to rank-deficiency, the inverse of $B^T \Pi B$ suffers from high numerical instability. The least-square estimation essentially requires to orthogonalise a set of factors. In this perspective, we could obtain $\bar{f}$ and $x^{j,FMP}$ by explicitly applying the Gram-Schmidt procedure to ΠB. Though equivalent with inverting $B^T \Pi B$, this alternative would benefit from better numerical properties.

2.1.2 Issues with full replication

The reverse-engineering stance that we took so far to find the factor mimicking portfolio could be extended to the case where B does not have full column rank.

However, such a factor-mimicking portfolio may be useless for practical considerations. Indeed, it does not allow for realistic constraints. This is a long-short portfolio, without bound on total weight or turnover constraint. It is entirely defined as an output of a regression at time t. The factor returns need to be estimated at a given frequency (daily, weekly, monthly) and the factor-mimicking portfolio has to be updated at the same frequency. It may completely change from one period to another.

The gross performance of this portfolio should equal the performance of the factor, however, the net performance (*i.e.* including trading costs, stamp duties, borrow fees...) may be significantly lower.

Hence, the full-replication portfolio is not a suitable option to trade a factor.

2.1.3 Minimum risk and maximum exposure

More realistic options to trade a factor consist in either:

- Minimizing the portfolio risk, subject to constraints on factors exposures.
- Maximizing the exposure to the factor that is tracked, subject to constraints on other exposures and risk.

The minimum-risk portfolio tracking factor $j \in \{1, ...m\}$, solves the problem:

$$\begin{array}{lrcll} \min\limits_{x} & & x^T \Sigma_r x & & \\ \text{s.t.} & x^T B e_j & \geq & 1 & \\ & |x^T B e_k| & \leq & \alpha_k & \forall k \neq j \\ & x & \in & \mathcal{F} & \end{array} \tag{P1}$$

The maximum-exposure problem for factor j is:

$$\begin{array}{lrcll} \max\limits_{x} & & x^T B e_j & & \\ \text{s.t.} & |x^T B e_k| & \leq & \alpha_k & \forall k \neq j \\ & x^T \Sigma_r x & \leq & \sigma^2_{max} & \\ & x & \in & \mathcal{F} & \end{array} \tag{P2}$$

$e_j = (0, , ..., 0, 1, 0..., 0)$ is a size m vector of zeros, with 1 in position j only.
$\alpha = [\alpha_1, ..., \alpha_{j-1}, \alpha_{j+1}, ..., \alpha_m]^T$ is a size $m-1$ vector of slack, $\alpha_k \ll 1$.
$\mathcal{F}$ represents additional constraints on the portfolio weights (e.g. long-only or turnover constraints).

Both those problems strip-out the regression part of factor-modelling and entirely focus on the loading matrix B. Hence, the factor returns are no longer required for the portfolio construction.

(P1) and (P2) take a different perspective to approach the factor portfolio construction.

(P1) is similar to the factor-mimicking portfolio problem, in a sense that unit exposure to factor j and minimal exposure to other factors are ensured. However, holding this portfolio no-longer guarantees a gain exactly equal to the factor return.

On the other hand (P2) is very similar to the maximum expected return problem. In fact, taking $\mu = B e_j$ and C obtained by removing column j to B, (P2) may be set as:

$$\begin{array}{lrcl} \max\limits_{x} & & x^T \mu & \\ \text{s.t.} & |C^T x| & \leq & \alpha \\ & x^T \Sigma_r x & \leq & \sigma^2_{max} \\ & x & \in & \mathcal{F} \end{array} \tag{P2'}$$

The portfolio obtained by solving (P2') does not aim at replicating factor j, it simply maximises exposure to the factor. We can however expect that the portfolio returns will be correlated with the factor returns.

This formulation also naturally ties up with the implementation of risk-premium portfolios.

For example, assume that the empirical outperformance of small caps over large caps as shown by Fama and French [20] remains true in the future. A first alternative may be to capture this risk-premium by constructing a portfolio that goes long small caps and short large cap.

In a more sophisticated approach, the formulation (P2') can be used. As a way of example, take:

$$\mu = [\mu^1, ..., \mu^n]^T \quad \text{with} \quad \mu^i = \frac{cap^i - \bar{cap}}{\sigma(cap)}$$

cap^i is a proxy for size of company i (e.g. market cap or log of market cap).
$\bar{cap}$ and $\sigma(cap)$ are the cross-sectional average and standard-deviation.
μ^i is then a z-score measuring the relative size of company i.

In this case, the vector μ is simply one of the columns of the matrix B.

The portfolio obtained by solving (P2') has the maximum exposure to the size factor.

In this setting, we no-longer work in the return space (in fact, due of the standardization process, μ is unitless). However, with the additional assumption that the expected return is a strictly increasing function of the exposure to the risk factor (size in our previous example), there is perfect equivalence between (P2) and Markowitz's maximum-return problem.

Obviously, the strict monotonicity assumption is a non-trivial one, but we have seen in section 1.3 that risk premia should have a such property (at least over the long-term). The more an investor accepts to be exposed to a source of risk, the more she should be compensated.

Hence, there is coherence between maximising returns and maximising exposures, the problem (P2') will be our base tool to construct risk-premia portfolios.

2.2 Top-down and bottom-up factor investing

One of the valuable properties of equity factors is their different behaviours throughout the business cycle (Bender *et al.* [5]). As a result, their returns correlation matrix can be exploited to enhance diversification.
This leads to multi-factor investing.

Following the discussions of the previous section, a natural approach to combine factors is to maximise the aggregate exposure:

$$\begin{array}{lrcl} \max\limits_{x} & & w^T B^T x & \\ \text{s.t.} & x^T \Sigma_r x & \leq & \sigma^2_{max} \\ & x & \in & \mathcal{F} \end{array} \tag{P3}$$

w is a size n weight vector that gathers the relative preferences between factors.

This is essentially a bottom-up approach where the sum of the asset-level exposures is maximised. Indeed, the decision vector x chooses for asset-level weights.

In a second approach, assume that the univariate factor portfolios have been constructed with expected factors returns $\mu = [\mu_1, ..., \mu_m]$, where μ_i is the expected return for factor i (slightly abusively, we call *factor return* what should really be called *return of a univariate factor portfolio*).

As discussed earlier, this main drawback of this approach is to require expected factor returns, notoriously difficult to estimate. The estimated returns should be considered cautiously. Also, getting a solid understanding of the dependency of the optimal solution to the input is critical.

The problem is then to choose a vector x of factor-level weights that maximises the total expected return:

$$\begin{array}{rlll} \max\limits_{x} & & x^T\mu & \\ \text{s.t.} & x & \in & \mathcal{G} \end{array} \tag{P4}$$

This is more of a top-down approach, which presumes the existence of tradable factors. It also requires knowledge of the expected factor returns (though this could be avoided with minimum-variance optimisation or risk-parity construction).

Both the bottom-up and the top-down approaches may benefit from robust optimisation, as both the factor exposures and the expected factor returns can be considered as uncertain. Uncertainty arises in the factor exposures and factor returns because they are stochastic and subject to estimation error (the former may be more important for factors exposures while the latter may be more important for factor returns).

As a result it seems appropriate to consider uncertain versions of the problems (P2), (P3) and (P4). Uncertainty arises in the objective functions for all three problems; on top that, there are uncertain constraints in problem (P2), via $|x^T Be_k| \leq \alpha_k$.

So as the discussions apply equivalently to the all the factor exposures and the factor returns problem, uncertainty is only modelled in the objective function (we do not include uncertain constraints in the robust formulations). This should not be considered as a problem as uncertainty in the objective is anyway treated as constraint in equivalent tractable formulations. Thus, for a given uncertainty structure, the complexity of the problem remain the same. Also, in section 3.4.3, we discuss a more general setting with uncertain constraints.

The discussions in sections 3 and 4 refer to both the bottom-up (factor exposures) and top-down (factor returns) portfolios, while in section 5 specific numerical experiments are run. The μ vector either refers to a column of the B matrix of factor exposures or to the expected factor returns.

2.3 Trade-off between turnover and alpha

A standard issue in the construction of a portfolio that aims at getting exposure to a factor, a risk premium or a signal is that it needs to be rebalanced at a certain periodicity. Rebalancing is required as the asset-level exposures changes with time. The rebalance process required trading, hence this is a costly process.

As a result, there is a trade-off between getting the maximum exposure and avoiding transaction cost. Some factors (like a short-term reversal factor) show impressive back-test returns, but the performance of a portfolio tracking such factors may be significantly lower because of elevated transaction costs.

Alpha decay is concept frequently used to approach this issue (Gârleanu and Pedersen [25]). Referring to a signal predicting returns, Gârleanu and Pedersen distinguish between the prediction strength (called *alpha*) and the decay of this prediction (*alpha decay*). A signal with strong alpha decay requires the portfolio to be rebalanced frequently, while a portfolio exposed to a slow-decaying signal benefits from positive returns for a long period of time. In the equity factors spectrum, the value factor is typically seen as a slow-decaying signal with low prediction power. Conversely, the momentum factor has a faster decay but higher prediction power.

Optimal rebalancing remains an open question with many possible approaches (e.g. periodic rebalancing with constraint, continuous trading, rebalance regions....). The periodic rebalancing is the setting that fits the best in the Markowitz framework, it has been well-studied and it is a standard in the industry.

In this case, the portfolio is rebalanced at fixed dates, and the holdings are obtained as the solution of a programming problem with constraints.

Those constraints may be:

1. Turnover constraint.
With y the vector of previous holdings, and TO_{max} the turnover-budget, the constraint is:

$$e^T|x-y| \leq TO_{max}.$$

2. Upper and lower bound on portfolio weights.
With x_{min} and x_{max} the vectors of bounds on portfolio weights:

$$x_{min} \leq x \leq x_{max}$$

3. Trading-costs constraint.
For example, with square-root market impact, take $\Lambda = \begin{bmatrix} d_1 & 0 & \\ 0 & ... & \\ & & d_n \end{bmatrix}$ and d_i quantifying the market-impact to trade asset i. TC_{max} is the trading-cost budget.

$$e^T \Lambda |x-y|^{\frac{3}{2}} \leq TC_{max}$$

We focus on turnover constraint as it has the advantage of not requiring the design of a transaction-cost model (unlike option 3.), while not being too problem-dependent (unlike option 2.)

Introducing such a constraint still comes at a price. It reduces the feasible set and it obviously lowers the optimal objective value.

In an extreme case, the μ vector of expected returns changes faster than what the optimisation can capture given the turnover-budget. The constraint may lock the portfolio in an undesirable position (if the expected return for an asset changes from a large positive value to a negative value from one rebalance to the next, the constraint may force the portfolio to keep positive exposure to that asset).

Such a constraint is nonetheless necessary for any real-world implementation of a factor portfolio. We will see that a key feature of robust optimisation is to make a better use of the turnover-budget.

2.4 Statistical properties of factors

Before further analysis, it should be noted that both the factor exposures and the expected factor returns are subject to stochasticity. The fundamental metrics driving the factor exposures are subject to periodic revisions (e.g. a company's quarterly results enable to update the return-on-equity ratio used in a quality factor. Quarterly results may not be predicted with absolute certainty). Also, as discussed in section 1.3, the expected factor returns change throughout the business cycle, itself being difficult to forecast.

2.4.1 Factor returns

Factor returns, or more specifically risk premia factor returns should be expected to deviate from the Gaussian distribution, generally seen as a decent approximation for equity (log) returns. Indeed, as they represent compensation for assuming certain risks, the returns are likely to be skewed and with excess kurtosis. Schematically, if a momentum factor has excess return over the long-term, this is because once in a while, it may suffer from severe drawdown.

In a first approximation, risk could be identified as standard deviation of returns (therefore well-explained by the Markowitz model). However, empirical research has shown that the second-moment of factor returns may not explain the excess compensation. A recent piece of research by Lempérière *et. al.* [33] has shed light on the particular statistical features of risk premia. They argue that such premia exhibit tail-risk skewness, but no significant excess volatility. This is observed not only in equities risk premia, but across most asset-classes.

Lempérière *et. al.* also design a skewness measure based on the ranked returns that well captures asymmetry in the risk premia performance. Risk premia consistently have negative skewness and and there is a linear relationship between this skewness and excess returns. It is also argued that other measures of skewness would lead to the same conclusions.

In order to gain some insight about the statistical properties of factors, we get returns from the Fama and French database [24].

More precisely, we retrieve the daily returns for the US market for 5 equity risk premia since 1990:
- a size factor (Small Minus Big - SMB)
- a value factor (High Minus Low - HML)
- a momentum factor (Up Minus Down - UMD)
- a quality factor (Robust Minus Weak - RMW)
- an investment-style factor (Conservative Minus Aggressive - CMA)

Table 2: Fama-French factor returns 1990-2015
μ annualised average return, σ annualised standard deviation, γ Pearson's skewness
p forward deviation, q backward deviation, r rank-one autocorrelation

	SMB	HML	RMW	CMA	UMD
μ	0.017	0.025	0.045	0.028	0.076
σ	0.091	0.095	0.068	0.066	0.137
γ	-0.165	0.126	0.349	-0.521	-0.946
p	0.011	0.010	0.010	0.006	0.017
q	0.010	0.012	0.007	0.014	0.020
r	0.027	0.088	0.144	0.106	0.180

Figure 3: Fama-French cumulative factor returns 1990-2015

As expected, all the factors have positive cumulative performance over the 25 years of data. The momentum factor has the strongest average return $\bar{\mu}$, but it also shows a drastic drawdown in 2009. The statistical measures in table 2 are only partially consistent with the results of Lempérière *et. al.* [33]. Pearson's skew is negative for only three of the five factors (SMB, CMA and UMD). However, alternative measures of asymmetry show negative skewness. In particular, the backward deviation of returns q, formally defined in definition 4.2 and further discussed in section 4.4, is higher than the standard deviation for all factors.

Lastly, an important property of factor returns can be observed with the autocorrelation $\bar{r}$. Autocorrelation is significantly positive for all the factors (as confirmed with a Ljung-Box test). This indicates that the factor are more stable than the underlying assets (that have autocorrelation non-significantly different from 0 [17]), and supports the use of factor models.

2.4.2 Factor exposures

Unlike factor returns, which have a common economic rationale, there is no obvious reasons as to why the exposures should share the same behaviour. It may nonetheless be seen that exposures, considered in *level,* should be close to random-walks (i.e. they should show strong and slowly-decaying autocorrelation). This is crucial to support the construction of maximum-exposure factor portfolios. Indeed, if the one-day autocorrelation was null, with a monthly rebalancing for example, maximising exposure would be meaningless. The optimised portfolio exposures would be random for most of the month. This is not the case however and the exposures typically move progressively.

For example, consider a momentum factor, defined for time T as:

$$mom_T^i = \sum_{j=0}^{l-1} r_{T-j}^i$$

mom^i is the momentum for company $i \in \{1, ..., n\}$
l is the look-back window.
r^i is the log-return for company i.

Assume that the log-returns are independent and normally distributed $\mathcal{N}(\mathrm{E}(r^i), \sigma^2(r^i))$. The portfolio is rebalanced at a τ-day frequency:
Then:

$$\sigma(mom_{T+\tau}^i, mom_T^i) = \begin{cases} 0 & \text{if } \tau > l \\ (l-\tau)\sigma^2(r^i) & \text{if } \tau \leq l \end{cases}$$

Hence:

$$\rho(mom_{T+\tau}^i, mom_T^i) = \begin{cases} 0 & \text{if } \tau > l \\ \frac{l-\tau}{l} & \text{if } \tau \leq l \end{cases}$$

There are two limiting cases. On the fist hand, if the portfolio is frequently rebalanced, with a long look-back window, the exposure changes very slowly between two rebalances and the required turnover to keep good exposure to the signal is low. On the other hand, if $\tau > l$, the exposure has no autocorrelation and is may be completely different from one rebalance to another.

A medium-term momentum could take $l = 252$, $\tau = 22$, *i.e.* the factor is the annualised return and the portfolio is rebalanced monthly. This yields an autocorrelation of 91% from one rebalance to the next.

This example may be simplistic, but it is quite representational of the typical sticky behaviour of factor exposures. Such behaviour could be also observed for most factors built on fundamental metrics (in which case averages are frequently used to smoothen the time series - *e.g.* a quality factor with the average return-on-equity over the last 5 fiscal exercises).

Also, most often, the exposures are winsorized and standardised (within the n-asset population and for a fixed time). The process aims at reducing sensitivity to outliers (much required if there is least-square estimation later) and it aims at providing unit-less values for effective comparison between factors. This may be a strongly non-linear transformation of the dataset, the final values remain nonetheless strongly autocorrelated.

3 Beyond standard optimisation

Though we only focus on (M1), there are three equivalent formulations of the Markowitz mean-variance problems and they are introduced below:

- Maximum-return problem:

$$\begin{array}{llll} \max\limits_{x} & & \mu^T x & \\ \text{s.t.} & x^T \Sigma_r x & \leq & \sigma_{max}^2 \\ & x & \in & \mathcal{F} \end{array} \tag{M1}$$

- Minimum-risk problem:

$$\begin{array}{llll} \min\limits_{x} & & x^T \Sigma_r x & \\ \text{s.t.} & \mu^T x & \geq & r_{min} \\ & x & \in & \mathcal{F} \end{array} \tag{M2}$$

- Risk-aversion problem:

$$\begin{array}{llll} \max\limits_{x} & & \mu^T x - \frac{\lambda}{2} x^T \Sigma_r x & \\ \text{s.t.} & x & \in & \mathcal{F} \end{array} \tag{M3}$$

σ_{max} is the risk-budget parameter.
r_{min} is the minimum expected return.
λ is the risk-aversion parameter.

These problems are equivalent in a sense that they lead to the same efficient frontier. More precisely, for a fixed choice of σ_{max} in (M1), there exists values r_{min} and λ such that (M1), (M2) and (M3) have the same solution. The same applies if one decides to fix r_{min} or λ.

For example, with $\mathcal{F} = \mathcal{R}^n$ and assuming that the solution for (M1) consumes the risk-budget (*i.e.* $x^T \Sigma_r x = \sigma_{max}^2$), we set:

$$r_{min} = \frac{\sigma_{max}}{\sqrt{\mu^T \Sigma_r \mu}}$$

$$\lambda = \frac{\sqrt{\mu^T \Sigma_r \mu}}{\sigma_{max}}$$

With the above parameters, the solutions of (M1), (M2) and (M3) reach the same point $(\, x^T \Sigma_r x \,;\, \mu^T x\,)$ on the efficient frontier.

There exists a forth problem similar to the previous formulations, namely the Sharpe ratio maximisation problem.

- Sharpe ratio maximisation problem:

$$\begin{array}{llll} \max\limits_{x} & & \frac{\mu^T x - r_f}{\sqrt{x^T \Sigma_r x}} & \\ \text{s.t.} & x & \in & \mathcal{F} \end{array} \tag{M4}$$

r_f is the risk-free rate.

