ROBUST PORTFOLIO OPTIMIZATION: A STUDY OF BSE30 AND BSE100

A Project Report Submitted

for the Course

MA498 Project I

by

Mohammed Bilal Girach

(Roll No. 150123024)

Shashank Oberoi

(Roll No. 150123047)

to the

DEPARTMENT OF MATHEMATICS

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI - 781039, INDIA

November 2018

CERTIFICATE

This is to certify that the work contained in this project report entitled "**Robust Portfolio Optimization: A Study of BSE30 and BSE100**" submitted by **Mohammed Bilal Girach (Roll No. 150123024)** and **Shashank Oberoi (Roll No. 150123047)** to the Department of Mathematics, Indian Institute of Technology Guwahati towards partial requirement of **Bachelor of Technology** in Mathematics and Computing has been carried out by them under my supervision.

It is also certified that, along with literature survey, a few new results are established using computational implementations carried out by the students under the project.

Turnitin Similarity: 10 %

Guwahati - 781 039 (Dr. Siddhartha Pratim Chakrabarty)

November 2018 Project Supervisor

ABSTRACT

We begin with a discussion on the classical Markowitz portfolio optimization, its drawbacks and consequent motivation of the alternate approach of robust portfolio optimization. This is followed by presenting several robust optimization models. Using uncertainty sets, we then present computational results with BSE30 and BSE100 followed by a simulation study using true mean and covariance of asset returns. We undertake a comparison of performance of the robust optimization approaches as compared to Markowitz optimization. We finally discuss the advantages of the robust optimization from the standpoint of number of stocks, number of samples and types of data.

Contents

List of Figures **v**

List of Tables **vi**

1 Introduction **1**

2 Robust Portfolio Optimization Models using Uncertainty Sets **5**

2.1 Uncertainty Sets 7

2.1.1 Uncertainty in Expected Returns 8

2.1.2 Separable Uncertainty Set 10

2.1.3 Joint Uncertainty Set 11

3 Computational Results **13**

3.1 Performance with $N = 31$ assets 15

3.2 Performance with $N = 98$ assets 19

4 Conclusions and Comments **23**

4.1 From the Standpoint of Number of Stocks 23

4.2 From the Standpoint of Number of Samples Generated 25

4.3 From the Standpoint of Type of the Data 26

Bibliography **27**

List of Figures

3.1 Efficient Frontier plot and Sharpe ratio plot for different portfolio optimization models in case of Simulated Data with 1000 samples (31 assets) 16

3.2 Efficient Frontier plot and Sharpe ratio plot for different portfolio optimization models in case of Simulated Data with same number of samples as market data (31 assets) 17

3.3 Efficient Frontier plot and Sharpe ratio plot for different portfolio optimization models in case of Market Data (31 assets) . 18

3.4 Efficient Frontier plot and Sharpe ratio plot for different portfolio optimization models in case of Simulated Data with 1000 samples (98 assets) 20

3.5 Efficient Frontier plot and Sharpe ratio plot for different portfolio optimization models in case of Simulated Data with same number of samples as market data (98 assets) 21

3.6 Efficient Frontier plot and Sharpe ratio plot for different portfolio optimization models in case of Market Data (98 assets) . 22

List of Tables

3.1 Comparison of different portfolio optimization models in case of Simulated Data with 1000 samples (31 assets) 16
3.2 Comparison of different portfolio optimization models in case of Simulated Data with same number of samples as market data (31 assets) 17
3.3 Comparison of different portfolio optimization models in case of Market Data (31 assets) 19
3.4 Comparison of different portfolio optimization models in case of Simulated Data with 1000 samples (98 assets) 19
3.5 Comparison of different portfolio optimization models in case of Simulated Data with same number of samples as market data (98 assets) 21
3.6 Comparison of different portfolio optimization models in case of Market Data (98 assets) 22

4.1 The maximum average Sharpe ratio compared by varying the number of stocks in different kinds of scenarios. 24
4.2 The maximum average Sharpe ratio compared by varying the number of stocks in different kinds of scenarios. 25