This problem is non-convex, it can be however be reformulated as a convex problem using homogeneity.

The focus is given to problem (M1) in the rest of the dissertation.

3.1 Flaws in the Markovitz framework

Though it is one of the most widely used model in the investment management community, the Markowitz model has been criticised for a variety of reasons. In this section, we review the critics that are most relevant to our factor-portfolio construction perspective.

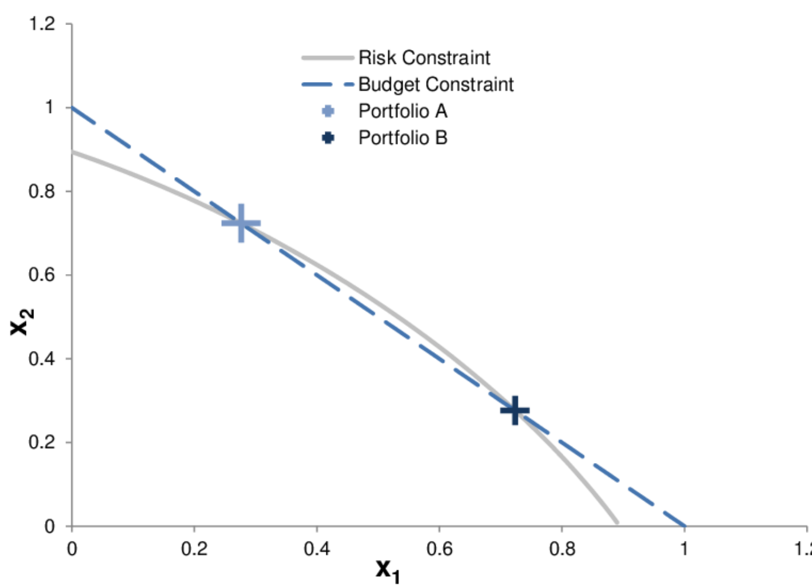
As an introductory example, consider a situation where:

- one solves the problem (M1), with two assets x_1 and x_2.
- the risk model Σ_r is such that $\Sigma_r = \begin{bmatrix} 0.05 & 0.025 \\ 0.025 & 0.05 \end{bmatrix}$
- the risk budget σ_{max} is given by $\sigma_{max}^2 = 0.04$
- there are budget and long-only constraints, i.e. $\mathcal{F} = \{x \in \mathcal{R}^2,\ \ x_1, x_2 \geq 0,\ \ x_1 + x_2 = 1\}$.

In a first configuration (A), $\mu = \begin{bmatrix} 0.09 \\ 0.11 \end{bmatrix}$, whereas in a second configuration (B), $\mu = \begin{bmatrix} 0.11 \\ 0.09 \end{bmatrix}$.

The below plot shows the optimal asset-mix in the two configurations.

Figure 4: Optimal mean-variance two-asset portfolios



The optimised portfolios in configuration (A) and (B) are respectively $x^A = \begin{bmatrix} 0.28 \\ 0.72 \end{bmatrix}$ and $x^B = \begin{bmatrix} 0.72 \\ 0.28 \end{bmatrix}$. With a slight change in μ, the optimal portfolio is completely different.

In fact, as long as $\mu_1 < \mu_2$, the optimal portfolio is x^A and it changes to x^B when $\mu_1 > \mu_2$. Indeed, what matters here is the relative position of μ_1 *vs.* μ_2, their absolute level does not really matter.

Although this example is much simplified, extreme sensitivity is a typical feature of standard Markowitz optimisation. Sensitivity would not be a problem *per se* if there was no trading cost. Similarly, it would not be a problem if the input parameters (μ and Σ_r in the previous example) were set once and for all.

However, none of these conditions holds in our perspective. As discussed before, trading costs are a key issue when building factor portfolios. One could not afford to turn over $|0.72 - 0.28| + |0.28 - 0.72| = 88\%$ of the portfolio should the estimate for μ changes from (A) to (B).

μ and Σ_r cannot be set once and for all as those parameters vary over time and they suffer from estimation error.

3.2 Robust returns estimators, portfolio resampling and stochastic optimisation

3.2.1 Robust estimators

A first approach to reduce the overall sensitivity is a robust statistics approach that starts by building *better* estimators for μ and Σ_r. *Better* may be understood in at least two senses: reduced sensitivity to outliers and lower mean squared error (MSE). To reduce sensitivity to outliers, standard statistical tools may be used, *e.g.* replacing means by medians, winsorization... Also, shrinkage enables to cut down the estimation error. In particular, for estimation of the mean of a Gaussian vector, the James-Stein estimator dominates the standard sample mean estimator in a sense that it has lower MSE.

Tailored estimates for portfolio management have also been designed. For example, the Black-Litterman combines an investor's view with the market equilibrium to improves the expected return estimates [13]. This is a Bayesian model where the prior distribution comes from the market equilibrium. The investor obtains a full probability distribution for the return by augmenting the prior with her own forecast. The resulting estimate can further be used in the optimisation procedure, enabling to take into account the confidence that the investor puts in her view.

3.2.2 Portfolio resampling

A heuristic developed by Michaud [38] specifically aims at stabilising the optimised portfolio with respect to changes in the inputs (expected returns and risk model). Put simply, the inputs are resampled and for each sample, the optimisation problem is solved. Further, the optimal solution is the average of all the sample solutions. The only additional input required in this case is the distribution of the estimation error (or at least partial information on the distribution). This is much of a Monte-Carlo approach. Obviously, the resampled efficient frontier is below the frontier obtained with the nominal inputs, but the realised portfolio performance should improve. Indeed, the optimal portfolio is more stable and its composition should be closer to the optimal composition (given realised parameters).

However, this technique suffers from severe drawbacks. First it has the typical Monte-Carlo feature of slow converge and requires extensive numerical simulation (simulating the range of possible input errors for a 1000-asset portfolio is not a trivial task). Also, the averaging process for the resampled optimal solution does not guarantee feasibility in the case of non-convex constraints.

Robust optimisation, discussed in more detail in section 3.3, may somehow be seen as an extension of Michaud's resampling. Instead of the average, focus is on the worst case, but the idea of varying the input around a nominal estimate and running optimisation for each situation is the same. The direct advantage of robust optimisation over resampling is that the robust optimisation uncertainty set needs not to be discrete (while the sample set has a finite number of estimates by nature).

3.2.3 Stochastic optimisation

As discussed previously, the input parameters in portfolio optimisation are affected by estimation error and they may exhibit a stochastic behaviour. A natural reaction following this observation is to model them as random variables and assign them a probability distribution. To get a well-defined problem, optimisation should be performed over an *expected* value and the constraints become *chance* constraints.

A stochastic version of problem (M1) is:

$$\begin{array}{llcl} \max\limits_{x} & & \mathrm{E}(\mu^T x) & \\ \text{s.t.} & \Pr(x^T \Sigma_r x \leq \sigma^2_{max}) & \geq & \rho \\ & x & \in & \mathcal{F} \end{array} \tag{S1}$$

In the above, the expectation and the chance constraint have different underlying probability sets.

The problem (S1) and other stochastic formulations are known to be very difficult to solve in the general case.

They require further assumptions to be simplified or approximated. A very prominent issue for such formulations in the case of factor portfolios is the absence of equivalent tractable formulations. Solving a stochastic program usually requires to generate scenarios and to solve a sub-problem for each scenario. This suffers from the curse of dimensionality: the number of scenarios required to obtained a fixed precision explodes when increasing the number of assets (or the number of time periods in the multi-period case).

As mentioned previously, tracking a risk premium requires a diversified portfolio, itself necessitating investment in many assets. Stochastic optimisation may be the ideal approach, but is computationally burdensome and impractical for high-dimensional problems.

Bertsimas *et al.* [11] have shown examples where most of the benefits of stochastic optimisation may be captured by robust optimisation, with drastic decrease in the computational effort. Robust optimisation is further investigated in the next section.

3.3 Robust portfolio optimisation

3.3.1 Literature review

The field of robust optimisation has grown tremendously over the past two decades. In this section, we quickly review the seminal research pieces and further focus on the portfolio-management applications.

Robust optimisation appears at the end of the 1990s decade. It arises both from the robust control literature (El Ghaoui and Lebret [19]), and as a separate framework, mainly designed by Ben-Tal and Nemirovski [3, 4, 5]. The main idea is to allow for uncertainty in the parameters of a convex optimisation problem. Uncertainty is associated with hard constraints (that have to be satisfied, for any realisation of the uncertain parameters). Ben-Tal and Nemirovski show that in the case of ellipsoidal uncertainty sets, the generic convex optimisation problems (*e.g.* linear and quadratic programming) have a robust equivalent with the same level of complexity. As such, they are tractable and can be solved efficiently with interior point algorithms.

The robust optimisation technique is thereafter used in the portfolio-construction context by Goldfarb and Iyengar [27]. They develop a linear factor model for the asset returns, allow for uncertainty in the different parts of the factor model and solve the equivalent Markowitz problems. They show that uncertainty coming from statistical estimation of the factor model takes the form of either hyper-rectangles or ellipsoids. Hence, the problem can be reformulated in the form of Second-Order Cone Programs (SOCP).

It is worth noting that the factor structure chosen by Goldfarb and Iyengar is different to ours. They take the factor returns as exogenous, while we consider them as endogenous (section 1.4). Further, they estimate the factor loading matrix (B in section 1.4) with linear regression. In our case, B is directly observable from market data. It may also be subject to error, but this is not regression noise.

Unlike Goldfarb and Iyengar, Ceria and Stubbs [14] think that uncertainty matters mainly for the expected returns. Or equivalently, that the estimation of expected returns is more error-prone than the estimation of variances and covariances. Hence, they separate the estimated covariance matrix of asset returns and the covariance matrix of estimation errors in expected returns. They show that with an ellipsoidal uncertainty set for the error in expected returns, the robust equivalent of the Markowitz problem simply adds a correction term. This term corresponds to a penalty for estimation error (the assets with large estimation error are more penalised than the assets for which one is sure about the future returns). Noting that the robust optimisation tends to be too conservative they also propose an adjustment to address this issue (the Net-Zero Alpha Adjustment). This adjustment is reviewed in more detail in section 4.3.

This present work is essentially in the setting of Ceria and Stubbs, where uncertainly is only considered at the expected returns level.

Stubbs and Vance [50] give guidance on the estimation of the parameters that define ellipsoidal uncertainty sets. More precisely, a scalar and an error estimation matrix need to be computed. The former controls the desired level of robustness while the latter models error in the expected returns construction process. The authors advise use of a diagonal matrix for the error estimation matrix. They also show that an exact estimation error matrix can be derived in several cases (simple historical mean estimate, regression-based estimate and Bayesian estimate).

The next set of important applications of robust optimisation in a portfolio-management context comes from the robust Value-at-Risk (VaR) literature. The main idea is that properly defined uncertainty sets enable to construct coherent risk measures (in the sense of Artzner *et al.* [2]*)*. The link between coherent risk measures and robust optimisation is detailed in Natarajan *et al.* [40, 41] and Bertsimas and Brown [10]. More relevant to our perspective, Natarajan *et al.* [40] propose the use of tailored deviation measures to optimise worst-case VaR when the distribution of uncertain parameters is asymmetric. The deviation measures, called forward and backward deviations, were first introduced by Chen *et al.* [15]. As such asymmetric deviation measures are utmost relevant to the construction of factor portfolios, they will be discussed further in section 4.4.

Another way to avoid getting pessimal solutions from robust optimisation problems is suggested by Hauser *et al.* [29]. They elaborate on the relative robust optimisation methodology as first introduced by Kouvelis and Yu [32] . *Relatively robust* should be understood as *relatively robust compared to the solution with hindsight on the uncertain parameters*. The cost for this more sophisticated approach is the increased complexity in the program. Indeed, it is a 3-layer problem (while the typical robust optimisation framework only has 2 layers). However, the authors exhibit tractable approximations for a large class of mean-variance portfolio optimisation problems. In particular, with ellipsoidal uncertainty in the expected returns and a polyhedral feasible set, a conic inner approximation is derived. This formulation is proven tight in the cases with zero or one inequality constraint.

A recent trend in the robust optimisation research has focused on modelling uncertainty sets directly from data (for example Bertsimas *et al.* [12] and Bertsimas and Thiele [8]). Unlike traditional methods that build on heuristics to shape the uncertainty set, the authors propose an integrated approach that focuses on data for the construction. This approach relies on mild assumptions about the distribution of uncertain parameters (*e.g.* independent marginals), the data and hypothesis testing. The resulting uncertainty set should offer a probabilistic guarantee (of constraint non-violation) and it should be computationally tractable. One of the aims is also that the set is not overly conservative. The authors derive sets with those features for numerous tests (e.g. χ^2 and Kolmogorov-Smirnov tests).

3.3.2 Uncertainty set

Before further discussions on robust problems, we review the concept of *uncertainty sets.*

Consider a generic robust optimisation program, with uncertainty in the objective only:

$$\max_{x\in\mathcal{F}} \min_{\mu\in\mathcal{U}} \quad f(x,\mu) \tag{R0}$$

For x decision variable and μ uncertain input parameter, the set $\mathcal{U}$ is the *uncertainty set.* Put simply, the uncertainty set is the set of possible values for μ. The formulation (R0) ensures that the objective function has the maximum value for the worst-possible realisation of μ within $\mathcal{U}$.

Appropriate modelling of the uncertainty is critical to obtain a tractable formulation. It should take into account the supposed distribution of the uncertain parameters (which guides the uncertainty *shape*) as well as the desired level of robustness (which guides the *size* of the uncertainty set*)*.

Standard uncertainty sets are:

1. $\mathcal{U} = \{\mu_1, ..., \mu_k\}$ - a discrete set of possible values.
2. $\mathcal{U} = \text{conv}\{\mu_1, ..., \mu_k\}$ - the convex hull of a discrete set.
3. $\mathcal{U} = \{\underline{\mu} \leq \mu \leq \overline{\mu}\}$ - a hyper-rectangle.
4. $\mathcal{U} = \{\mu \ : \ A\mu \leq b\}$ - a polyhedra.
5. $\mathcal{U} = \{\mu \ : \ \mu = \mu_0 + v \ : \ v^T W v \leq 1\}$ - an ellipsoid.

Apart for the first one, all these sets are convex - which is of great help to obtain tractable formulations.

In the remaining of the dissertation, we use variants of the ellipsoidal uncertainty set. To support the use of such sets, we first take a closer look at data and see how uncertain parameters may fit in an ellipsoid.

Assuming $\mu \sim \mathcal{N}(\mu_0, \Sigma_\mu)$, a standard result in statistics (*e.g.* [26]) states that we have the ρ-level confidence region:

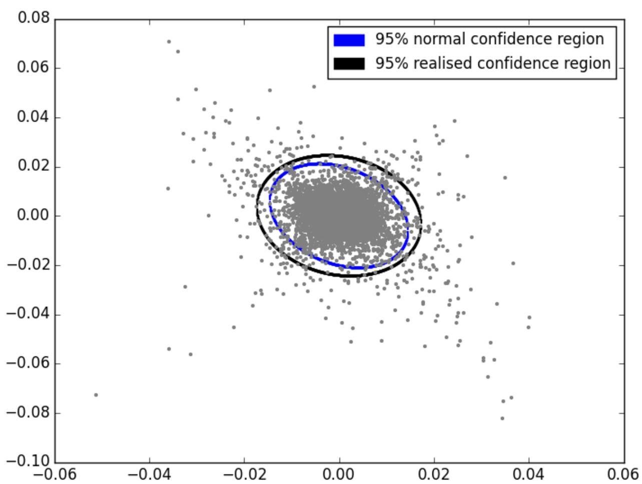
$$\Pr(\, \mu \in \{ \; \mu + v \; : \; v^T \Sigma_\mu^{-1} v \leq \chi_n^2(\rho) \; \} \,) = \rho \tag{9}$$

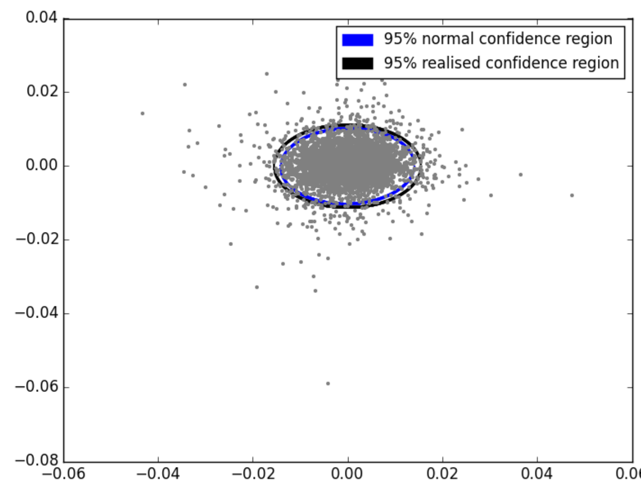
χ_n^2 is the inverse cumulative distribution function of the chi-squared distribution with n degrees of freedom.

In section 2.4, it was shown that the normality assumption for factor returns was likely to be violated. To test the ellipsoidal goodness of fit for factor returns, we:

- Draw scatter plots for pairs of factor returns as obtained in section 2.4.
- Fit a normal uncertainty set at 95% confidence using the empirical covariance and equation (9).
- Fit an ellipsoidal uncertainty set that contains 95% of the sample points.

Figure 5: Uncertainty in factor returns
Ellipsoidal confidence regions for the HML and UMD factors (left).
Ellipsoidal confidence regions for the SMB and CMA factors (right).





The first observation is that the factor returns are concentrated around a nominal value - 0 - and that the distributions seem closer to normal than to *e.g.* uniform. In both cases however , the 95% realised confidence region is larger than what could be predicted by the normal distribution, highlighting that factor returns have fat tails. This is especially true for the momentum (HML) and value (UMD) factors, and slightly less observable for the size (SMB) and investment-style (CMA) factors.

As only some two-dimensional projections of the μ vector are considered, the experiment may seem partial. However, it shows that the ellipsoidal uncertainty set remains a valid first-order approximation for the possible values of μ. Even if the factor returns distribution deviates from the Gaussian distribution, one can adjust the desired level of robustness by augmenting the size of the ellipsoid (one may inflate the value of $\chi_n^2(\rho)$ to account for fat tails). Regarding potential adjustments to the shape of uncertainty, we will consider a generalisation of the ellipsoidal uncertainty set that captures asymmetry in section 4.4.

3.3.3 The robust risk budget problem

In this section, a tractable formulation for the robust equivalent of problem (M1) is derived. We focus on the case with ellipsoidal uncertainty set for the expected return. The risk model Σ_r is assumed to be fixed.

The robust equivalent of problem (M1) is:

$$\begin{array}{lccc} \max\limits_{x} & \min\limits_{\mu\in\mathcal{U}} & \mu^T x & \\ \text{s.t.} & x^T\Sigma_r x & \leq & \sigma^2_{max} \\ & x & \in & \mathcal{F} \end{array} \tag{R1}$$

$\mathcal{U}$ is the uncertainty set. In this section, it is assumed that $\mathcal{U} = \{\mu = \mu_0 + \nu \ : \ \nu^T W \nu \leq 1\}$.
μ_0 is the the centre of the ellipsoid, it may be chosen as expected value for μ.
$W \succeq 0$ defines the size of the ellipsoid. It enables to adjust the confidence that should be put in μ estimates.

The inner part of problem (R1) can be solved on a stand-alone basis.

For a fixed $x \in \mathcal{F}$, we look for:

$$\begin{array}{lccc} \min\limits_{\nu} & & \nu^T x & \\ \text{s.t.} & \nu^T W \nu & \leq & 1 \end{array} \tag{10}$$

The pessimal value for v may be obtained with the Karush-Kuhn-Tucker (KKT) conditions (which are detailed in the annexes):

$$\nu = \frac{-W^{-1}x}{\sqrt{x^T W^{-1} x}} \tag{11}$$

Hence, the pessimal value for μ is:

$$\mu = \mu_0 - \frac{W^{-1}x}{\sqrt{x^T W^{-1} x}} \tag{12}$$

The problem (R1) becomes:

$$\begin{array}{lccc} \max\limits_{x} & \mu_0^T x & - & \sqrt{x^T W^{-1} x} \\ \text{s.t.} & x^T\Sigma_r x & \leq & \sigma^2_{max} \\ & x & \in & \mathcal{F} \end{array} \tag{R1'}$$

Though it is not readily solvable, the formulation (R1') highlights the point made by Ceria and Stubbs: in its simplest form, the robust equivalent simply penalises the objective function for error in expected returns. The penalty is a decreasing function of the confidence that is put in the estimate for μ.

In order to get a problem which can be readily solved by an off-the-shelf optimiser, we first need to define *Second-Order Cone Programming*.

Definition 3.1. ***Second-order cone***
A second-order cone (also called quadratic cone) L^k is a set such that:

$$L^k = \{x \in \mathcal{R}^k \ : \ x_1 \geq 0, \ x_1^2 \geq \sum_{l=2}^{k} x_l^2\}$$

Definition 3.2. ***Second-Order Cone Programming***
A Second-Order Cone Program (SOCP) takes the form:

$$\begin{array}{lcccc} \min\limits_{x\in\mathcal{R}^n} & & c^T x & & \\ \text{s.t.} & b^i - \Theta^i x & \in & L^{k_i} & i \in \{1, ..., m\} \end{array}$$

Θ^i are linear maps $\mathcal{R}^n \mapsto \mathcal{R}^{k_i}$
b^i are vectors of $\mathcal{R}^{k_i}$

In order to get a linear objective function, we rewrite (R1'):

$$\begin{array}{lrcl} \max\limits_{x,s} & & s & \\ \text{s.t.} & s & \leq & \mu_0^T x - \sqrt{x^T W^{-1} x} \\ & x^T \Sigma_r x & \leq & \sigma_{max}^2 \\ & x & \in & \mathcal{F} \end{array}$$

Since:

$$s \leq \mu_0^T x - \sqrt{x^T W^{-1} x} \Leftrightarrow \begin{bmatrix} \mu_0^T x - s \\ Ux \end{bmatrix} \in L^{n+1}$$

U is the Cholesky factorisation of W^{-1}, *i.e.* $W^{-1} = U^T U$.

We can finally cast (R1) into SOCP form:

$$\begin{array}{lrcl} \max\limits_{x,s} & & s & \\ \text{s.t.} & \begin{bmatrix} \mu_0^T x - s \\ Ux \end{bmatrix} & \in & L^{n+1} \\ & x^T \Sigma_r x & \leq & \sigma_{max}^2 \\ & x & \in & \mathcal{F} \end{array} \tag{R1''}$$

3.3.4 Illustrative example

In order to illustrate the salient features of robust optimisation, we go back over the example introduced in section 3.1.

Both the (M1) problem and its robust equivalent (R1) are solved with two assets, budget and long-only constraints, the same risk model Σ_r and risk budget σ_{max}.

It is also assumed that the uncertainty set is given by $W = \begin{bmatrix} 71 & -24 \\ -24 & 71 \end{bmatrix}$.

Call x_{M1} and x_{R1} the solutions of the non-robust and robust problems respectively.

Similarly, call $\check{\mu}_{M1} = \check{\mu}_{M1}(x_{M1})$ the pessimal value for the expected return within the uncertainty set (the same definition holds for the (R1) problem). Its value is found by solving equation (12).

In addition to configurations (A) and (B), we also compare the two formulations in configurations (C) and (D), where μ is resp. such that $\mu = \begin{bmatrix} 0.0999 \\ 0.1001 \end{bmatrix}$ and $\mu = \begin{bmatrix} 0.05 \\ 0.15 \end{bmatrix}$.

Configuration (A) - $\mu = \begin{bmatrix} 0.09 \\ 0.11 \end{bmatrix}$

$$x_{M1} = \begin{bmatrix} 0.28 \\ 0.72 \end{bmatrix} x_{R1} = \begin{bmatrix} 0.40 \\ 0.60 \end{bmatrix} \qquad \check{\mu}_{M1} = \begin{bmatrix} 0.014 \\ -0.01 \end{bmatrix} \check{\mu}_{R1} = \begin{bmatrix} -0.002 \\ -0.002 \end{bmatrix}$$

$$x_{M1}^T \mu = 0.104 \quad x_{R1}^T \mu = 0.102 \qquad x_{M1}^T \check{\mu}_{M1} = -0.003 \quad x_{R1}^T \check{\mu}_{R1} = -0.002$$

Configuration (B) - $\mu = \begin{bmatrix} 0.11 \\ 0.09 \end{bmatrix}$

$$x_{M1} = \begin{bmatrix} 0.72 \\ 0.28 \end{bmatrix} x_{R1} = \begin{bmatrix} 0.60 \\ 0.40 \end{bmatrix} \qquad \check{\mu}_{M1} = \begin{bmatrix} -0.01 \\ 0.014 \end{bmatrix} \check{\mu}_{R1} = \begin{bmatrix} -0.002 \\ -0.002 \end{bmatrix}$$

$$x_{M1}^T \mu = 0.104 \quad x_{R1}^T \mu = 0.102 \qquad x_{M1}^T \check{\mu}_{M1} = -0.003 \quad x_{R1}^T \check{\mu}_{R1} = -0.002$$

Configuration (C) - $\mu = \begin{bmatrix} 0.0999 \\ 0.01001 \end{bmatrix}$

$$x_{M1} = \begin{bmatrix} 0.28 \\ 0.72 \end{bmatrix} x_{R1} = \begin{bmatrix} 0.50 \\ 0.50 \end{bmatrix} \qquad \check{\mu}_{M1} = \begin{bmatrix} 0.024 \\ -0.020 \end{bmatrix} \check{\mu}_{R1} = \begin{bmatrix} -0.003 \\ -0.003 \end{bmatrix}$$

$$x_{M1}^T\mu = 0.1000 \quad x_{R1}^T\mu = 0.1000 \qquad x_{M1}^T\check{\mu}_{M1} = -0.008 \quad x_{R1}^T\check{\mu}_{R1} = -0.003$$

Configuration (D) - $\mu = \begin{bmatrix} 0.05 \\ 0.15 \end{bmatrix}$

$$x_{M1} = \begin{bmatrix} 0.28 \\ 0.72 \end{bmatrix} x_{R1} = \begin{bmatrix} 0.28 \\ 0.72 \end{bmatrix} \qquad \check{\mu}_{M1} = \begin{bmatrix} -0.026 \\ -0.030 \end{bmatrix} \check{\mu}_{R1} = \begin{bmatrix} -0.026 \\ -0.030 \end{bmatrix}$$

$$x_{M1}^T\mu = 0.122 \quad x_{R1}^T\mu = 0.122 \qquad x_{M1}^T\check{\mu}_{M1} = 0.014 \quad x_{R1}^T\check{\mu}_{R1} = 0.014$$

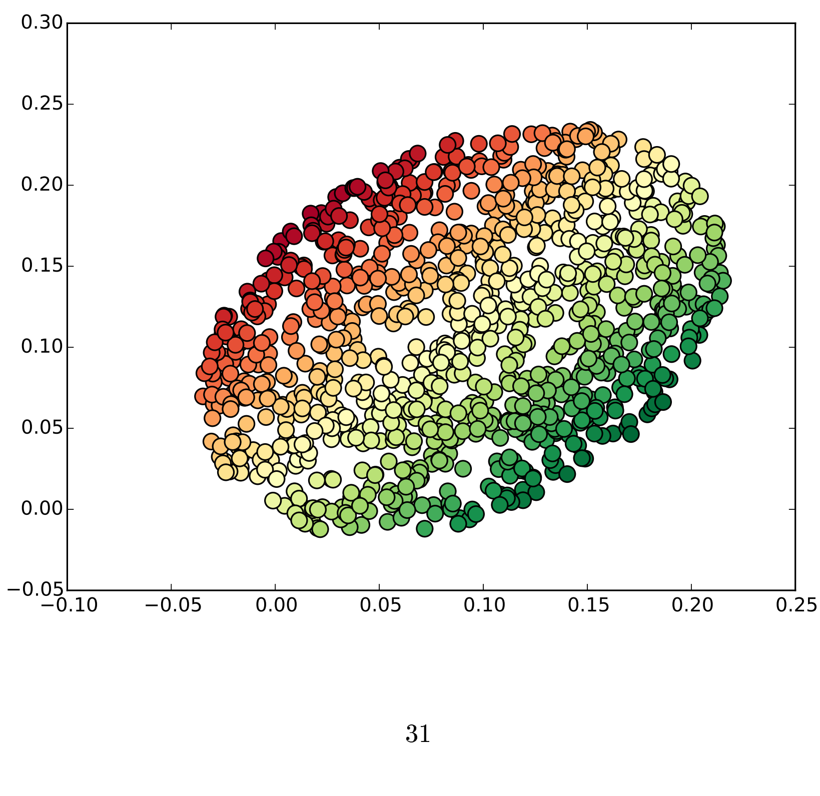
In an additional experiment, we compare the performance of the robust and non-robust problems for many realisations of μ within the uncertainty set.
More precisely:

1. The (M1) and (R1) problems are solved with the setting of configuration (A). This fixes values for x_{R1} and x_{M1}.
2. The difference in objective value $x_{R1}^T\mu - x_{M1}^T\mu$ is computed, assuming that μ drifts away from configuration (A). It takes some random value within $\mathcal{U} = \{\mu = \begin{bmatrix} 0.09 \\ 0.11 \end{bmatrix} + \nu \; : \; \nu^T \begin{bmatrix} 71 & -24 \\ -24 & 71 \end{bmatrix} \nu \leq 1\}$.

This experiment basically illustrates the behaviour of the two formulations when the realised return is different from the expected return.

Figure 6: Difference in objective value for samples μ

1000 uniform random samples for $\mu = \begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix}^T$ are taken in $\mathcal{U}$.
The colour represents the difference, with a green dot for a positive value and a red dot for a negative value.
The horizontal axis corresponds to μ_1, the vertical axis corresponds to μ_2.
x_{R1} and x_{M1} are obtained by solving the programs with value $\mu = \begin{bmatrix} 0.09 & 0.11 \end{bmatrix}^T$.



A few comments on the two sets of results are in order:

- As expected, the robust optimisation is less aggressive that its non-robust counterpart in configurations (A), (B), (C). When the components of the μ vector are similar, the robust optimisation takes this into account and weighs the optimal portfolio accordingly. In particular, in configuration (C), the robust portfolio is equally-weighted as μ_1 is almost the same as μ_2.
- When the components of the μ vector are very dissimilar as in configuration (D), the robust and non-robust solutions are the same. The solutions are different only if the relative preference between the different components of μ changes within the uncertainty set . Loosely speaking, the robust procedure seeks an optimum where the relative preferences are stable. It captures the potential change in the expected vs. the realised ranking (of μ components).
- In all the configurations, the value $x^T\mu$ is higher in the non-robust setting. This should be expected as this objective value of the non-robust problem. Similarly, the robust procedure guarantees that its solution is optimal, should μ reaches its worst-case value.
  As a result, in some parts of the uncertainty sets (namely, when $\mu_1 \geq \mu_2$), the robust formulation performs better, while in other parts it performs worse (figure 6). However, this approach remains silent as to which of the robust or non-robust formulation is better *on average.* The answer to this question would require some distributional assumptions of μ over the uncertainty set.

3.4 Theoretical properties and tractability of robust problems

In this section, we review results of robust optimisation that are important to provide a general understanding of the field. Particular focus is given to conservativeness and tractability issues.

3.4.1 Conservativeness of robust linear optimisation

Consider the generic linear-programming (LP) problem:

$$\min_x \{c^T x \mid Ax \leq b\} \tag{LP}$$

A is a $m \times n$ matrix
b is a size m vector
c and x are size n vectors.

It is assumed that uncertainty has been cast into the A matrix while b and c are known with certainty.
The robust counterpart (RLP) of this problem is:

$$\min_x \{c^T x \mid Ax \leq b \;\; \forall A \in \mathcal{U}\} \tag{RLP}$$

$\mathcal{U}$ is a set of $m \times n$ matrices.

Note that this setting is general enough to encompass a linearised version of (R1) (removing the risk-budget constraint).
Indeed:

$$\max_{x\in\mathcal{F}} \min_{\mu\in\mathcal{U}} \{\mu^T x\} \Leftrightarrow \max_{x,t} \{\; t \mid t \leq \min_{\mu\in\mathcal{U}} \mu^T x, \;\; x \in \mathcal{F}\} \Leftrightarrow \max_{x,t} \{\; t \mid t \leq \mu^T x \;\; \forall \mu \in \mathcal{U}, \;\; x \in \mathcal{F}\}$$

An important question raised by Ben-Tal and Nemirovski [4] addresses the structure that $\mathcal{U}$ may have. In the pioneering work of Soyster [49], uncertainty is modelled column-wise (i.e. each of the a_i columns of the A matrix varies in uncertainty set $\mathcal{U}_i$). This uncertainty structure is very conservative, as it assumes that all the uncertain coefficients in the matrix A simultaneously reach their worst-case value. Ben-Tal and Nemirovski give guidance to construct more reasonable uncertainty sets.

Consider b, c and $\mathcal{U}$ as fixed and call $\mathcal{P}$ the family of the (LP) programs with those fixed parameters and some $A \in \mathcal{U}$. The goal here is to find conditions on $\mathcal{U}$, which guarantee that the solution to (RLP) is no worst than the worst instance of $\mathcal{P}$.

As shown in [4], with a general $\mathcal{U}$, it may indeed be the case that:

- all the instances of $\mathcal{P}$ are feasible while the robust counterpart is infeasible
- or the robust counterpart has a worse objective value than all instances of $\mathcal{P}$

Such situations create a gap between a problem and its robust counterpart and should be avoided.

The primary condition to avoid such a gap is that $\mathcal{U}$ only has constraint-wise (*i.e.* row-wise) uncertainty. More precisely, with $\mathcal{U}_i$ the projection of $\mathcal{U} \subset \mathcal{R}^{m \times n} = \mathcal{R}^n \times \mathcal{R}^n .... \times \mathcal{R}^n$ onto the i-th element of the right-hand side, the uncertainty is said to be constraint-wise if:

$$\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times .... \times \mathcal{U}_m \tag{Condition 1}$$

A second condition required to get a suitable uncertainty set is what Ben-Tal and Nemirovski call the *boundedness assumption*. Call $\mathcal{F}_A$ the feasible set for the instance of $\mathcal{P}$ with uncertain matrix A. The boundedness assumption requires that there exists a convex compact set $\mathcal{C}$ that contains the feasible sets of all instances of $\mathcal{P}$.
That is:

$$\exists \mathcal{C} \subset \mathcal{R}^n, \quad \mathcal{C} \text{ convex compact}, \quad \mathcal{F}_A \subset \mathcal{C} \quad \forall A \in \mathcal{U} \tag{Condition 2}$$

Proposition 3.3. *Assume that Condition 1 and Condition 2 hold, then:*
(i) (RLP) is infeasible if and only if there exists an infeasible instance (LP)$\in \mathcal{P}$
(ii) If (RLP) is feasible with c^* its optimal value, we have:

$$c^* = \sup\{c^*(\text{LP}) \mid (\text{LP}) \in \mathcal{P}\}$$

A proof may be found in [4].

All the uncertainty sets that are exhibited in this dissertation verify those conditions and Proposition (3.3) ensures that there is no gap in the considered problems and their robust equivalents.

3.4.2 Complexity and tractability of robust problems

Consider a generic optimisation problem:

$$\begin{array}{llll} \min\limits_x & & f(x) & \\ \text{s.t.} & g_i(x,u) & \leq & 0 \qquad \forall i \in \{1,...,m\} \end{array} \tag{NLP}$$

x is the decision variable, while u is a vector of problem parameters.

The robust counterpart is:

$$\begin{array}{llll} \min\limits_x & & f(x) & \\ \text{s.t.} & g_i(x,u_i) & \leq & 0 \qquad \forall u_i \in \mathcal{U}_i \ , \ \forall i \in \{1,...,m\} \end{array} \tag{RNLP}$$

A natural question when considering the robust option is: what is the incremental complexity of the (RNLP) problem compared to the (NLP) problem? Indeed, a more difficult-to-solve problem should be expected as the (RNLP) problem is a semi-infinite program (assuming that any of the $\mathcal{U}_i$ sets is non-discrete).

We start by giving a bit more substance to the notion of *complexity*, in the specific context of optimisation.

Generally speaking, optimisation problems are solved with computers using numerical algorithms. Computers do a fixed number of operations per unit of time; complexity should hence be understood as the number of operations that the algorithms need to do. To differentiate the problems which are easy-to-solve from the problems that are difficult-to-solve (easy and difficult referring to the number of required operations), researchers have introduced the concept of *tractability*.

Definition 3.4. ***Tractable, intractable***
A *tractable* problem is such that it can be reformulated in a equivalent problem for which there exist solution algorithms with worst-case running time polynomial in the input size (Bertsimas *et al.* [11]). An *intractable* problem is a problem for which there exists instances that cannot be solved in polynomial time.

Tractable problems include LP, some forms of Quadratic Programs (QP), SOCP and Semi-Definite Programs (SDP).

As factor portfolios tend to invest in a large number of assets (in order to be well-diversified), it is crucial to get tractable formulations in our perspective.

3.4.3 Tractable formulations of robust counterpart of standard problems

In general, the robust equivalent of a tractable problem is not necessarily tractable. However, with careful choice of g_i and $\mathcal{U}_i$ in (RNLP), tractable formulations may be obtained. Such examples are reviewed in this section.