4.3 The maximum average Sharpe ratio compared by varying the type of the data in different kinds of scenarios. 26

Chapter 1

Introduction

Investment in an individual security always has an associated risk, which can be minimized through diversification, a process involving investment in a portfolio consisting of several securities. In accordance with this notion, Markowitz [20, 21] introduced the mean-variance model for optimal allocation of weights to the securities comprising the portfolio. Mean and covariance matrix of returns of securities are used as the measures for giving a quantitative sense to the return and the risk, respectively, of the portfolio. Despite the theoretical simplicity of the approach of Markowitz model based on risk-return trade-off, there are several drawbacks associated with incorporating it in a practical setup.

Theoretically, Markowitz based portfolio optimization can result in assigning extreme weights to the securities comprising the portfolio. However, investment in securities can not be made in such extreme positions like large short positions if one takes active trading into account. Such kind of scenarios can be avoided by introducing appropriate constraints on the weights. Black and Litterman [9] argued that there is an added disadvantage since there are high chances of the optimal portfolio lying in the neighborhood of the imposed constraints. Thus, imposition of constraints leads to strong dependence of the constructed portfolio upon the constraints. For example, disallowing short sales often results in assigning zero weights to many securities and largely positive weights to the securities having small market capitalization.

One of the most significant demerits of the mean-variance model is the sensitivity issue associated with risk and return parameters of the individual securities of a portfolio. These parameters are estimated using mean and standard deviation of returns. While computing these estimates, historical data of returns is usually taken into account in order to calculate the sample mean and the sample variance. The historical data neglects various other market factors and is not an accurate representation for estimates of future returns. Taking into account the above reasons, Michaud [22] argued that the mean-variance analysis tends to maximize the impact of estimation errors associated with the return and the risk parameters for the securities. As a result, Markowitz portfolio optimization often overweighs (underweighs) the securities having higher (lower) expected return, lower (higher) variance of returns and negative (positive) correlation between their returns. Labelling the model as "estimation-error maximizers", he stated that it often leads to financially counter-intuitive portfolios, which, in some cases, perform worse than the equal-weighted portfolio. Broadie [10] investigated the error maximization property of mean-variance analysis. Accordingly, he conducted a simulation based study to compare the estimated efficient frontier with the actual frontier computed using true parameter values. He observed that points on the estimated efficient frontier show superior performance as compared to the corresponding points on the actual frontier. He supported his argument of over-estimation of expected returns of optimal portfolios through his simulated results of obtaining the estimated frontier lying above the actual frontier. Additionally, he pointed out that non-stationarity in the data of returns can further increase the errors in computing the efficient frontier. Chopra and Ziemba [12] performed the sensitivity analysis of performance of optimal portfolios by studying the relative effect of estimation errors in means, variances and covariances of security returns, taking the investors' risk tolerance into consideration as well. They observed that at a high risk tolerance (to be defined in later chapters) of around fifty, cash equivalent loss for estimation errors in means is about eleven times greater than that for errors in variances or covariances. Accordingly, they pointed out that if the investors have superior estimates for means of security returns, they should prefer using them over the sample means calculated from historical data. Best and Grauer [6, 7] also arrived at similar conclusions by studying the sensitivity of weights of optimal portfolios with respect to changes in estimated means of returns on individual securities. Further, on imposition of no short selling constraint on the securities, they observed that a small change in estimated mean return of an individual security can assign zero weights to almost half the securities comprising the portfolio.

The discussed literature arrives at a common conclusion that the optimal portfolios are extremely sensitive towards the estimated values of input parameters, particularly expected returns of individual securities. In order to address this issue, there has been significant progress in recent years in the area of robust portfolio optimization. Several methods have been proposed in this area. **We are particularly interested in the approaches falling in the category related to enhancing robustness by optimizing the portfolio performance in worst-case scenarios.**