1. Polyhedral uncertainty for robust linear programming

With constraint-wise polyhedral uncertainty, (RLP) becomes:

$$\begin{array}{llcll} \min\limits_{x} & & c^T x & & \\ \text{s.t.} & a_i^T x & \leq & b_i & \forall a_i \in \mathcal{U}_i, \quad i \in \{1, ..., m\} \\ & \mathcal{U}_i & = & \{a_i \mid D_i a_i \leq d_i\} & i \in \{1, ..., m\} \end{array} \tag{PolRLP}$$

The inner problem can be simplified using the primal-dual relationship:

$$\begin{array}{lccc} \max\limits_{a} & & a_i^T x & \\ \text{s.t.} & D_i a_i & \leq & d_i \end{array} \tag{Primal}$$

$$\begin{array}{lccc} \min\limits_{s} & & s_i^T d_i & \\ \text{s.t.} & s_i^T D_i & = & x \\ & s_i & \geq & 0 \end{array} \tag{Dual}$$

We can then reformulate (PolRLP) as a linear program:

$$\begin{array}{lcccl} \min\limits_{x,s} & & c^T x & & \\ \text{s.t.} & s_i^T d_i & \leq & b_i & i \in \{1, ..., m\} \\ & s_i^T D_i & = & x & \\ & s_i & \geq & 0 & \end{array} \tag{PolRLP'}$$

2. Ellipsoidal uncertainty for robust linear programming

The problem (R1) is an example of robust linear programming with ellipsoidal uncertainty set.
A tractable formulation for the general case is shown below - this may be used in one wants to augment our model with constraint-wise uncertainty.

The problem with ellipsoidal uncertainty is:

$$\begin{array}{llcll} \min\limits_{x} & & c^T x & & \\ \text{s.t.} & a_i^T x & \leq & b_i & \forall a_i \in \mathcal{U}_i, \quad i \in \{1, ..., m\} \\ & \mathcal{U}_i & = & \{a_i = a_i^0 + \nu, \quad \nu^T W_i \nu \leq 1\} & \end{array} \tag{ElliRLP}$$

Exactly as in section (3.3.3), the use of the Karush-Kuhn-Tucker conditions on the inner problems yields:

$$\begin{array}{lccl} \min\limits_{x} & c^T x & & \\ \text{s.t.} & \begin{bmatrix} x^T a_i^0 - b_i \\ U_i x \end{bmatrix} \in & L^{n+1} & i \in \{1, ..., m\} \end{array} \tag{ElliRLP'}$$

U_i corresponds to the Cholesky decomposition of the W_i matrix.

The robust problem is a SOCP, it can be efficiently solved.

3. Robust quadratic programming

Robust counterparts of quadratically constrained problems are in general intractable (this is the case in particular for polyhedral and general ellipsoidal uncertainty sets - Bertsimas *et al.* [11]). However, in the case of a single ellipsoid uncertainty, the robust problems has a tractable SDP formulation. Such a formulation could be put in practice in the portfolio management context if one wants to have a robust risk constraint (as it translates into a quadratic constraint).

Consider the Quadratically-Constrained Linear program:

$$\begin{array}{llll} \min\limits_x & c^T x & & \\ \text{s.t.} & \|Ax\|^2 \leq & b^T x + c & \forall (A,b,c) \in \mathcal{U} \end{array} \tag{ElliQP}$$

$$\mathcal{U} = \{(A,b,c) = (A^0, b^0, c^0) + \sum_{i=1}^{k} u_i (A^k, b^k, c^k), \ \ \|u\| \leq 1\}$$

The important feature here is that the uncertain components (A, b, c) belong to the same ellipsoid - *i.e.* they are scaled by the same u in the above definition for $\mathcal{U}$. If we remove this assumption, we do not get a tractable formulation.

In order to go from a semi-infinite program to a tractable formulation, the below inner problem needs to be solved:

$$\begin{array}{ll} \max\limits_{A,b,c} & \|Ax\|^2 - b^T x - c \\ \text{s.t.} & (A,b,c) \in U \end{array} \tag{QCQP}$$

As shown by Ben-Tal and Nemirovski[3], the problem (QCQP) may be simplified with the use of the S-lemma.

Theorem 3.5. ***S-Lemma***
Consider F and G symmetric matrices of the same size.
Assume that the quadratic form $z^T F z + 2 f^T z + \phi$ is strictly positive for some z.
Then, the implication:

$$z^T F z + 2 f^T z + \phi \Rightarrow z^T G z + 2 g^T z + \gamma$$

holds true if and only if:

$$\exists \tau \geq 0, \ \ \begin{bmatrix} G - \tau F & g^T - \tau f^T \\ g - \tau f & \gamma - \tau \phi \end{bmatrix} \succeq 0$$

We apply the S-lemma to (QCQP), take the dual of the resulting semi-definite problem and obtain the below equivalent formulation for (ElliQP) (the full derivation may be found in [3]):

$$\begin{array}{lll} \min\limits_x & c^T x & \\ \text{s.t.} & \tau \geq & 0 \end{array} \tag{ElliQP'}$$

$$\begin{bmatrix} c^0 + 2x^T b^0 - \tau & \frac{1}{2} c^1 + x^T b^1 & \cdots & c^k + x^T b^k & (A^0 x)^T \\ \frac{1}{2} c^1 + x^T b^1 & \tau & & 0 & (A^1 x)^T \\ \vdots & & \ddots & & \vdots \\ \frac{1}{2} c^k + x^T b^k & 0 & & \tau & (A^k x)^T \\ A^0 x & A^1 x & \cdots & A^k x & I \end{bmatrix} \succeq 0$$

The problem (ElliQP') can readily be input into a numerical solver. The SDP formulation guarantees that it will be efficiently solved.

4 Robust formulations for factor portfolios

4.1 Where does uncertainty matter?

Interestingly, the Markowitz framework leads to deterministic optimisation problems. Though the asset returns are modelled as random variables, the hypotheses about investors preferences imply that only the two first moments of the distribution matter. As these moments are assumed to known, there is no stochastic component in the mean-variance optimisation. Robust optimisation can be seen as a way to reintroduce some uncertainty about the random processes.

Predicting assets returns is hard, to state the obvious. This difficultly has been well documented by academics (a comprehensive survey on the subject has been done by Welsh and Goyal [52]), in fact, this is the *raison d'etre* of a large portion of the financial industry.

Each approach to predict assets returns has to cope with specific hurdles. The quantitative approach suffers from non-stationary expected returns, Markovian prices path and efficient markets amidst others.

Robust optimisation can then be used as a tool to deal with uncertainty in the expected asset return. In fact, in mean-variance optimisation, errors in expected returns should be the primary focus.

A natural extension could be to model error in the second moment of the returns distribution (in a similar fashion as Goldfarb and Iyengar [27]). Alternatively, it could be considered to allow for uncertainty in the entire distribution (this approach is called *distributional uncertainty*).

In this work, choice is made to only focus on uncertainly in the μ vector. We do not allow for uncertainty in higher moments of the distribution. This choice is made because risk is typically easier to predict than returns, because dynamic risk adjustment is possible (by downsizing positions for example), because the performance of an investment manager is first and foremost judged on returns, and because allowing for higher moments uncertainty makes the problems less tractable and transparent.

As discussed in section 2.2, the μ vector is used to refer to two different things in our perspective. It may either be the factor exposures or the expected factor returns. Both factor returns and factor exposures are subject to uncertainty, which makes robust optimisation relevant in both cases. The estimation of factor returns faces the same hurdles as the estimation of assets returns, as detailed previously in this section.

The uncertainty around factor exposures takes multiple forms. Any factor has many different proxies, and there is no absolute choice for a proxy (*e.g.* a P/B ratio performs very well for a value strategy in the recovery period of the business cycle while an EV/EBITDA ratio performs better in the crash part of the business cycle). Is is likely that a stock appearing with a high factor exposure for one proxy has high exposure with other proxies but there is no one-to-one mapping between proxies. This constitutes a first source of uncertainty.

Uncertainty should also be taken into account because of the dynamic behaviour of factor exposures. As mentioned previously, some factor exposures can be modelled as random walks. In this sense, factor exposures are similar to asset prices (and dissimilar to factor returns, for whom the autocorrelation is closer to 0). Therefore, it should be expected that factor exposures change from one rebalance to another. The change may be drastic (*e.g.* in the case of a very short-term momentum factor) or it may be mild (*e.g.* in the case of a value factor that uses historical data). Therefore, there is uncertainty about the behaviour of the factor exposures between two rebalances. This should also be taken into account via the robust optimisation procedure.

A last reason to take uncertainty into account in the factor exposures can be found in measurement errors. Factor models are typically data-intensive and a factor built on numerous accounting metrics is error-prone (be it because of differences in accounting standards between countries, errors in currency conversion or poor handling of outliers).

For the sake of transparency, we only model uncertainty in the objective function. Should one want to build a factor portfolio with constraints on the (undesirable) factor exposures, it could be advised to also consider constraints robustness. This would not add major difficulty and we could recall the results of section 3.4.3.

4.2 Robust mean-variance problem

As seen in section 3.3.3, a tractable formulation for the ellipsoidal robust equivalent of problem (M1) is:

$$\begin{array}{rrcl} \max\limits_{x,s} & & & s \\ \text{s.t.} & \begin{bmatrix} \mu_0^T x - s \\ Ux \end{bmatrix} & \in & L^{n+1} \\ & x^T \Sigma_r x & \leq & \sigma_{max}^2 \\ & x & \in & \mathcal{F} \end{array} \tag{R1”}$$

In our setting, the μ_0 vector refers either to the expected factor returns or to the factor exposures, while U captures uncertainty around the nominal value μ_0.

4.3 Net-Zero Alpha Adjustment

The net-zero alpha adjustment formulation has been introduced by Ceria and Stubbs [14], the stated goal being a reduction of the conservativeness of robust optimisation. Indeed, as shown by Scherer [46], the standard robust optimisation procedure is equivalent to a Bayesian shrinkage estimation. Scherer shows that robustness simply drags the expected returns vector towards 0. Each component of the μ vector is shrunk by a amount proportional to its uncertainty.

A way to avoid this behaviour is to introduce a special structure in the uncertainty set $\mathcal{U}$. In this section, a tractable formulation for the robust equivalent of problem (M1) with net-zero alpha adjustment is derived.

4.3.1 Uncertainty set

The uncertainty set is now given by:

$$\mathcal{U} = \{ \ \mu = \mu_0 + \nu \ : \ \nu^T W \nu \leq 1, \quad e^T \nu = 0 \ \}$$

The additional constraint $e^T\nu = 0$ ensures that for any downward adjustment in the μ vector, there is an offsetting upward adjustment. It guarantees that the vector is not unilaterally shrunk towards 0. $\mathcal{U}$ remains a convex set with this additional constraint.

4.3.2 Equivalent tractable formulation

The net-zero alpha adjustment robust equivalent of problem (M1) is then:

$$\begin{array}{rrcl} \max\limits_{x} & \min\limits_{\mu \in \mathcal{U}} & \mu^T x & \\ \text{s.t.} & x^T \Sigma_r x & \leq & \sigma_{max}^2 \\ & x & \in & \mathcal{F} \end{array} \tag{R2}$$

We need to solve the inner problem to get the pessimal value for μ.
For a fixed $x \in \mathcal{F}$, we look for:

$$\begin{array}{rrcl} \min\limits_{\nu} & & \nu^T x & \\ \text{s.t.} & \nu^T W \nu & \leq & 1 \\ & e^T \nu & = & 0 \end{array} \tag{13}$$

The Lagrangian of this problem is:

$$L(\nu, \delta, \gamma) = \nu^T x + \delta e^T \nu + \gamma(\nu^T W \nu - 1) \tag{14}$$

The KKT conditions yield:

$$\nu = \frac{-W^{-1}(x - \frac{e^T W^{-1} x}{e^T W^{-1} e} e)}{\sqrt{(x - \frac{e^T W^{-1} x}{e^T W^{-1} e} e)^T W^{-1} (x - \frac{e^T W^{-1} x}{e^T W^{-1} e} e)}} \tag{15}$$

Hence, the pessimal value for μ is:

$$\mu = \mu_0 - \frac{W^{-1}(x - \frac{e^T W^{-1} x}{e^T W^{-1} e} e)}{\sqrt{(x - \frac{e^T W^{-1} x}{e^T W^{-1} e} e)^T W^{-1} (x - \frac{e^T W^{-1} x}{e^T W^{-1} e} e)}} \tag{16}$$

The problem (R2) becomes:

$$\begin{array}{rrcl} \max\limits_{x} & \mu_0^T x & - & \sqrt{(x - \frac{e^T W^{-1} x}{e^T W^{-1} e} e)^T W^{-1} (x - \frac{e^T W^{-1} x}{e^T W^{-1} e} e)} \\ \text{s.t.} & x^T \Sigma_r x & \leq & \sigma_{max}^2 \\ & x & \in & \mathcal{F} \end{array} \tag{R2'}$$

The square-root term may further be input into a second-order cone.
Since:

$$x^T (W^{-1} - \frac{W^{-1} e e^T W^{-1}}{e^T W^{-1} e}) x = (x - \frac{e^T W^{-1} x}{e^T W^{-1} e} e)^T W^{-1} (x - \frac{e^T W^{-1} x}{e^T W^{-1} e} e) \tag{17}$$

Take the Cholesky decomposition V such that:

$$V^T V = W^{-1} - \frac{W^{-1} e e^T W^{-1}}{e^T W^{-1} e} \tag{18}$$

The problem (R2') is then ready to be input into a numerical solver:

$$\begin{array}{rrcl} \max\limits_{x} & \mu_0^T x & - & s \\ \text{s.t.} & \begin{bmatrix} s \\ Vx \end{bmatrix} & \in & L^{n+1} \\ & x^T \Sigma_r x & \leq & \sigma_{max}^2 \\ & x & \in & \mathcal{F} \end{array} \tag{R2''}$$

4.4 Robust deviation

The robust deviation formulation has been introduced by Chen *et al.* [15], and further transposed to the portfolio management context by Natarajan *et al.* [40]. Again, the main goal here is to combat the conservativeness of robust optimisation. More precisely, when uncertain parameters with asymmetric distribution are modelled with ellipsoidal uncertainty sets the sets tend to be too wide. Ellipsoid are symmetric by nature and the robust ellipsoidal formulation is too conservative in this case.

Chen *et al.* introduced deviation measures, called *forward* and *backward* deviations, that enable to capture distributional asymmetry. In this section, we first present an asymmetric uncertainty set, then derive the robust deviation counterpart of problem (M1) and lastly discuss some features of the forward and backward deviations.

4.4.1 Uncertainty set

The uncertainty set is given by:

$$\mathcal{U} = \{\ \mu = \mu_0 + Mz\ :\ z \in \mathcal{H}\ \}$$

$$\mathcal{H} = \{\ z = y - w\ :\ y, w \geq 0,\ \|P^{-1}y + Q^{-1}w\| \leq \Omega_{max}\ \}$$

$M^T M = \Sigma_\mu$ where Σ_μ is the covariance matrix of μ.
μ_0 is the expected value for μ.
z is a size n random vector.
$P = diag(p_1, ..., p_n)$ and $Q = diag(q_1, ..., q_n)$, $p_i, q_i \geq 0 \quad i = 1, ..., n$
Ω_{max} is the budget of uncertainty.

In this setting, the uncertain μ is more easily understood as a random variable, and this is why we take z as a random vector (the distributional properties of μ were factored in the W matrix previously). Ω_{max} represents the maximum total deviation allowed and it scales the size of the uncertainty set.

z needs to be constructed as a standardized random variable, such that:

$$\mathrm{E}(z) = 0 \quad \text{and} \quad \mathrm{E}(zz^T) = I$$

4.4.2 Equivalent tractable formulation

The robust equivalent for problem (M1) is:

$$\begin{array}{rrcl} \max_{x} & \min_{\mu\in\mathcal{U}} & \mu^T x & \\ \text{s.t.} & x^T\Sigma_r x & \leq & \sigma^2_{max} \\ & x & \in & \mathcal{F} \end{array} \tag{R3}$$

It may be reformulated as:

$$\begin{array}{rrcl} \max_{x,t} & & t & \\ \text{s.t.} & 0 & \leq & \mu_0^T x - t + \min_{z\in\mathcal{H}} z^T M^T x \\ & x^T\Sigma_r x & \leq & \sigma^2_{max} \\ & x & \in & \mathcal{F} \end{array} \tag{19}$$

Further, we apply the transformations $y = Py$ and $w = Qw$.
The problem then becomes:

$$\begin{array}{rrcl} \max_{x,t} & & t & \\ \text{s.t.} & 0 & \leq & \mu_0^T x - t + \min_{\{y,w\geq 0,\ \|y+w\|\leq\Omega_{max}\}} (Py)^T M^T x - (Qw)^T M^T x \\ & x^T\Sigma_r x & \leq & \sigma^2_{max} \\ & x & \in & \mathcal{F} \end{array} \tag{20}$$

Equivalently:

$$\begin{array}{rrcl} \max_{x,t} & & t & \\ \text{s.t.} & 0 & \leq & \mu_0^T x - t - \max_{\{y,w\geq 0,\ \|y+w\|\leq\Omega_{max}\}} (Qw)^T M^T x - (Py)^T M^T x \\ & x^T\Sigma_r x & \leq & \sigma^2_{max} \\ & x & \in & \mathcal{F} \end{array} \tag{R3'}$$

As shown by Chen *et al.*, the inner problem may be simplified using the below lemma.

Lemma 4.1. *Consider the optimisation problem:*

$$\begin{array}{rrrcl} f^* = & \max_{y,w} & a^T y + b^T w & & \\ & s.t. & \|y+w\| & \leq & \Omega_{max} \\ & & y, w & \geq & 0 \end{array} \tag{DevMax}$$

Then

$$f^* = \Omega_{max}\|u\|,\ \ u_j = \max\{a_j, b_j, 0\}\ \ \forall j \in \{1, \ \ , n\}.$$

Proof. A proof is given in annexes A.3. □

As a result, problem (R3') is equivalent to:

$$\begin{array}{rrcl} \max_{x,t,u} & & t & \\ \text{s.t.} & 0 & \leq & \mu_0^T x - t - \Omega_{max}\|u\| \\ & u & \geq & -PM^T x \\ & u & \geq & QM^T x \\ & u & \geq & 0 \\ & x^T\Sigma_r x & \leq & \sigma^2_{max} \\ & x & \in & \mathcal{F} \end{array} \tag{21}$$

For an optimal solution, $u = \max\{-PM^T x, QM^T x, 0\}$ hence the inner approximation is tight.