Significant efforts have been made towards formulating these kinds of approaches from Markowitz based mean-variance analysis. However, no rigorous framework has been observed that could establish these approaches at par with Markowitz optimization in terms of portfolio performance. Ceria and Stubbs [11] introduced a methodology for robust portfolio optimization, taking into account the estimation errors in input parameters while formulating the optimization problem. The approach involves minimizing the worst case return for a given confidence region. They observed that the constructed robust portfolios perform superior in comparison to those constructed using mean-variance analysis in most of the cases but not in each month with certainty. Utilizing the above framework, Scherer [24] showed that robust methods do not lead to significant change in the efficient set. Constructing an example, he showed that robust portfolio underperforms out of the sample in comparison to Markowitz portfolio, especially in the case of low risk aversion and high uncertainty aversion. He also argued that performance of robust portfolio is dependent upon the consistency between uncertainty aversion and risk aversion which is quite complicated. Santos [23] performs similar experiments and concluded, stating better performance of robust optimization in comparison to the portfolios constructed using mean-variance analysis in the case of simulated data unlike the real market data.

Chapter 2

Robust Portfolio Optimization Models using Uncertainty Sets

All the real world optimizing problems inevitably have uncertain parameters embedded in them. In order to tackle such problems, a framework called "Stochastic Programming" [8] is used, which can model such problems having uncertain parameters. These models take the probability distributions of the underlying data into consideration. To improve the stability of the solutions, robust methods such as re-sampling techniques, robust estimators and Bayesian approaches were developed. One of the approaches is **robust optimization**, which is used when the parameters are known to lie in a certain range. In this chapter, we discuss some robust models with worst-case optimization approaches for a given objective function within a predefined "uncertainty" set.

The concept of uncertainty sets was introduced by Soyster [25], where he uses a different definition for defining a feasible region of a convex programming problem. In this definition, the convex inequalities are replaced by convex sets with a condition that the finite sum of convex sets again should be within another convex set. In another way, he defines a new linear programming problem (LPP) with uncertain truth value, but it is bound to lie within a defined convex set. Later, El Ghaoui and Lebret [15] extend these uncertainty sets to define a robust formulation while tackling the least-squares problem having uncertain parameters, but they are bounded matrices. In their work, they describe the problem of finding a worst-case residual and refer the solution as a robust least-squares solution. Furthermore, they show that it can be computed via semi-definite or second order cone programming. El Ghaoui, Oustry, and Lebret [16] further study how to integrate bounded uncertain parameters in semidefinite programming. They introduce robust-formulations for semidefinite programming and provide sufficient conditions to guarantee the existence of such robust solutions. Ben-Tal and Nemirovski [4] mainly focus on the uncertainty related with *hard* constraints and which are *ought* to be satisfied, irrespective of the representation of the data. They suggest a methodology where they replace an actual uncertain linear programming problem by its robust counterpart. They show that the robust counterpart of an LPP with the ellipsoidal uncertainty set is computationally attractive, as it reduces to a polynomial time solvable conic quadratic program. Additionally, they use interior points methods [5] to compute the solutions efficiently. Along the same lines, Goldfarb and Iyengar [17], focus on the robust convex quadratically constrained programs which are a subclass of the robust convex programs of Ben-Tal and Nemirovski [4]. They mainly work on finding uncertainty sets which structures this subclass of programs as second-order cone programs.

In its early phases, the major directions of research were to introduce robust formulations and to build uncertainty sets for robust counterparts of the LPP as they are computationally attractive. Once the basic framework of robust optimization was established, it is now applied across various domains such as learning, statistics, finance and numerous areas of engineering.

2.1 Uncertainty Sets

The determination of the structure of the uncertainty sets, so as to obtain computationally tractable solutions has been a key step in robust optimization. In the real world, even the distribution of asset returns has an uncertainty associated. In order to address this issue, a most frequently used technique is to find an estimate of the uncertain parameter and to define a geometrical bound around it. Empirically, historical data is used to compute the estimates of these uncertain parameters. For a given optimization problem, determining the geometry of the uncertainty set is a difficult task. Accordingly, in this section, we discuss a couple of popular types of uncertainty sets which are used in portfolio optimization.