Lastly, the problem is reformulated as a SOCP:

$$
\begin{array}{lrcl}
\max\limits_{x,t,u,s} & & & t \\
\text{s.t.} & 0 & \leq & \mu_0^T x - t - s\Omega_{max} \\
& \begin{bmatrix} s \\ u \end{bmatrix} & \in & L^{n+1} \\
& u & \geq & -PM^T x \\
& u & \geq & QM^T x \\
& u & \geq & 0 \\
& x^T \Sigma_r x & \leq & \sigma_{max}^2 \\
& x & \in & \mathcal{F}
\end{array}
\tag{R3''}
$$

4.4.3 The forward and backward deviations

While many deviation measures could be used to fill the P and Q matrices (including the standard deviation), Chen *et al.* [15] present candidates that enable to get a probabilistic guarantee of constraint violation for a broad range of distributions. In their setting, p_i (resp. q_i) is the forward (backward) deviation of the i^{th}component of z.

Definition 4.2. ***Forward and backward deviations***
Let ζ be a (univariate) centred random variable.
The forward deviation p and backward deviation q for ζ are:

$$p = \sup_{t>0} \frac{1}{t} \sqrt{2 \ln(E[\exp(t\zeta)])}$$

$$q = \sup_{t>0} \frac{1}{t} \sqrt{2 \ln(E[\exp(-t\zeta)])}$$

Given a set of independent samples of the variable ζ, this definition may directly be used to construct estimators for the forward and backward deviations.

Also, as shown in [15], with the additional assumptions that the components of z are independent and with bounded support, we have the probabilistic guarantee of constraint violation:

$$\Pr(t \leq \mu_0^T x + z^T M^T x) \geq 1 - \exp(-\Omega_{max}^2/2)$$

This yields a useful lower bound for the optimal objective value, with only mild assumptions about the distribution of μ.

The forward and backward deviations also have intuitive connections with the standard deviation.

Proposition 4.3. *For ζ (univariate) centred random variable with standard deviation σ:*
(i) $p \geq \sigma$ and $q \geq \sigma$
(ii) if ζ is normally distributed, $p = q = \sigma$

Proof. A proof is given in annexes A.4. □

4.4.4 Robust deviation and ellipsoidal uncertainty sets

Adding some assumptions on the P and Q matrices, it may be shown that the robust deviation uncertainty set is equivalent to the ellipsoidal uncertainty set. In this light, the former appears as an asymmetric generalisation of the latter.

Equivalent uncertainty sets also guarantee that the solutions to the problems (R1) and (R3) are the same (when the conditions on P and Q are validated). We will see that this is indeed the case in our numerical experiments.

Recall that the ellipsoidal uncertainty set for problems (R1) and (R3) are defined by:

$$\mathcal{U}_1 = \{ \ \mu = \mu_0 + \nu \ : \ \nu^T W \nu \leq 1 \ \}$$

$$\mathcal{U}_3 = \{ \ \mu = \mu_0 + Mz \ : \ z = y - w \ : \ y, w \geq 0, \ \|P^{-1}y + Q^{-1}w\| \leq \Omega_{max} \ \}$$

Assume that $W = \frac{1}{\Omega_{max}^2}\Sigma_\mu^{-1}$ and that $P = Q = I$, with I the identity matrix, then:

$$\mathcal{U}_1 = \{\ \mu = \mu_0 + Mz\ :\ \|z\| \leq \Omega_{max}\ \}$$

$$\mathcal{U}_3 = \{\ \mu = \mu_0 + Mz\ :\ z = y - w\ :\ y, w \geq 0,\ \|y + w\| \leq \Omega_{max}\ \}$$

We show that $\mathcal{U}_1 = \mathcal{U}_3$.

Proof. For $\mu \in \mathcal{U}_1,\ \mu = \mu_0 + Mz$, and $i = 1, ..., n$, choose:

$$(y_i, w_i) = \begin{cases} (z_i, 0) & \text{if } z_i \geq 0 \\ (0, -z_i) & \text{otherwise} \end{cases}$$

This choice guarantees $\mu \in \mathcal{U}_3$.

Conversely, for $\mu \in \mathcal{U}_3$:

$$\|y + w\| \leq \Omega_{max} \Rightarrow \|y - w\| \leq \Omega_{max} \Rightarrow \|z\| \leq \Omega_{max}$$

Thus, $\mu \in \mathcal{U}_1$. $\square$

5 Numerical experiments

5.1 Literature review on computational results for robust portfolios

Most common optimisation problems do not have a closed-form solution. The solution needs to be computed numerically with an optimiser. Naturally, adding a layer of complexity by enforcing robustness does not allow for more intelligible solutions (except in some limiting cases). When advising the use of a new robust formulation, researchers have thus generally provided numerical results to illustrate (the benefits of) their approach! Some of those results are commented here.

The first set of results with specific focus on portfolio optimisation comes from Ben Tal *et al.* in 2000 [5]. They apply robust optimisation to multi-period asset allocation (*i.e.* maximising the terminal value of the investment in n assets over T periods). They show that linear programming is no longer a valid option if asset returns are considered as random variables. On top of this, the standard stochastic programming formulation is intractable most often. Hence, they advise robust optimisation with ellipsoidal uncertainty (given partial momentum information on the returns).

Further, they experiment with simulated data and compare the robust, stochastic and *nominal* formulations (*nominal* meaning that the random variables are simply replaced with their expected values). They find that the robust formulation reduces risk of losses, that is has more balanced terminal values and that, in most cases, it provides higher average returns. The robust problem has lower risk is and this is clear from their results; on the other hand, the superiority of this formulation in term of average returns is arguable. They chose a high trading cost parameter (set to 10% for buying and selling), this may artificially create an unfair advantage to the robust formulation as it has more stable solutions.

Another set of important empirical results comes from Goldfarb and Iyengar [27], whose approach was previously discussed in section 3.3.1. They derive the robust formulations for the Markowitz problems when the expected returns and covariance matrix are estimated via a factor model. As a by-product of estimation, they obtain confidence regions for the parameters. For a fixed confidence level, these regions may directly be used as uncertainty sets.

Applying this to simulated data for the maximum Sharpe ratio problem, Goldfarb and Iyengar find that the average Shape ratio of the standard problem is higher than the average Sharpe ratio of the robust problem. More precisely, the ratio average robust Sharpe / average standard Sharpe is below 1, and decreasing for an increasing confidence level (is goes to 0.8 for a confidence level near 1). On the other hand, the worst-case ratio is better for the robust problem (ratio worst-case robust / worst-case standard is above 1 and increasing). They conclude: "*Thus, at a modest 20% reduction in the mean performance, the robust framework delivers an impressive 200% increase in the worst-case performance*".

Goldfarb and Iyengar also run tests on real market data, but their results are inconclusive: for high and low confidence levels, the robust portfolio has better performance, but it does not do as good as the standard portfolio for a medium confidence level.

Ceria and Stubbs [14] build on the concept of *true, actual* and *estimated efficient frontiers* to investigate the empirical properties of robust optimisation. The *true* frontier corresponds to the efficient frontier obtained by solving the Markowitz problems with the true (unknown) values for the expected returns and covariance matrix. The *estimated* frontier corresponds to the frontier obtained using the estimated (*i.e.* subject to error) parameters. Lastly, the *actual* frontier is obtained by feeding the true parameters in the estimated portfolios (portfolio in the estimated efficient frontier). The actual frontier is below the true frontier by construction, whereas the estimated frontier is likely to be above the actual and true frontiers (provided the mean estimation error is null).

Running numerical tests with simulated data, Ceria and Stubbs show that both the estimated and actual robust frontier are closer to the true frontier (than their non-robust equivalent). These results are in contradiction with those of Goldfarb and Iyengar: indeed, that the actual robust frontier lies above the actual standard frontier means that the average robust Sharpe ratio is higher than the average standard Sharpe.

The crux of the results of Ceria and Stubbs is the assumption that the mean estimation error of the expected return is null. Indeed, with null mean error, some assets see their expected return overestimated while others have their expected return underestimated. In the standard Markowitz formulation, the assets with positive estimation error are then overweighted (compared to the optimal solution with true parameters) while the assets with negative estimation error are underweighted. In the robust formulation, the effect of the estimation error is dampened by allowing uncertainty in the expected return. Further, this reduces overweighting/underweighting and brings the robust frontier closer to the true frontier. The assumption of null mean estimation error is contentious and we do not make it in our numerical tests.

In contrast to most previous results which focus on the *performance*, Perchet *et al.* [43] investigate the *composition* of robust portfolios, by comparing risk-based portfolios and robust portfolios. Risk-based portfolios are portfolios that are constructed without explicit values for the expected returns; these can be the equally-weighted, risk-parity, inverse-variance and minimum-variance portfolios. Perchet *et. al.* show that in the limit of low uncertainty, robust portfolios correspond to the standard Markowitz portfolios. Adding some assumptions on the structure of the uncertainty set, they also show that in the limit of high uncertainty, robust portfolios converge toward appropriate risk-based portfolios. In most cases, a weighted combination of the standard and risk-based portfolios provides a good representation of the robust portfolio. Because of the uncertainty structure that has been chosen, we will find the same type of convergence in our experiments.

Though we may leverage these results when appropriate, our perspective is quite different to existing empirical research as we focus on factor portfolios.

5.2 Data set

As discussed in section 2.2 and section 2.4, optimising for factors exposures is quite different to optimising for factor returns, even if both may be fed into the μ vector of problem (M1).

Hence, two distinct data sets are used:

1. Simulated data for factor exposures. The goal here is to check whether robust portfolio may get *better* factor exposures than standard portfolios, in a context where the asset-level factor exposures are stochastic. *Better* refers to either higher exposures, lower turnover or more stable exposures. The asset-level factor exposures are simulated, and we study how the robust portfolio may capture them, rebalance after rebalance.

2. Market data for factor returns. The goal here is to check whether robust combinations of factor portfolios may outperform their standard equivalent, on a risk-return adjusted basis. Focus is given to the distributional properties of the returns of the optimal portfolio. The data for the factor returns are taken from the Fama-French Data Library [24].

Choice is made to simulate data for experiments on factor exposures because getting real market data is cumbersome (standard fundamental data providers like I.B.E.S., Worldscope or FactSet do not provide free sample) and with limited added value. Simulating the data also enables to extract the salient features of robustness, in particular with regard to the trade-off between maximizing exposure and reducing turnover.

On the other hand, using empirical distributions for the factor returns provides straightforward comparison between the robust and non-robust portfolios (*e.g.* performance analysis during the momentum drawdown of 2009 can be conducted).

Fama and French provide research data that can be used to proxy factor returns for a long period of time and with good geographical coverage. It should nonetheless be pointed out that their approach to estimate factor returns differs to what we developed in section 1.4. Their factor returns are essentially the returns of long-short portfolios. The stocks are ranked according to their score on a fundamental metric, and portfolios are built by going long the top percentiles and short the bottom percentiles. Taking such returns remains a valid option in our perspective as they show similar distributional properties as the *pure* factor returns, as discussed in section 2.4.

5.3 Procedures for numerical experiments

5.3.1 Testing factor exposures portfolios

As focus is given to both the exposures and the turnover, we take a portfolio rebalancing approach to compare the different problem formulations. It enables to capture the dynamic behaviour of the portfolio when asset-level exposures vary.

Of course, the most adapted problem formulation largely depends on the modelling of asset-level factor exposures. Choice is made here to simulate mean-reverting processes for the factor exposures paths (variants of Ornstein-Uhlenbeck). Though it may be questionable (*e.g.* because some factors may exhibit random-walk or trending types of behaviour), mean-reversion ensures that the relative preferences in the optimisation process change, hence pinpointing a careful use of the turnover budget. A behaviour close to random-walk may also be encompassed in this setting by taking a very low mean-reversion speed.

The procedure in details:

1. Generate the asset-level factor exposures paths $\{\mu_1, ..\mu_i, ..., \mu_T\}$ using Monte-Carlo simulation.
2. Let m be the number of rebalances, such that $T = 0 \pmod m$, and $\tau = \frac{T}{m}$ the rebalance frequency.
   At each rebalance date $t = k\tau,\ k = 1, ..., m$, get the optimal portfolios x_t^* by solving the optimisation problems (M1), (R1), (R2) and (R3).
   Use the expected asset-level factor exposures at mid-rebalance - *i.e.* the current exposure discounted for mean-reversion up to half of the rebalance frequency, that is $\mathrm{E}(\mu_{t+\frac{\tau}{2}})$.
3. Compute portfolio performance metrics by taking averages over the paths $\{\mu_1, ..., \mu_T\}$.

In order to reduce the Monte-Carlo noise, a lengthy path $\{\mu_1, .., \mu_T\}$ is simulated ($T = m\tau$, with $m = 10,000$ portfolio rebalances). This is equivalent to simulating many shorter paths. The same path is used to fairly compare the different problem formulations (the random seed is fixed).

To improve convergence speed, antithetic variates are used where appropriate - *i.e.* if taking symmetrical $(\mu, -\mu)$ does not distort the distribution.

This setting allows to model a concrete situation where the exposures move at a higher frequency than the portfolio rebalance process. Typically, factor exposures are computed on a daily basis while the portfolio rebalance occurs on a weekly, monthly or quarterly basis.

Simulation of the factor exposure paths

The driving equation for μ is:

$$\Delta\mu_i = \mu_i - \mu_{i-1} = \Psi(\theta - \mu_{i-1}) + \epsilon_i$$

$\mu_i,\ i = 1, ..., T$ are vectors of size n.
Ψ is a $n \times n$ diagonal matrix of mean-reversion speed.
θ is a size n vector of long-term factor exposures.
$\epsilon_i,\ i = 1, ..., T$ are size n vectors of random independent increments.

1. Symmetric increment
   For $\Sigma_{\Delta\mu} = H^T H$ the covariance matrix of $\Delta\mu$ and ξ_i Brownian increment $\mathcal{N}(0,1)$ whom components are independent, take:
   $$\epsilon_i = H^T \xi_i$$
2. Asymmetric increment
   Let $\xi_i = [\xi_i^1, ..., \xi_j^n]^T$ be a vector of independent Bernoulli increments, and $\omega = [\omega_1, ..., \omega_n]$ a vector of parameters controlling the skewness of the distribution, such that:
   $$\begin{array}{rcl} \Pr(\xi_i^j = \omega_j) & = & \beta \\ \Pr(\xi_i^j = \frac{-\beta\omega_j}{1-\beta}) & = & 1-\beta \end{array} \quad \forall i \in \{1, ..., T\},\ j \in \{1, ..., n\}$$
   Further, take:
   $$H^T H = \frac{1-\beta}{\beta}\mathrm{diag}(\frac{1}{\omega_1^2}, ..., \frac{1}{\omega_n^2})\Sigma_{\Delta\mu}^T \quad \text{and} \quad \epsilon_i = H^T \xi_i$$

By construction, in both the symmetric and asymmetric case:

$$E(\epsilon_i) = 0 \quad E(\epsilon_i \epsilon_i^T) = \Sigma_{\Delta\mu}$$

However, the distributions differ on their higher-order moments. Positive skew (i.e. forward deviation) can be modelled in the Bernoulli increments by taking $\omega_j \geq 0$ and β close to 0. Conversely, negative skew (backward deviation) can be modelled by taking $\omega_j \leq 0$ and β close to 0.

Calibration of the uncertainty sets

An important issue that has been omitted so far is the calibration of the robustness parameters, namely W for problems (R1) and (R2) and P, Q and Ω_{max} for (R3). In real-life, the calibration exercise much depends on the μ estimation process (several examples have been discussed by Stubbs and Vance [50]).

In our case, ultimately, we want to control the probability that μ remains within a certain set *between two rebalance dates.*
As τ is the rebalance frequency, two successive rebalance dates are t and $t+\tau$. The cumulative increment between two rebalances is:

$$\mu_{t+\tau} - \mu_t = \sum_{i=1}^{\tau} (I - \Psi)^{i-1} \Psi \theta + ((I - \Psi)^{\tau} - I)\mu_t + \sum_{i=1}^{\tau} (I - \Psi)^{i-1} \epsilon_{t+\tau+1-i}$$

I is the $n \times n$ identity matrix.

Thus, the time t conditional variance V_t is:

$$V_t(\mu_{t+\tau} - \mu_t) = \sum_{i=1}^{\tau} (I - \Psi)^{2(i-1)} V_t(\epsilon_{t+\tau+1-i}) = \sum_{i=1}^{\tau} (I - \Psi)^{2(i-1)} \Sigma_{\Delta\mu}$$

1. Symmetric increment
   In this case, $\mu_{t+\tau} - \mu_t$ is normally distributed. A standard result in statistics [26] is that the ρ-level confidence region around $E_t(\mu_{t+\tau} - \mu_t)$ is given by:

   $$v \ : \ v^T (\sum_{i=1}^{\tau} (I - \Psi)^{2(i-1)} \Sigma_{\Delta\mu})^{-1} v \leq \chi_n^2(\rho)$$

   χ_n^2 is the inverse cumulative distribution function of the chi-squared distribution with n degrees of freedom. Therefore, in problems (R1) and (R2), W is chosen such that:

   $$W = (\chi_n^2(\rho) \sum_{i=1}^{\tau} (I - \Psi)^{2(i-1)} \Sigma_{\Delta\mu})^{-1}$$

   With such symmetric increments, we also have $P = Q = I$ (proposition 4.3).
   Thus, there is equivalence between the uncertainty sets of problems (R1) and (R3), provided the choice in problem (R3):

   $$\Omega_{max} = \sqrt{\chi_n^2(\rho)}$$

2. Asymmetric increments
   Exact calibration in this case is more involved. Parameters need to be approximated, keeping in mind that it is crucial to ensure equivalence between the uncertainty sets of problems (R1) and (R3) in the limiting case with symmetry ($\beta = 0.5$) for fair comparison.
   We keep the same value for W as the one found in the symmetric case (noting that the confidence region is no longer exact).
   The forward and backward deviations P and Q, are computed in-sample, using definition (4.2). The budget of uncertainty $\Omega_{max}$is calibrated such that the uncertainty sets of problems (R1) and (R3) are the same for $\beta = 0.5$. More precisely,

   $$\Omega_{max} = \sqrt{\frac{\chi_n^2(\rho)}{\frac{1}{4}(e^T P e + e^T Q e)^2}} \tag{22}$$

5.3.2 Testing factor returns portfolios

This experiment compares the performance of the different formulations directly in the expected return space. Factor returns are downloaded from the Fama French Data Library [24] - this is the very same set of data as in section 2.4 (daily returns for 5 US risk premia).

Further:

1. The data set is split into a training set and a test set of equal size. There are 6531 observations for all the factors, each set has about 12 years of data.
2. The optimisation parameters are calibrated on the test set. The problems (M1), (R1) and (R3) are solved with those inputs.
3. The optimised portfolios' performance is evaluated on the test set.

The second step requires a bit more details. The parameters that need to be calibrated are the expected return μ, the risk model Σ_r, the covariance matrix of the error in expected return Σ_μ, the forward and backward deviation of the error and the budget of uncertainty Ω_{max}.