In literature, there are many extensions of uncertainty sets varying from simple polytopes to statistically derived conic-representable sets. A *polytopic* [14] uncertainty set which resembles a "box", it is defined as

$$U_\delta(\hat{\mathbf{a}}) = \{\mathbf{a} : |a_i - \hat{a_i}| \leq \delta_i, i = 1, 2, 3, ..., N\}, \tag{2.1}$$

where $\mathbf{a} = (a_i, a_2, ..., a_N)$ is a vector of values of uncertain parameters of dimension N and $\hat{\mathbf{a}} = (\hat{a_1}, \hat{a_2}, ..., \hat{a_N})$ is generally the estimate for $\mathbf{a}$.

There is another class of uncertainty sets which takes the second moment of the distribution into account. Such type of sets are called *ellipsoidal* uncertainty sets. One of the most popular way of defining them [14] is

$$U(\hat{\mathbf{a}}) = \left\{\mathbf{a} : \mathbf{a} = \hat{\mathbf{a}} + \mathbf{P}^{1/2}\mathbf{u}, ||\mathbf{u}|| \leq 1\right\}, \tag{2.2}$$

where the choice of $\mathbf{P}$ is driven by the optimization problem. The main motivation behind the use of such kind of sets is that they come up naturally when one tries to estimate uncertain parameters using techniques like regression. Additionally, these sets take probabilistic properties into account. We further discuss how to model the uncertainties for some of the financial indicators.

2.1.1 Uncertainty in Expected Returns

Recently, many attempts have been made to model the uncertainty in the expected returns because of several reasons. When compared with variances and covariances, it is known that the effect on the performance of portfolio due to the estimation error is more in case of expected returns. Though it is unlikely that the future returns of the assets are equal to the estimated value of expected return, one can foresee that they will be within a certain range of the estimated return. Accordingly, one can define uncertainty sets in such a way so that expected values lie inside the geometric bound around the estimated value, say $\hat{\boldsymbol{\mu}}$.

In a simple scenario, one can define possible intervals for the expected returns of each individual asset by using box uncertainty set. Mathematically, it can be expressed as

$$U_{\boldsymbol{\delta}}(\hat{\boldsymbol{\mu}}) = \left\{\boldsymbol{\mu} : |\mu_i - \hat{\mu}_i| \leq \delta_i, i = 1, 2, 3, ..., N\right\}, \tag{2.3}$$

where N represents the number of stocks and δ_i represents the value which determines the confidence interval region for individual assets. Clearly from the above expression, for the asset i, the estimated error has an upper bound limit of δ_i. On incorporating the box uncertainty set in the robust formulation of Markowitz model having the constraints that involve no short selling and the sum of the weights equalling unity, the following max-min problem

$$\max_{\mathbf{x}} \left\{ \min_{\boldsymbol{\mu} \in U_{\boldsymbol{\delta}}(\hat{\boldsymbol{\mu}})} \boldsymbol{\mu}^{\top}\mathbf{x} - \lambda \mathbf{x}^{\top}\Sigma\,\mathbf{x} \right\} \text{such that } \mathbf{x}^{\top}\mathbf{1} = 1 \text{ and } \mathbf{x} \geq 0, \tag{2.4}$$

transforms to a maximization problem

$$\max_{\mathbf{x}} \quad \hat{\boldsymbol{\mu}}^{\top}\mathbf{x} - \lambda \mathbf{x}^{\top}\Sigma\,\mathbf{x} - \boldsymbol{\delta}^{\top}|\mathbf{x}| \quad \text{such that } \mathbf{x}^{\top}\mathbf{1} = 1 \text{ and } \mathbf{x} \geq 0, \tag{2.5}$$

where λ represents the risk-aversion of the investor and $\mathbf{1}$ represents the unity vector.

The most popular choice is to use ellipsoidal uncertainty set, as it takes the second moments into account. Uncertainty in expected return using ellipsoidal uncertainty set is expressed as

$$U_{\delta}(\hat{\boldsymbol{\mu}}) = \left\{ \boldsymbol{\mu} : (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^{\top} \Sigma_{\boldsymbol{\mu}}^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \delta^2 \right\}, \tag{2.6}$$

where Σ_{μ} is a variance-covariance matrix of the estimation error of expected returns of the assets. Again, solving (2.4) with ellipsoid uncertainty set yields

$$\max_{\mathbf{x}} \left\{ \hat{\boldsymbol{\mu}}^{\top}\mathbf{x} - \lambda \mathbf{x}^{\top}\Sigma_{\boldsymbol{\mu}}\,\mathbf{x} - \delta\sqrt{\mathbf{x}^{\top}\Sigma_{\boldsymbol{\mu}}\,\mathbf{x}} \right\} \text{such that } \mathbf{x}^{\top}\mathbf{1} = 1 \text{ and } \mathbf{x} \geq 0. \tag{2.7}$$

While dealing with the box uncertainty, it is assumed that the returns follow normal distribution as it eases the task of computing the desired confidence intervals for each individual asset. We define δ_i for $100(1-\alpha)\%$ confidence level as follows:

$$\delta_i = \sigma_i\, z_{\frac{\alpha}{2}}\, n^{-\frac{1}{2}} \tag{2.8}$$

where $z_{\frac{\alpha}{2}}$ represents the inverse of standard normal distribution, σ_i is the standard deviation of returns of asset i and n is the number of observations of returns for asset i.

For the same reason, if the uncertainty set follows ellipsoid model, the underlying distribution is assumed to be tracing a χ^2 distribution with the number of assets being the degrees of freedom (df). Accordingly, for $100(1-\alpha)\%$ confidence level, δ is defined in the following manner

$$\delta^2 = \chi^2_N(\alpha) \tag{2.9}$$

where $\chi^2_N(\alpha)$ is the inverse of a chi square distribution with N degrees of freedom.

2.1.2 Separable Uncertainty Set

As mentioned earlier, portfolio performance is more sensitive towards estimation error in mean returns of assets in comparison to variances and covariances of asset returns. This is one of the major reasons behind research works laying less emphasis upon the uncertainty set for covariance matrix of asset returns. Box uncertainty set for covariance matrix of returns is defined on similar lines as that for expected returns. Lower bound $\underline{\Sigma}_{ij}$ and upper bound $\overline{\Sigma}_{ij}$ can be specified for each entry Σ_{ij} of the covariance matrix. Using this methodology, the constructed box uncertainty set for covariance matrix is expressed in the following form:

$$U_\Sigma = \{\Sigma : \underline{\Sigma} \leq \Sigma \leq \overline{\Sigma},\ \Sigma \succeq 0\}. \tag{2.10}$$

In the above equation, the condition $\Sigma \succeq 0$ implies that Σ is a symmetric positive semidefinite matrix. This condition is necessary in most of the robust optimization approaches, particularly those involving Markowitz model as the basic theoretical framework.

Tütüncü and Koenig [26] discuss a method to solve the robust formulation of Markowitz optimization problem having non-negativity constraints upon the weights of assets. They define the uncertainty set for covariance matrix as per above equation and uncertainty set for expected returns as $U_{\boldsymbol{\mu}} = \{\boldsymbol{\mu} : \underline{\boldsymbol{\mu}} \leq \boldsymbol{\mu} \leq \overline{\boldsymbol{\mu}}\}$, where $\underline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\mu}}$ represent lower and upper bounds on mean return vector $\boldsymbol{\mu}$ respectively . Accordingly, the robust optimization problem

$$\max_{\mathbf{x}} \left\{ \min_{(\boldsymbol{\mu},\Sigma)\in(U_{\boldsymbol{\mu}},U_{\Sigma})} \boldsymbol{\mu}^{\top}\mathbf{x} - \lambda\mathbf{x}^{\top}\Sigma\,\mathbf{x} \right\} \text{such that } \mathbf{x}^{\top}\mathbf{1} = 1 \text{ and } \mathbf{x} \geq 0, \quad (2.11)$$

can be formulated as following:

$$\max_{\mathbf{x}} \left\{ \underline{\boldsymbol{\mu}}^{\top}\mathbf{x} - \lambda\mathbf{x}^{\top}\,\overline{\Sigma}\,\mathbf{x} \right\} \text{such that } \mathbf{x}^{\top}\mathbf{1} = 1 \text{ and } \mathbf{x} \geq 0. \quad (2.12)$$

Above approach involves the use of “separable” uncertainty sets, which implies, the uncertainty sets for mean returns and covariance matrix are defined independent of each other.