The expected return and the risk model are calibrated by computing respectively the sample mean and covariance matrix on the full training data. To get the parameters required for the robust formulations, we split the training set into smaller subperiods and compute the sample mean for each. We then calculate the sample covariance matrix, and the forward and backward deviation of the sample means.

Alike in the previous experiment, W is obtained using the confidence region for Gaussian vector:

$$W = (\chi_n^2(\rho)\Sigma_\mu)^{-1}$$

The budget of uncertainty is obtained using formula (22).

The risk budget σ_{max} is set to 20%, and there is a long-only constraint.

Lastly, some performance metrics are reported below to provide comparison for the behaviour of the factors between the training set and the test set.

Table 3: Fama-French factor returns - Training set (1990-2002) - Test set (2002-2015)
μ annualised average return, σ annualised standard deviation
p forward deviation, q backward deviation

	Training set					Test set				
	SMB	HML	RMW	CMA	UMD	SMB	HML	RMW	CMA	UMD
μ	0.0024	0.0366	0.0724	0.0487	0.1411	0.0316	0.0131	0.0177	0.0076	0.0111
σ	0.0925	0.1006	0.0776	0.0820	0.1209	0.0905	0.0881	0.0579	0.0463	0.1509
p	0.0058	0.0063	0.0049	0.0052	0.0076	0.0057	0.0056	0.0036	0.0029	0.0095
q	0.0059	0.0063	0.0049	0.0052	0.0077	0.0057	0.0055	0.0036	0.0029	0.0097

5.4 Relevant parameters and performance criteria

Relevant parameters

With regard to the simulated data, we compare the output of the different problem formulations while varying the number of assets in the portfolio, adding long-only and turnover constraints and modelling the asset-level exposures with different behaviours (as discussed in the previous section).

As shown by Scherer [46], a critical parameter of for robust problems is the budget of uncertainty (factored in W for problems (R1) and (R2), and explicitly modelled as Ω_{max} for problem (R3)). Hence, different levels of uncertainty are tested. In our perspective, the budget of uncertainty controls the potential drift in μ between two rebalance dates that the optimisation prepares for.

In the numerical experiments, the budget of uncertainty is inferred from a confidence level ρ, itself calibrated such that between two rebalances μ remains in the uncertainty set $\mathcal{U}$ with probability ρ.

Performance criteria

The problem of interest is the maximum exposure problem (M1); logically, the main performance criteria is the average portfolio exposure, defined as:

$$\bar{\mu} = \frac{1}{T}\sum_{i=1}^{T} \mu_i^T x_i^*$$

x_i^* is the prevailing optimal portfolio a time i. Between two rebalances, $x_i^* = x_t^*$ with $t = \lfloor \frac{i}{\tau} \rfloor \tau$.

Similarly, the average turnover is computed as:

$$\bar{TO} = \frac{1}{m}\sum_{t=2}^{m} e^T |x_t^* - x_{t-1}^*|$$

The average portfolio risk is:

$$\bar{\sigma} = \frac{1}{m}\sum_{t=1}^{m} \sqrt{x_t^{*T}\Sigma_r x_t^*}$$

Note that the risk does not change between two rebalances.

Ideally, the turnover and risk budgets are entirely consumed at each rebalance, such that $\bar{TO} \simeq TO_{max}$ and $\bar{\sigma} \simeq \sigma_{max}$. This enables easier comparison of average exposures.

The portfolio risk may also be supplemented by a diversification measure $\bar{d}$. Building on the Herfindahl-Hirschmann concentration index, define:

$$\bar{d} = 1 - \frac{1}{m}\sum_{t=1}^{m} \frac{\frac{x_t^{*T} x_t^*}{(e^T|x_t^*|)^2} - \frac{1}{n}}{1 - \frac{1}{n}}$$

A perfect diversification with equal-weighting at each rebalance leads to $\bar{d} = 1$. Conversely, a portfolio systematically invested in a single-asset has diversification $\bar{d} = 0$.

In order to get further insight into the distributional properties of the optimised portfolio, the standard deviation $\bar{\sigma_\mu}$ and Pearson's skewness $\bar{\gamma_\mu}$ of the portfolio exposures are computed:

$$\bar{\sigma_\mu} = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(\mu_i^T x_i^* - \bar{\mu})^2} \quad \text{and} \quad \bar{\gamma_\mu} = \frac{1}{T}\sum_{i=1}^{T}\left[\frac{\mu_i^T x_i^* - \bar{\mu}}{\bar{\sigma_\mu}}\right]^3$$

The minimum exposure $\bar{\mu}_{min}$, maximum exposure $\bar{\mu}_{max}$ and 5^{th}-percentile exposure $\bar{\mu}_{5\%}$ may also be of interest.

Let F be the empirical cumulative distribution function for $(\mu_1^T x_1^*, ..., \mu_T^T x_T^*)$:

$$\bar{\mu}_{min} = \min(\mu_1^T x_1^*, ..., \mu_T^T x_T^*) \qquad \bar{\mu}_{max} = \max(\mu_1^T x_1^*, ..., \mu_T^T x_T^*) \qquad \bar{\mu}_{5\%} = \min(y \,|\, F(y) \geq 0.05)$$

Lastly, the portfolio forward and backward deviations are calculated as:

$$\bar{p_\mu} = \sup_{t>0} \frac{1}{t}\sqrt{2\ln(\frac{1}{T}\sum_{i=1}^{T}\exp(t(\mu_i^T x_i^* - \bar{\mu})))} \quad \text{and} \quad \bar{q_\mu} = \sup_{t>0} \frac{1}{t}\sqrt{2\ln(\frac{1}{T}\sum_{i=1}^{T}\exp(-t(\mu_i^T x_i^* - \bar{\mu})))}$$

6 Results

Numerical experiments are run with Python, using the Mosek© optimiser. Mosek also provides Fusion©, a high-level object-oriented interface that proves very useful to do fast model development.

6.1 Simulated data

In this section, focus is given to the factors exposures portfolio.

As a reminder:
- (M1) corresponds to the standard maximum exposure problem.
- (R1) corresponds to the robust equivalent with ellipsoidal uncertainty set.
- (R2) corresponds to the robust equivalent with ellipsoidal uncertainty set and net-zero alpha adjustment.
- (R3) corresponds to the robust deviation problem.

6.1.1 Experiment 1

Context
The first experiment compares the standard and robust problems for different numbers of assets. The results with turnover and long-only constraints and symmetric increments for μ are gathered in table 4. The same experiment was run, removing the turnover and long-only constraints and the results are provided in annex.
Both the long-only and long-short case represents plausible situations, while the case with unlimited turnover budget is more for expositional purposes.

Results

Table 4: Results of experiment 1

Long-only portfolio, $TO_{max} = 10\%$, $\sigma_{max} = 20\%$, $\rho = 70\%$, $\theta = [0, ..., 0]^T$, symmetric increments for μ

	Problem (M1)								Problem (R1)							
n	$\bar{\mu}$	$\bar{\sigma}$	$\bar{TO}$	$\bar{d}$	$\bar{\sigma_\mu}$	$\bar{\mu}_{min}$	$\bar{\mu}_{5\%}$	$\bar{\mu}_{max}$	$\bar{\mu}$	$\bar{\sigma}$	$\bar{TO}$	$\bar{d}$	$\bar{\sigma_\mu}$	$\bar{\mu}_{min}$	$\bar{\mu}_{5\%}$	$\bar{\mu}_{max}$
3	0.161	0.194	0.097	0.748	0.789	-2.802	-1.126	3.350	0.171	0.190	0.099	0.838	0.773	-2.706	-1.096	3.377
5	0.210	0.188	0.100	0.730	0.780	-2.776	-1.061	3.795	0.222	0.180	0.100	0.870	0.737	-2.594	-0.977	3.598
7	0.239	0.184	0.100	0.750	0.749	-2.862	-0.974	3.199	0.254	0.176	0.100	0.892	0.701	-2.580	-0.887	3.079
10	0.262	0.180	0.100	0.779	0.703	-2.565	-0.873	3.106	0.276	0.172	0.100	0.910	0.655	-2.374	-0.788	2.906
15	0.290	0.176	0.100	0.808	0.691	-2.814	-0.837	3.171	0.302	0.168	0.100	0.927	0.638	-2.484	-0.744	3.128
20	0.317	0.174	0.100	0.834	0.667	-2.220	-0.768	3.261	0.332	0.167	0.100	0.933	0.626	-2.114	-0.695	3.089
30	0.350	0.172	0.100	0.848	0.651	-2.331	-0.704	2.955	0.357	0.165	0.100	0.943	0.608	-2.113	-0.632	2.913
40	0.371	0.170	0.100	0.860	0.641	-2.100	-0.682	3.428	0.378	0.164	0.100	0.949	0.600	-1.901	-0.608	2.978

Comments
This first experience shows promising results for robust optimisation in the turnover-constrained, long-only case.

Indeed, the robust problem displays better average factor exposure $\bar{\mu}$ for a lower risk $\bar{\sigma}$. On top of this, the factor exposure is more stable (lower $\bar{\sigma_\mu}$), and subject to less extreme values as highlighted by $\bar{\mu}_{min}$, $\bar{\mu}_{5\%}$ and $\bar{\mu}_{max}$. This is combined with better diversification $\bar{d}$.

It is worthwhile noting that those properties are verified consistently for all assets (though the difference in average factor exposure tends to reduce with an increasing number of assets). Apart for a slight difference in the 3-asset case, the portfolios have same turnover (the turnover budget is entirely consumed).

However, the results are not as indisputable when removing the long-only or turnover constraint (tables 9 and 10). The difference in average factor exposure changes sign when varying the number of assets for the long-short case, while it is clearly in favour of the non-robust formulation when there is unlimited turnover budget. Consistently with the constrained case though, the robust portfolio remains more diversified (especially in the unlimited turnover case) and less risky. The factor exposure also varies less and it has less severe extreme negative values $\bar{\mu}_{min}$.

It seems that the robust formulation has unilaterally better performance in the long-only, turnover-constrained case, while in the unconstrained cases, robustness results in lower but more stable exposure. Additional tests around the turnover have shown that the effects of robustness are more visible when the constraint is tight. We argue that the robust formulation makes a better use of the turnover constraint.

6.1.2 Experiment 2

Context

In this case, the exposures paths are distorted and asymmetry is introduced. Some of the assets are skewed while others are not, and the experiment aims at checking whether a robust deviation formulation can adjust.

More precisely, we take skewed increments for $\lfloor \frac{n}{3} \rfloor$assets, with $\beta = 0.001$ while other $n - \lfloor \frac{n}{3} \rfloor$ assets have $\beta = 0.5$. Take $\omega = [-1, ..., -1]$ such that the skewed assets have a negative Pearson's skewness (indeed, there are rare large negative increments and frequent small positive increments).

There are long-only and turnover constraints, a fixed confidence level $\rho = 70\%$ and different numbers of assets.

Results

Table 5: Results of experiment 2

Long-only portfolio, $TO_{max} = 10\%$, $\sigma_{max} = 20\%$, $\rho = 70\%$, $\theta = [0, ..., 0]^T$

Asymmetric increments for $\lfloor \frac{n}{3} \rfloor$*assets, with* $\beta = 0.001$, $\omega = [-1, ..., -1]$

	Problem (M1)								Problem (R1)								Problem (R3)							
n	$\bar{\mu}$	$\bar{\sigma}$	$\bar{d}$	$\sigma_{\bar{\mu}}$	$\gamma_{\bar{\mu}}$	$q_{\bar{\mu}}$	$\bar{\mu}_{min}$	$\bar{\mu}_{5\%}$	$\bar{\mu}$	$\bar{\sigma}$	$\bar{d}$	$\sigma_{\bar{\mu}}$	$\gamma_{\bar{\mu}}$	$q_{\bar{\mu}}$	$\bar{\mu}_{min}$	$\bar{\mu}_{5\%}$	$\bar{\mu}$	$\bar{\sigma}$	$\bar{d}$	$\sigma_{\bar{\mu}}$	$\gamma_{\bar{\mu}}$	$q_{\bar{\mu}}$	$\bar{\mu}_{min}$	$\bar{\mu}_{5\%}$
3	0.114	0.193	0.764	0.758	-0.795	0.964	-3.482	-1.319	0.121	0.189	0.855	0.752	-0.913	1.058	-4.362	-1.331	0.125	0.194	0.746	0.770	-0.231	0.873	-3.183	-1.167
5	0.201	0.189	0.713	0.729	-0.507	0.868	-3.258	-1.112	0.216	0.181	0.860	0.704	-0.558	0.931	-3.805	-1.062	0.222	0.185	0.782	0.731	-0.100	0.809	-3.038	-0.982
7	0.230	0.185	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.252	0.176	0.894	0.657	-0.796	0.860	-3.106	-0.956	0.248	0.181	0.812	0.686	-0.386	0.782	-2.588	-0.944
10	0.236	0.181	0.762	0.688	-0.551	0.830	-3.006	-1.027	0.254	0.172	0.907	0.635	-0.730	0.807	-2.715	-0.933	0.240	0.176	0.848	0.671	-0.363	0.768	-2.619	-0.923
15	0.302	0.178	0.785	0.674	-0.367	0.802	-2.910	-0.853	0.301	0.169	0.921	0.632	-0.557	0.760	-2.675	-0.838	0.301	0.171	0.883	0.656	-0.377	0.748	-2.648	-0.833
20	0.293	0.174	0.826	0.671	-0.386	0.772	-2.800	-0.870	0.312	0.167	0.932	0.620	-0.534	0.747	-2.595	-0.787	0.311	0.169	0.908	0.638	-0.448	0.741	-2.566	-0.814
30	0.313	0.171	0.855	0.669	-0.386	0.768	-2.773	-0.865	0.320	0.165	0.945	0.624	-0.522	0.749	-2.576	-0.781	0.319	0.166	0.929	0.634	-0.474	0.747	-2.564	-0.800
40	0.354	0.171	0.856	0.657	-0.354	0.755	-2.568	-0.759	0.337	0.164	0.951	0.620	-0.510	0.739	-2.493	-0.751	0.341	0.165	0.939	0.625	-0.475	0.740	-2.498	-0.756

Comments

For all three portfolios and as expected, skewness in the asset-level exposures translates into portfolio-level skewness. Pearson's $\gamma_{\bar{\mu}}$ is negative, the backward deviation is higher than the standard deviation and the downside risk measures $\bar{\mu}_{min}$ and $\bar{\mu}_{5\%}$ are lower than in the case with symmetric increments (the comparison may not be entirely fair as the random numbers are not the same).

Alike in experiment 1, the robust portfolio are more diversified. There are quite a few discrepancies with experiments 1 though.

First, the comparison of average factor exposures is inconclusive (the robust formulations seem better for a low number of assets while they are not as good for a large number of assets). Also, the robust problem with ellipsoidal uncertainty set (R1) is more affected than the non-robust formulation by the skewness in asset-level exposures. Indeed, it shows worst backward deviation and worst-case exposures $\bar{\mu}_{min}$ and $\bar{\mu}_{5\%}$.

On the other hand, asymmetry is better dealt with by the robust deviation formulation (R3) (this is obvious with a low number of assets but it may be more arguable with many assets). It shows better worst-case exposures than both (M1) and (R1). The backward deviation is also lower for (R3) (though it has more negative skewness for a high number of assets).

6.1.3 Experiment 3

Context
In a third experiment, the number of assets is fixed to $n = 7$, and we compare all the formulations as a function of the confidence level ρ. On top of being the same for the different formulations, the random seed is fixed for the different confidence levels. The results are shown for both the symmetric and asymmetric increments (more precisely, in this case, $\lfloor \frac{n}{3} \rfloor$ assets are skewed, while the remaining $n - \lfloor \frac{n}{3} \rfloor$ assets have symmetric increments with $\beta = 0.5$).

The first four plots show the portfolio exposure $\bar{\mu}$, while the table 6 displays risk measures (this table has the results for the asymmetric case only, the results with symmetric increments may be found in annex, table 11). There are long-only and turnover constraints.

Results

Figure 7: Results of experiment 3 - Average exposure

Plot the value of $\bar{\mu}$ for varying confidence levels.

Long-only portfolio, $TO_{max} = 10\%$, $\sigma_{max} = 20\%$, $n = 7$, $\theta = [0, ..., 0]^T$

Top plots: symmetric increments

Bottom plots: asymmetric increments for $\lfloor \frac{n}{3} \rfloor$ assets, with $\beta = 0.001$. Remaining assets have $\beta = 0.5$.

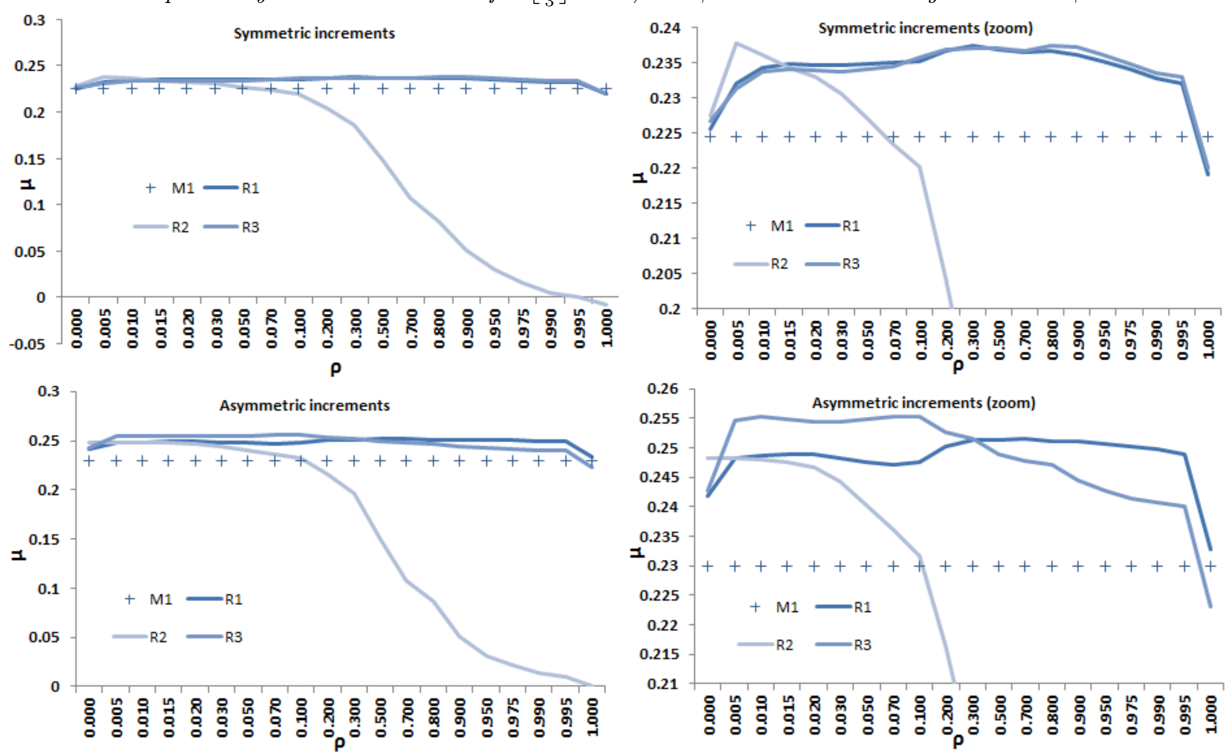


Table 6: Results of experiment 3 - Risk measures

Long-only portfolio, $TO_{max} = 10\%$, $\sigma_{max} = 20\%$, $n = 7$, $\theta = [0, ..., 0]^T$

Asymmetric increments for $\lfloor \frac{n}{3} \rfloor$ assets, with $\beta = 0.001$. Remaining assets have $\beta = 0.5$.