2.1.3 Joint Uncertainty Set

There are certain drawbacks associated with separable uncertainty sets. Lu [19] argues that such kind of uncertainty sets don’t take the knowledge of actual confidence level into consideration. Secondly, separable uncertainty sets don’t incorporate the joint behavior of mean returns and covariance matrix. As a result, these uncertainty sets are completely or partially similar to box uncertainty sets. This is one of the major reasons behind robust portfolios being conservative or highly non-diversified as observed in numerous computations. In order to address these drawbacks, Lu proposes a "joint uncertainty set". This uncertainty set is constructed as per desired confidence level using a statistical procedure that takes the factor model [18] for asset returns into consideration.

Delage and Ye [13] define a joint uncertainty set that takes into consideration the uncertainty in distribution of asset returns as well as moments (mean returns and covariance matrix of returns). The proposed uncertainty set having confidence parameters, $\gamma_1 \geq 0$ and $\gamma_2 \geq 1$, is given by:

$$\begin{aligned} &(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\top \ \hat{\Sigma}^{-1} \ (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \gamma_1, \\ &E[(\mathbf{r} - \hat{\boldsymbol{\mu}})(\mathbf{r} - \hat{\boldsymbol{\mu}})^\top] \leq \gamma_2 \hat{\Sigma}. \end{aligned} \tag{2.13}$$

In the above equation, $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ represent the estimates of mean return vector and covariance matrix of asset returns respectively, and $\mathbf{r}$ is the random return vector. Using this uncertainty set, they formulate the portfolio optimization problem as a Distributionally Robust Stochastic Program (DRSR). Accordingly, they demonstrate that the problem is computationally tractable by solving it as a semidefinite program.

Chapter 3

Computational Results

We analyze the performance of robust portfolio optimization methods discussed in the preceding chapter in a practical setup involving domestic market data and simulated data. The analysis is performed under two scenarios, namely, the number of stocks N being 31 and the number of stocks N being 98. This is done in order to observe the effect of increase in number of stocks on the performance of robust methods with respect to the Markowitz model.

For the first scenario, we use the daily log-returns based on daily adjusted close price of the 31 stocks comprising BSE30 (data source: Yahoo Finance [3]). Accordingly, we have considered the period from December 18, 2017 to September 30, 2018 (both inclusive) comprising of a total of 194 active trading days. Corresponding to this market data, we prepare two sets of simulated data for the 31 assets by sampling returns from a multivariate normal distribution with mean and covariance matrix set equal to those obtained from the S&P BSE30 data. The first set comprises of number of samples same as the daily log-return observations in S&P BSE30 market data, namely, 193, whereas the second set comprises of 1000 samples. We use these two types of simulated data in order to study the impact of number of samples in simulated data on the portfolio optimization approaches. Additionally, we compare the performance of robust optimization approaches vis-à-vis Markowitz approach in the case of market data as well as simulated data in each scenario in order to analyze whether worst-case portfolio optimization approaches are useful in real market setup.

For the second scenario, we use the log-returns based on daily adjusted close price data of the 98 stocks comprising S&P BSE100 (data source: Yahoo Finance [3]) with the period spanning from December 18, 2016 to September 30, 2018. Two sets of simulated data are constructed using multivariate normal distribution, on the similar lines as the first scenario.

The robust portfolio optimization approaches that we have taken into consideration for analyzing their performance with respect to Markowitz model without short-selling (**Mark**) are as follows:

1. Robust Model involving box uncertainty set in expected return without short-selling (**Box**).

2. Robust Model involving ellipsoidal uncertainty set in expected return without short-selling (**Ellip**).

3. Robust Model involving separable uncertainty set without short-selling (**Sep**).

For Box and Ellip model, we construct uncertainty sets in expected mean return with $100(1-\alpha)\%$ confidence level by considering $\alpha = 0.05$. Separable uncertainty set in Sep model is constructed as a $100(1-\alpha)\%$ confidence interval for both μ and Σ using Non-parametric Boostrap Algorithm with same α as in other robust models and assuming β, $i.e$ the number of simulations, equal to 8000.

The performance of the robust optimization approaches mentioned above are studied by taking into consideration the "Sharpe Ratio" of the portfolios constructed having λ representing risk-aversion in the ideal range [14] *i.e.*, $\lambda \in [2, 4]$. Since the yield for Treasury Bill in India from 2016 to 2018 has been found to oscillating around 6% [2], so, we have assumed the annualized riskfree rate to be equal to 6%. In the following sections, we present the computational results observed in case of two scenarios as discussed above.

3.1 Performance with $N = 31$ assets

In Figure 3.1 and Table 3.1, we present the efficient frontier and performance of portfolios constructed by applying the Marko model and three robust models to the simulated returns with 1000 samples. We observe that efficient frontiers for Ellip and Sep models lie below the one for Mark model. This supports the argument made by Broadie [10] regarding over-estimation of efficient frontier in case of Mark model. Overlap of efficient frontiers for Mark and Box models indicates that utilizing box uncertainty set for robust optimization does not prove to be of much use in this case. This claim is supported quantitatively from Table 3.1 as well as since the average Sharpe ratio for portfolios constructed in the ideal range of risk-aversion is same in case of both the models. From Table 3.1, we infer that Sep model performs at par with Mark model if we take into consideration the average Sharpe Ratio. This is evident from Figure 3.1 as well, since Mark model starts outperforming Sep model in terms of Sharpe ratio after the risk-aversion crosses 3. On the other hand, Ellip model outperforms all the models including Mark model in the entire ideal range of risk-aversion.

On performing the simulation study with same number of samples as in

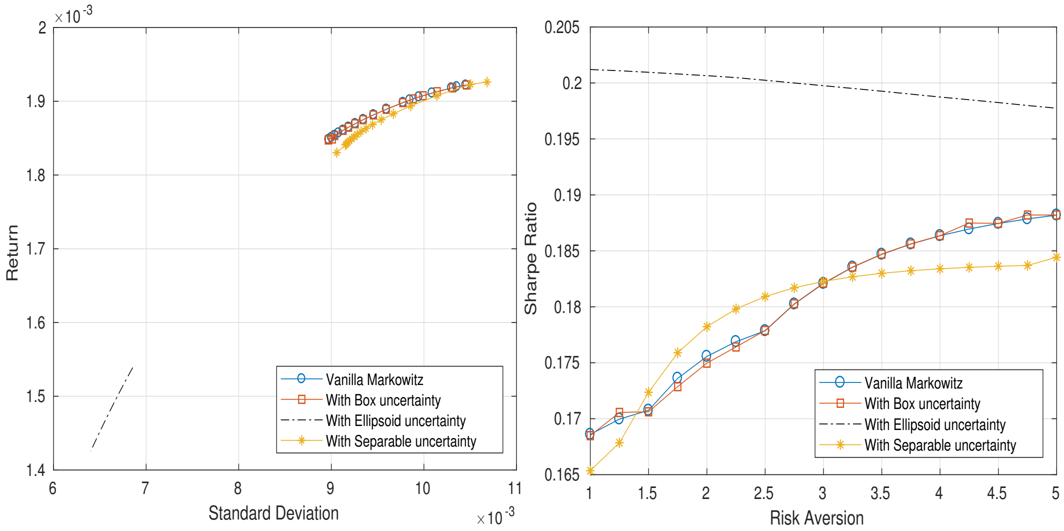

Figure 3.1: Efficient Frontier plot and Sharpe ratio plot for different portfolio optimization models in case of Simulated Data with 1000 samples (31 assets)

λ	SR_{Mark}	SR_{Box}	SR_{Ellip}	SR_{Sep}
2	0.176	0.175	0.201	0.178
2.5	0.178	0.178	0.2	0.181
3	0.182	0.182	0.2	0.182
3.5	0.185	0.185	0.199	0.183
4	0.186	0.186	0.199	0.183
Avg	0.181	0.181	0.2	0.182