	Problem (M1)						Problem (R1)						Problem (R3)					
ρ	$\bar{d}$	$\bar{\sigma_\mu}$	$\bar{\gamma_\mu}$	$\bar{q_\mu}$	$\bar{\mu}_{min}$	$\bar{\mu}_{5\%}$	$\bar{d}$	$\bar{\sigma_\mu}$	$\bar{\gamma_\mu}$	$\bar{q_\mu}$	$\bar{\mu}_{min}$	$\bar{\mu}_{5\%}$	$\bar{d}$	$\bar{\sigma_\mu}$	$\bar{\gamma_\mu}$	$\bar{q_\mu}$	$\bar{\mu}_{min}$	$\bar{\mu}_{5\%}$
0.000	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.780	0.686	-0.614	0.836	-2.794	-0.983	0.744	0.694	-0.500	0.827	-2.802	-0.986
0.005	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.821	0.676	-0.674	0.853	-2.996	-0.961	0.767	0.691	-0.426	0.813	-2.803	-0.962
0.010	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.829	0.674	-0.684	0.854	-2.982	-0.962	0.770	0.690	-0.415	0.808	-2.772	-0.960
0.015	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.835	0.672	-0.690	0.853	-2.968	-0.963	0.772	0.689	-0.407	0.805	-2.756	-0.958
0.020	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.839	0.671	-0.692	0.852	-2.970	-0.962	0.774	0.688	-0.402	0.802	-2.745	-0.954
0.030	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.844	0.669	-0.692	0.849	-2.965	-0.961	0.777	0.688	-0.397	0.799	-2.724	-0.953
0.050	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.851	0.667	-0.688	0.844	-2.937	-0.960	0.780	0.687	-0.394	0.796	-2.695	-0.952
0.070	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.856	0.667	-0.690	0.843	-2.939	-0.961	0.782	0.687	-0.392	0.793	-2.678	-0.952
0.100	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.861	0.666	-0.697	0.846	-2.985	-0.957	0.786	0.687	-0.392	0.791	-2.657	-0.955
0.200	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.869	0.663	-0.718	0.849	-3.031	-0.958	0.793	0.686	-0.396	0.787	-2.620	-0.957
0.300	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.874	0.662	-0.725	0.850	-3.042	-0.958	0.797	0.686	-0.392	0.786	-2.608	-0.955
0.500	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.885	0.660	-0.758	0.855	-3.075	-0.957	0.804	0.686	-0.384	0.782	-2.596	-0.948
0.700	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.894	0.657	-0.796	0.860	-3.106	-0.956	0.812	0.686	-0.386	0.782	-2.588	-0.944
0.800	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.900	0.655	-0.814	0.860	-3.109	-0.956	0.817	0.685	-0.390	0.782	-2.584	-0.941
0.900	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.906	0.652	-0.831	0.860	-3.120	-0.955	0.824	0.685	-0.391	0.782	-2.572	-0.945
0.950	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.911	0.650	-0.846	0.860	-3.129	-0.959	0.829	0.685	-0.394	0.780	-2.564	-0.948
0.975	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.915	0.647	-0.858	0.860	-3.132	-0.963	0.833	0.683	-0.397	0.779	-2.561	-0.950
0.990	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.919	0.645	-0.873	0.860	-3.126	-0.965	0.837	0.683	-0.397	0.778	-2.559	-0.948
0.995	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.922	0.643	-0.884	0.859	-3.115	-0.965	0.840	0.683	-0.399	0.778	-2.557	-0.949
1.000	0.738	0.706	-0.590	0.846	-2.743	-1.029	0.950	0.630	-0.966	0.855	-3.092	-0.962	0.868	0.675	-0.436	0.773	-2.555	-0.963

Comments

First, it may be observed that all the robust formulations converge to the standard formulation (M1) when decreasing the confidence level. Indeed, the performance criteria are all in line with ρ near 0 (in fact, it is simply the consequence of all the portfolios being the same). Such convergence was expected as the uncertainty set shrinks to $\{\mu_0\}$ when $\rho = 0$. There is just a point estimate and all the formulations are equivalent.

Conversely, with increasing confidence level, the diversification increases for the robust portfolios. In particular, for the robust ellipsoid formulation (R1), the diversification tends to 1, the portfolio is almost equally-weighted. This is consistent with the theoretical results derived by Perchet *et al.* [43], as discussed in section 5.1. The solution to problem (R1) with $\rho = 1$ is the equally-weighted portfolio.

Another result that could have been expect is the equivalence of problems (R1) and (R3) in the case with symmetric increments (figure 7 and table 11). Equivalence between the two problems has been derived in section 4.4.4.

Now, looking at the results with intermediate confidence levels:

- Superior average factor exposure of the robust formulations (R1) and (R3) is confirmed, and this consistently happens for most of the confidence levels (except for a confidence level near 1). In the asymmetric case, the robust ellipsoid formulation has lower exposure for low confidence levels while it has higher exposure for high confidence levels.

- The net-zero alpha adjustment seems to fail in its stated purpose (reducing conservativeness). In fact, the problem (R2) appears to be more conservative than the other robust formulations and it tends quickly towards the equally-weighted portfolio. In this case, robustness comes at the price of a much lower average optimal value.

- In the asymmetric case, the (R3) formulation keeps the desirable properties that were already observed in experiment 2. It has better skewness and extreme values than both the (M1) and (R1) problems, as shown by the values of $\bar{\gamma_\mu}$, $\bar{q_\mu}, \bar{\mu}_{min}$ and $\bar{\mu}_{5\%}$. Increasing the confidence level for (R3) also results in extreme values $\bar{\mu}_{min}$ and $\bar{\mu}_{5\%}$ of lower magnitude (while this is quite the opposite for (R1)).

6.2 Market data

The results for the experiment on factor returns portfolios are gathered in table 7 and table 8. The first table displays the holding portfolios for the three formulations, while the second table shows some performance metrics (we slightly amend the definition for $\bar{\mu}$, $\bar{\sigma}$, $\bar{q}$, $\bar{p}$ to connect them to the out-of-sample daily portfolio returns).

As the experiment focuses on returns, the risk metrics are augmented with maximum drawdowns, Calmar and Sharpe ratios (respectively mean return to maximum drawdown and mean return to volatility). Also, in this context the 5th percentile $\bar{\mu}_{5\%}$ may be understood as a Value At Risk measure (VAR).

Figure 8 shows the out-of-sample cumulative return of the three portfolios.

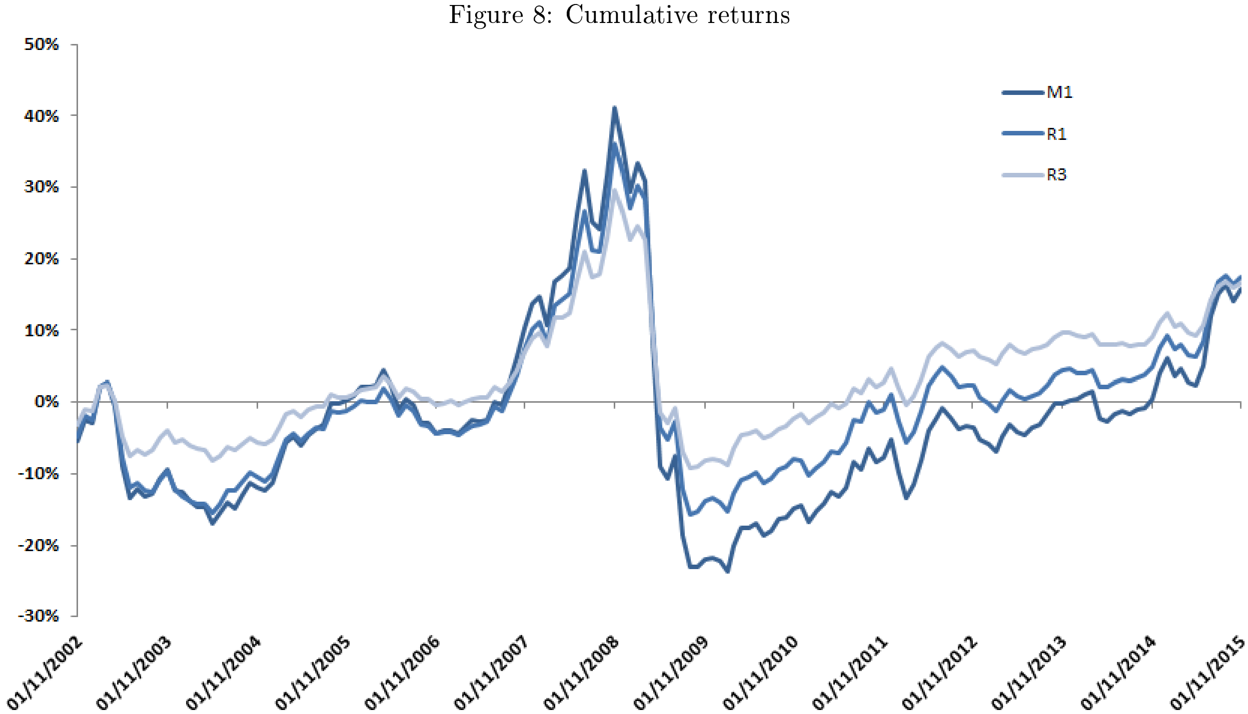
Table 7: Portfolios holdings

	SMB	HML	RMW	CMA	UMD
Problem (M1)	0.000	0.000	0.183	0.000	0.817
Problem (R1)	0.000	0.000	0.355	0.000	0.645
Problem (R3)	0.036	0.000	0.259	0.221	0.485

Table 8: Performance measures

	$\bar{\mu}$	$\bar{\sigma}$	Sharpe	$\bar{\mu}_{5\%}$	Drawdown	Calmar	$\bar{q}$	$\bar{p}$
Problem (M1)	0.012	0.126	0.098	-0.011	0.118	0.105	0.008	0.008
Problem (R1)	0.013	0.104	0.129	-0.009	0.097	0.139	0.007	0.007
Problem (R3)	0.013	0.079	0.163	-0.007	0.074	0.172	0.005	0.005

Figure 8: Cumulative returns



Comments
First, it should be observed that the optimal portfolios have significantly different holdings. A large portion of the non-robust portfolio is invested in the momentum factor whereas this portion is lower in the formulation (R1), and even further reduced in the robust deviation formulation (R3). The preponderance of the momentum factor in all the problems is simply explained by its elevated return in the test set (table 3). Both robust formulations are nonetheless more conservative: they give up some optimality by reducing the weight of the factor with maximum expected return and invest in a broader range of factors.

The robust deviation formulation down-weights the momentum as this factor shows the highest backward and forward deviations (in both the training and the test sets). As a result, the portfolio shows smoother returns during the 2008-2009 period, it then has the best Sharpe and Calmar ratios of all three portfolios. Further, it has lower (absolute) VAR, lower drawdown and lower backward deviation.

This experiment is subject to the weaknesses of testing with market data - as the out-of-sample period has seen quite different factor returns compared to the in-sample period. A cynical interpretation would conclude that robust optimisation performs well simply because the realised parameters have been closer to the worst case than to the average. Still, the experiment unarguably shows that robustness provides more diversified, lower-risk portfolios, and that the robust deviation formulation grants solutions less affected by distributional asymmetry.

Conclusion

This dissertation has primarily focused on integrating the robust optimisation framework into the construction of factor portfolios:

- We reviewed the concept of risk premium and detailed the economic rationale underpinning the existence of common equity risk premia.
- We presented different approaches to capture the factors excess returns, namely a bottom-up approach that focuses on factor exposures and a top-down approach that focuses on factor returns.
- We showed that the portfolio construction may benefit from robust optimisation as both exposures and expected returns are subject to estimation error and/or time-varying behaviours.
- We introduced different robust formulations and derived equivalent Second-Order Cone Programs for each (ellipsoidal uncertainty set, net-zero alpha adjustment and robust deviation).
- We compared the performance of the non-robust and robust problems on both simulated data (for the factors exposure problem) and market data (for the factor returns problem).

Our first contribution is therefore a consistent approach from the economics to the mathematical modelling of robust factor portfolios.

Our second contribution is a set of empirical results, some of which validate the theoretical features of robust optimisation, while others are less expected:

- in all the situations that we encountered, robustness grants more diversified, less risky portfolios.
- in some situations (*e.g.* with long-only constraint), robustness allows for a better use of a turnover constraint thus providing better average factor exposure. This is a direct effect of the enhanced solution stability (with respect to changes in the input).
- the net-zero alpha adjustment fails to provide less conservative robust solutions.
- the robust deviation proves capable of handling distributional asymmetry (as seen with both the factor exposures and the factor returns problems). It properly down-weights assets which show large backward deviation, hence reducing the overall asymmetry in the optimal portfolio.
- the confidence level controlling the uncertainty set's size is critical for the success or failure of robust optimisation.

It is quite impressive to see that a worst-case approach such as robust optimisation may actually push the efficient frontier upward (in the long-only, turnover-constrained case, the optimal portfolio has both higher return and lower risk). However, robust optimisation may not be advised blindly: we have seen that robustness comes at the price of a lower objective value in the case with unlimited turnover. Also, for high confidence levels (combined with a particular uncertainty set structure) the robust problem simply yields the equally-weighted portfolio. Researchers have shown that this portfolio has desirable properties, but this makes the investor's views on expected returns redundant and may then be problematic. While observing that the confidence level controlling the size of an uncertainty set is critical, we may not give an absolute number as to what the right level is. It seems that it should be fine-tuned on a case-by-case basis.

There are quite a few directions in which this work could be extended. First, the robust equivalents of the minimum risk (M2) and maximum Sharpe ratio (M4) problems could be considered. In this perspective, we could also allow for uncertainty in the risk model, and constraint-wise uncertainty (to have robust control over undesirable factor exposures).

Another direction could focus on different modellings for the exposures paths. We chose a multivariate mean-revering process as what seemed the most appropriate. Further experiments could take different mean-reversion speeds or different long-term levels for different assets. It would also be interesting to look a the results with different processes, *e.g.* random walks or trending processes.

Lastly, the robust modelling could be enriched with more granular transaction cost models (replacing the turnover constraint), or multi-period formulations. As it would include a view on the future values of factors, it is likely that multi-period optimisation supplements robustness and provides additional benefit on the usage of a trading-cost budget.

Annexes

A.1 Karush-Kuhn-Tucker (KKT) conditions

The KKT conditions consist in a set of first-order necessary conditions such that the solution to a constrained non-linear program is optimal. Those conditions are detailed below, following Cornuejols and Tütüncü [18].

Consider the problem:

$$\begin{array}{rll} \min\limits_{x \in \mathcal{R}^n} & f(x) & \\ \text{s.t.} & g_i(x) = 0 & i \in \mathcal{E} \\ & g_i(x) \geq 0 & i \in \mathcal{I} \end{array} \tag{P}$$

Assume that f and g_i, $i \in \mathcal{E} \cup \mathcal{I}$ are continuously differentiable functions.

Definition A.1. ***Regular point***
Let x belong to the feasible set of problem (P), with $\mathcal{J}$ the subset of $\mathcal{I}$ such that $g_i(x) = 0$.
Then, x is called a regular point if the gradients $\nabla g_i(x)$ are linearly independent for $i \in \mathcal{E} \cup \mathcal{J}$.

Theorem A.2. ***First-order KKT conditions***
Let x^*be a local minimiser for problem (P), and assume that x^* is a regular point.
Then, for $i \in \mathcal{E} \cup \mathcal{I}$, there exists λ_i that verify:

$$\nabla f(x^*) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i \nabla g_i(x^*) = 0$$

$$\forall i \in \mathcal{I} \quad \lambda_i \geq 0 \quad \textit{and} \quad \lambda_i g_i(x^*) = 0$$

A.2 Least-squares with equality constraints

The problem of interest is:

$$\begin{aligned} \min_{f \in \mathcal{R}^k} \quad & \|Af - b\|^2 \\ \text{s.t.} \quad & Cf = d \end{aligned} \tag{CLS}$$

There are n observations and p constraints, the decision variable f is of size k.
It is assumed that $n \geq p \geq k$ - which simply means in our context that there are more assets than factors, and more factors than constraints. In our case, there are only two constraints, so this should not be an issue. We also assume that the D matrix has full rank (no redundant constraint).

Take the QR decomposition for C^T:

$$Q^T C^T = \begin{bmatrix} R \\ 0 \end{bmatrix}$$

R is a size $p \times p$ upper triangular matrix.

We let:

$$AQ = \begin{bmatrix} A_1 & A_2 \end{bmatrix} \qquad \text{and} \qquad Q^T f = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$$

The split has been done such that:
A_1 has size $n \times p$.
A_2 has size $n \times (k - p)$.
f_1 has size p.
f_2 has size $k - p$.

Thus, we have:

$$Cf = \begin{bmatrix} R \\ 0 \end{bmatrix}^T Q^T f = R^T f_1 \qquad \text{and} \qquad Af = AQQ^T f = A_1 f_1 + A_2 f_2$$

The problem (CLS) is then equivalent to:

$$\begin{aligned} \min_{f_1, f_2} \quad & \|A_1 f_1 + A_2 f_2 - b\|^2 \\ \text{s.t.} \quad & R^T f_1 = d \end{aligned} \tag{CLS'}$$

As f_1 is uniquely determined by the equality $R^T f_1 = d$ (there are p unknown and p equalities), we can first solve for f_1. This gets us an unconstrained problem:

$$\min_{f_2} \quad \|A_1 f_1 + A_2 f_2 - b\|^2 \tag{CLS''}$$

The solution for (CLS”) is given by:

$$f_2 = (A_2^T A_2)^{-1} A_2^T (b - A_1 f_1)$$

And we finally reconstruct f via:

$$f = Q \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$$

□

A.3 Proof for lemma (4.1)

Though we only provide the proof for the Euclidean norm, lemma (4.1) actually remains valid for a broader class of norms (called *absolute norms* by Chen *et al.*).
We have:

$$\Omega_{max}\|u\| = \Omega_{max}\|u\|^* = \begin{array}{cc} \max\limits_{r} & \sum\limits_{j=1}^{n} \max\{a_j, b_j, 0\} r_j \\ \text{s.t.} & \|r\| \leq \Omega_{max} \end{array} \tag{23}$$

Both the self-duality of the Euclidean norm and the definition of the dual norm have been used.

Now, suppose r^* is an optimal solution to (23).

Let:

$$\begin{cases} y_j = w_j = 0 & \text{if} \quad \max\{a_j, b_j\} = 0 \\ y_j = |r_j^*|, \;\; w_j = 0 & \text{if} \quad a_j \geq b_j, \;\; a_j > 0 \qquad \forall j \in \{1, ..., n\} \\ w_j = |r_j^*|, \;\; y_j = 0 & \text{if} \quad a_j < b_j, \;\; b_j > 0 \end{cases}$$

With those settings for y and w, we have:

$$a_j y_j + b_j w_j \geq \max\{a_j, b_j, 0\}\, r_j^*$$

$$y_j + w_j \leq |r_j^*|$$

Hence $\|y_j + w_j\| \leq \|r_j^*\|, \;\; y \geq 0, \;\; w \geq 0$ and y, w are feasible to problem (DevMax).
Then:

$$f^* \geq \sum_{j=1}^{n} a_j y_j + b_j w_j \geq \sum_{j=1}^{n} \max\{a_j, b_j, 0\}\, r_j^* = \Omega_{max}\|u\|$$

Conversely, for (y^*, w^*) solution to problem (DevMax), define $r = y^* + w^*$.

By construction, $\|r\| \leq \Omega$. Also, we have:

$$r_j \, \max\{a_j, b_j, 0\} \geq a_j y_j^* + b_j w_j^* \quad j = 1, ..., n$$

As a result:

$$\Omega_{max}\|u\| \geq \sum_{j=1}^{n} \max\{a_j, b_j, 0\}\, r_j \geq \sum_{j=1}^{n} a_j y_j^* + b_j w_j^* = f^*$$

$\square$

A.4 Proof for proposition (4.3)

(i) We show that $p \geq \sigma$, the proof for $q \geq 0$ can be derived similarly.

A Taylor expansion near 0 gives:

$$E[\exp(t\zeta)] \underset{t \to 0}{\sim} E(1 + t\zeta + \frac{t^2\zeta^2}{2} + \frac{t^3\zeta^3}{6})$$

Using the fact that ζ is centred and doing another expansion for the logarithm:

$$\ln E[\exp(t\zeta)] \underset{t \to 0}{\sim} \frac{t^2 E(\zeta^2)}{2} + \frac{t^3 E(\zeta^3)}{6}$$

This yields:

$$\frac{1}{t}\sqrt{2\ln(E[\exp(t\zeta)])} \underset{t \to 0}{\sim} \sqrt{E(\zeta^2) + \frac{tE(\zeta^3)}{3}} \underset{t \to 0}{\longrightarrow} \sigma$$

Hence:

$$p = \sup_{t>0} \frac{1}{t}\sqrt{2\ln(E[\exp(t\zeta)])} \geq \sigma$$

(ii) With $\zeta \sim \mathcal{N}(0, \sigma)$, the moment-generating function is:

$$E[\exp(t\zeta)] = \exp(\frac{t^2\sigma^2}{2})$$

Thus

$$p = \sup_{t>0} \frac{1}{t}\sqrt{2\ln(E[\exp(t\zeta)])} = \sup_{t>0} \sigma = \sigma$$

$\square$

A.5 Additional results

A.5.1 Experiment 1

Table 9: Results of experiment 1 - No long-only constraint

Long-short portfolio, $TO_{max} = 10\%$, $\sigma_{max} = 20\%$, $\rho = 70\%$, $\theta = [0, ..., 0]^T$, symmetric increments for μ

	Problem (M1)								Problem (R1)							
n	$\bar{\mu}$	$\bar{\sigma}$	$\bar{TO}$	$\bar{d}$	$\bar{\sigma_{\mu}}$	$\bar{\mu}_{min}$	$\bar{\mu}_{5\%}$	$\bar{\mu}_{max}$	$\bar{\mu}$	$\bar{\sigma}$	$\bar{TO}$	$\bar{d}$	$\bar{\sigma_{\mu}}$	$\bar{\mu}_{min}$	$\bar{\mu}_{5\%}$	$\bar{\mu}_{max}$
3	0.158	0.195	0.098	0.774	0.802	-2.706	-1.165	3.301	0.165	0.190	0.099	0.850	0.779	-2.686	-1.117	3.301
5	0.249	0.196	0.100	0.869	0.816	-3.079	-1.061	3.720	0.245	0.184	0.100	0.896	0.751	-2.735	-0.971	3.535
7	0.264	0.196	0.100	0.912	0.787	-2.794	-1.028	3.788	0.269	0.179	0.100	0.924	0.695	-2.655	-0.869	3.254
10	0.308	0.197	0.100	0.941	0.800	-2.923	-1.003	3.663	0.315	0.176	0.100	0.945	0.683	-2.524	-0.798	3.225
15	0.324	0.198	0.100	0.962	0.806	-2.762	-0.993	3.980	0.336	0.173	0.100	0.963	0.657	-2.239	-0.730	3.613
20	0.400	0.199	0.100	0.973	0.830	-2.693	-0.958	3.996	0.379	0.172	0.100	0.972	0.670	-2.415	-0.719	3.645
30	0.447	0.199	0.100	0.981	0.847	-3.472	-0.930	4.060	0.417	0.170	0.100	0.981	0.672	-2.416	-0.665	3.291
40	0.467	0.199	0.100	0.986	0.829	-2.735	-0.900	4.005	0.446	0.169	0.100	0.985	0.646	-2.253	-0.630	2.932

Table 10: Results of experiment 1 - No turnover constraint

Long-only portfolio, $TO_{max} = \infty$, $\sigma_{max} = 20\%$, $\rho = 70\%$, $\theta = [0, ..., 0]^T$, symmetric increments for μ

	Problem (M1)								Problem (R1)							
n	$\bar{\mu}$	$\bar{\sigma}$	$\bar{TO}$	$\bar{d}$	$\bar{\sigma_{\mu}}$	$\bar{\mu}_{min}$	$\bar{\mu}_{5\%}$	$\bar{\mu}_{max}$	$\bar{\mu}$	$\bar{\sigma}$	$\bar{TO}$	$\bar{d}$	$\bar{\sigma_{\mu}}$	$\bar{\mu}_{min}$	$\bar{\mu}_{5\%}$	$\bar{\mu}_{max}$
3	0.422	0.200	0.516	0.600	0.769	-2.584	-0.850	3.657	0.404	0.196	0.475	0.688	0.767	-2.628	-0.857	3.657
5	0.662	0.200	0.789	0.500	0.765	-3.034	-0.602	4.197	0.610	0.191	0.667	0.671	0.758	-2.873	-0.627	4.197
7	0.803	0.200	0.902	0.467	0.766	-2.741	-0.462	4.196	0.711	0.187	0.739	0.700	0.756	-2.694	-0.529	4.196
10	0.931	0.200	1.023	0.444	0.749	-2.421	-0.296	4.142	0.791	0.182	0.810	0.744	0.728	-2.258	-0.376	4.142
15	1.081	0.200	1.150	0.429	0.730	-1.923	-0.139	3.982	0.886	0.178	0.862	0.791	0.694	-1.897	-0.250	3.765
20	1.183	0.200	1.187	0.421	0.726	-2.026	-0.013	4.331	0.941	0.175	0.887	0.822	0.676	-1.873	-0.148	4.215
30	1.295	0.200	1.293	0.414	0.729	-1.935	0.096	4.446	0.996	0.171	0.942	0.859	0.661	-1.927	-0.074	3.932
40	1.390	0.200	1.342	0.410	0.722	-1.852	0.202	4.768	1.052	0.169	0.969	0.879	0.640	-1.643	0.017	4.007

A.5.2 Experiment 3

Table 11: Results of experiment 3 (Symmetric increments) - Risk measures

Long-only portfolio, $TO_{max} = 10\%$, $\sigma_{max} = 20\%$, $n = 7$, $\theta = [0, ..., 0]^T$

	Problem (M1)						Problem (R1)						Problem (R3)					
ρ	$\bar{d}$	$\bar{\sigma_\mu}$	$\bar{\gamma_\mu}$	$\bar{q_\mu}$	$\bar{\mu}_{min}$	$\bar{\mu}_{5\%}$	$\bar{d}$	$\bar{\sigma_\mu}$	$\bar{\gamma_\mu}$	$\bar{q_\mu}$	$\bar{\mu}_{min}$	$\bar{\mu}_{5\%}$	$\bar{d}$	$\bar{\sigma_\mu}$	$\bar{\gamma_\mu}$	$\bar{q_\mu}$	$\bar{\mu}_{min}$	$\bar{\mu}_{5\%}$
0.000	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.795	0.746	-0.007	0.757	-2.633	-1.031	0.793	0.745	-0.009	0.755	-2.592	-1.030
0.005	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.830	0.729	0.031	0.733	-2.475	-0.987	0.828	0.729	0.030	0.732	-2.445	-0.983
0.010	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.834	0.728	0.040	0.730	-2.427	-0.982	0.833	0.728	0.040	0.729	-2.395	-0.978
0.015	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.838	0.727	0.041	0.728	-2.402	-0.980	0.836	0.727	0.044	0.728	-2.371	-0.977
0.020	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.840	0.726	0.040	0.727	-2.398	-0.977	0.838	0.726	0.045	0.727	-2.357	-0.976
0.030	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.844	0.723	0.037	0.725	-2.406	-0.972	0.842	0.724	0.042	0.724	-2.359	-0.974
0.050	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.849	0.719	0.037	0.721	-2.410	-0.966	0.847	0.721	0.042	0.721	-2.370	-0.966
0.070	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.854	0.717	0.036	0.719	-2.410	-0.958	0.851	0.718	0.040	0.719	-2.379	-0.960
0.100	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.859	0.714	0.029	0.717	-2.422	-0.954	0.856	0.716	0.038	0.717	-2.392	-0.954
0.200	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.868	0.711	0.008	0.718	-2.476	-0.946	0.866	0.712	0.023	0.715	-2.417	-0.945
0.300	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.874	0.710	-0.001	0.720	-2.519	-0.942	0.872	0.710	0.014	0.715	-2.446	-0.939
0.500	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.882	0.709	-0.001	0.720	-2.549	-0.938	0.881	0.708	0.009	0.716	-2.493	-0.936
0.700	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.892	0.707	0.005	0.716	-2.519	-0.931	0.890	0.706	0.007	0.715	-2.490	-0.931
0.800	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.897	0.705	0.009	0.712	-2.485	-0.928	0.895	0.705	0.009	0.711	-2.451	-0.926
0.900	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.905	0.702	0.013	0.708	-2.459	-0.922	0.902	0.703	0.013	0.708	-2.418	-0.920
0.950	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.911	0.700	0.014	0.706	-2.448	-0.917	0.908	0.701	0.016	0.704	-2.378	-0.918
0.975	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.915	0.699	0.015	0.704	-2.423	-0.916	0.913	0.699	0.017	0.702	-2.359	-0.916
0.990	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.920	0.697	0.018	0.701	-2.397	-0.914	0.918	0.697	0.019	0.700	-2.334	-0.914
0.995	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.924	0.696	0.020	0.699	-2.381	-0.914	0.921	0.696	0.021	0.698	-2.328	-0.912
1.000	0.761	0.757	-0.063	0.800	-3.002	-1.047	0.951	0.685	0.045	0.685	-2.235	-0.910	0.948	0.685	0.044	0.685	-2.204	-0.907

References

[1] Arnott, R. D., Bernstein, P. L., *What risk premium is "normal"* ? Journal of Portfolio Management, 2002.

[2] Artzner P., Delbaen F., Eber J.M., Heath D., *Coherent measures of risk*, Mathematical Finance, Vol. 9 No. 3, 1999.

[3] Ben-Tal A., Nemirovski, A., *Robust convex optimization,* Mathematics of Operation research, Vol. 23 No. 4, 1998.

[4] Ben-Tal A., Nemirovski, A., *Robust solutions of uncertain linear programs,* Operations Research Letters, Vol. 25 No. 1, 1999.

[5] Ben-Tal A., Margelit T., Nemirovski, A., *Robust modeling of multi-stage portfolio problems*, Technion Israel Institute of Technology, 2000.

[6] Ben-Tal A., El Ghaoui, L., Nemirovski, A., *Robust Optimization*, Princeton University Press, 2009.

[7] Bender, J., Briand, R., Melas, D., Subramanian, R. A., *Foundation of factor investing,* MSCI, 2013.

[8] Bertsimas, D., Thiele, A., *Robust and data-driven optimization: Modern decision-making under uncertainty*, Massachusetts Institute of Technology, 2006.

[9] Bertsimas D., Pachamanova D., *Robust multiperiod portfolio management in the presence of transaction costs*, Computer and Operations Research 35, 2008.

[10] Bertsimas, D., Brown, D.B., *Constructing uncertainty sets for robust linear optimization,* Operations Research, Vol. 57 No.6, 2009.

[11] Bertsimas D., Brown D.B., Caramanis, C., *Theory and applications of robust optimization*, SIAM Review, Vol. 53, No. 3, pp. 464-501, 2010.

[12] Bertsimas, D., Gupta, V., Kallus, N., *Data-driven robust optimization*, Operation Research, 2013.

[13] Back, F., Litterman, R., *Asset allocation: combining investor views with market equilibrium*, Goldman Sachs Fixed Income Research Note, 1990.

[14] Ceria, S., Stubbs, R., *Incorporating estimation errors into portfolio selection: robust portfolio construction*, Axioma Research Paper No. 003, 2006.

[15] Chen, X., Sim, M., Sun, P., *A robust optimization perspective on stochastic programming*. Operation Research, Vol. 55 No. 6, 2007.

[16] Cho, W. N., *Robust portfolio optimization using conditional value at risk*, Imperial College London, 2008.

[17] Cont, R., *Empirical properties of asset returns: stylized facts and statistical issues*, Quantitative Finance, Vol. 1, 2001.

[18] Cornuejols G., Tütüncü R., *Optimization methods in finance*, Carnegie Mellon University, Pittsburgh, 2006.

[19] El Ghaoui, L., Lebret, H., *Robust solutions to least-square problems with uncertain data,* SIAM Journal, Vol. 18 No. 4, 1997.

[20] Fama, E. F., French, K. R., *The cross-section of expected stock returns*, The Journal of Finance, Vol. 47 No. 2, 1992

[21] Fama, E. F., French, K. R., *Common risk factors in the returns on stocks and bonds,* Journal of Financial Economics, Vol. 33, 1993.

[22] Fama, E. F., French, K. R., *The Capital Asset Pricing Model: theory and evidence*, Journal of Economic Perspectives, Vol. 18 No. 3, 2004.

[23] Fama, E. F., French, K. R., *A five-factor asset pricing model, 2014.*

[24] Fama, E. F., French, K. R., Data Library *http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html,* 2016.

[25] Gârleanu, N., Pedersen, L. H., *Dynamic trading with predictable returns and transaction costs,* The Journal of Finance, Vol. 68. No. 6, 2013.

[26] Gatignon, H., *Statistical analysis of management data*, Springer, 2010.

[27] Goldfarb, D., Iyengar, G., *Robust portfolio selection problems,* Mathematics of Operation research, Vol. 28, 2003.

[28] Hauser, R., *Lecture notes for the optimisation module*, 2013.

[29] Hauser, R., Krishnamurthy, V., Tütüncü R., *Relative robust portfolio optimization*, 2013.

[30] Jegadeesh, N., Titman, S., *Returns to buying winners and selling losers: implications for stock market efficiency*, Journal of Finance, Vol. 48 No. 1, 1993.

[31] Kim, W. C., Kim, J. H., Ahn, S. H., Fabozzi, F. J., *What do robust equity portfolio models really do?*, Annals of Operations Research, Vol. 205 Issue 1, 2012.

[32] Kouvelis, P., Yu, G., *Robust discrete optimization and its applications.* Kluwer Academic Publishers, Dordtecht, The Netherlands, 1997.

[33] Lempérière, Y., Deremble C., Nguyen, T. T., Seager, P., Potters, M., Bouchaud, J. P, *Risk premia: asymmetric tail risks and excess returns,* Capital Fund Management. 2014.

[34] Lintner, J., *The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets,* The Review if Economics and Statistics, Vol. 47 No. 1, 1965.

[35] Markowitz, H. M., *Portfolio selection,* The Journal of Finance Vol. 7 No. 1, 1952.

[36] Menchero, J., *Characteristics of factor portfolios,* MSCI Barra, 2010.

[37] Michaud, R. O., *The Markowitz optimization enigma: is 'optimized' optimal?* Financial Analysts Journal, Vol. 45, 1989.

[38] Michaud, R. O., *Efficient asset management*, Harvard Business School Press. 1998

[39] Mossin, J., *Equilibrium in a capital asset market*, Econometrica, Vol. 34, No. 4, 1966.

[40] Natarajan, K., Pachamanova, D., Sim, M., *A tractable parametric approach to Value-at-Risk optimization*, Risk Management Institute, RMI Working Paper No. 06/16, 2006.

[41] Natarajan, K., Pachamanova, D., Sim, M., *Constructing risk measures from uncertainty sets*, Operations Research, Vol. 57 No.5, 2009.

[42] Pastor, L., Stambaugh, R. F., *Liquidity risk and expected stock returns*, The Journal of Political Economy, Vol. 111, 2003.

[43] Perchet, R., Xiao, L., Leote de Carvalho, R., Heckel, T. *Insights into robust portfolio optimization: decomposing robust portfolios into mean-variance and risk-based portfolios, 2015.*

[44] Renshaw, A., *Real world case studies in portfolio construction using robust optimization,* Axioma Research Paper No. 007, 2008.

[45] Ross, S. A., *The arbitrage theory of capital asset pricing,* Journal of Economic Theory, Vol. 13, 1976.

[46] Scherer B., *Can robust portfolio optimisation help to build better portfolios?*, Journal of Asset Management, Vol. 7, 2007.

[47] Sharpe, W. F., *Capital asset prices: a theory of market equilibrium under conditions of risk*, The Journal of Finance, Vol. 19 No. 3, 1964.

[48] Sivaramakrishnan, K., Jeet, V., Vandenbussche D., *Multi-period portfolio optimization with alpha decay*, Axioma Research Paper, No. 055, 2015.

[49] Soyster, A. L., *Convex programming with set-inclusive constraints and applications to inexact linear programming*, Operations Research, Vol. 21 No. 5, 1973.

[50] Stubbs, R. A., Vance, P., *Computing return estimation error matrices for robust optimization*, Axioma Research Paper No. 001, 2005.

[51] Treynor, J. L., *Market value, time and risk,* 1961.

[52] Welch, I., Goyal, A., *A comprehensive look at the empirical performance of equity premium prediction,* Review of Financial Studies, Vol. 21 No. 4, 2008.