Table 3.1: Comparison of different portfolio optimization models in case of Simulated Data with 1000 samples (31 assets)

case of market data (Figure 3.2 and Table 3.2), we observe similar results on comparing Box model with Mark model. However, we observe slight inconsistency in performance of Box model as evident from the Sharpe Ratio plot. The efficient frontiers for Sep and Ellip model lie below that for Mark model. We also infer that Sep and Ellip model outperform Mark model in terms of Sharpe Ratio in the ideal range of risk-aversion. However, it is difficult to compare the performance of Sep model with that of Ellip model in this case since the average Sharpe Ratio for both of them is almost the same.

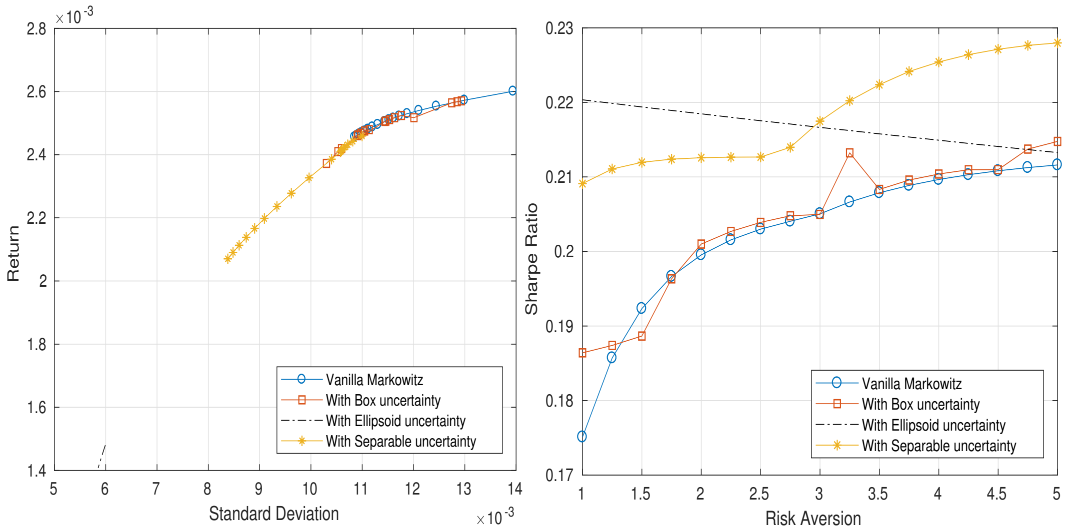

Figure 3.2: Efficient Frontier plot and Sharpe ratio plot for different portfolio optimization models in case of Simulated Data with same number of samples as market data (31 assets)

λ	SR_{Mark}	SR_{Box}	SR_{Ellip}	SR_{Sep}
2	0.2	0.198	0.218	0.213
2.5	0.203	0.204	0.218	0.213
3	0.205	0.207	0.217	0.217
3.5	0.208	0.209	0.216	0.222
4	0.21	0.21	0.215	0.225
Avg	0.205	0.206	0.217	0.218

Table 3.2: Comparison of different portfolio optimization models in case of Simulated Data with same number of samples as market data (31 assets)

In a “real” market setup involving stocks comprising BSE30, we observe that efficient frontier for Box model almost overlaps with that for Mark model. However, the performance of Box model in terms of Sharpe ratio is quite inconsistent as evident from the plot in Figure 3.3. Efficient frontier for Sep model lies below that of the Mark model and the gap between the plots widens to a great extent, incase of the Ellip model. We also observe that Sep model outperforms Mark model in the ideal range of risk-aversion on taking the Sharpe Ratio into consideration as the performance measure. This is not true in case of Ellip Model as evident from the Sharpe ratio plot in Figure 3.3. Even from Table 3.3, we observe that average Sharpe ratio for Ellip model is only slightly greater than that for Markowitz. Thus, unlike the simulated data, Sep model performs superior in comparison to Ellip model when applied to market data.

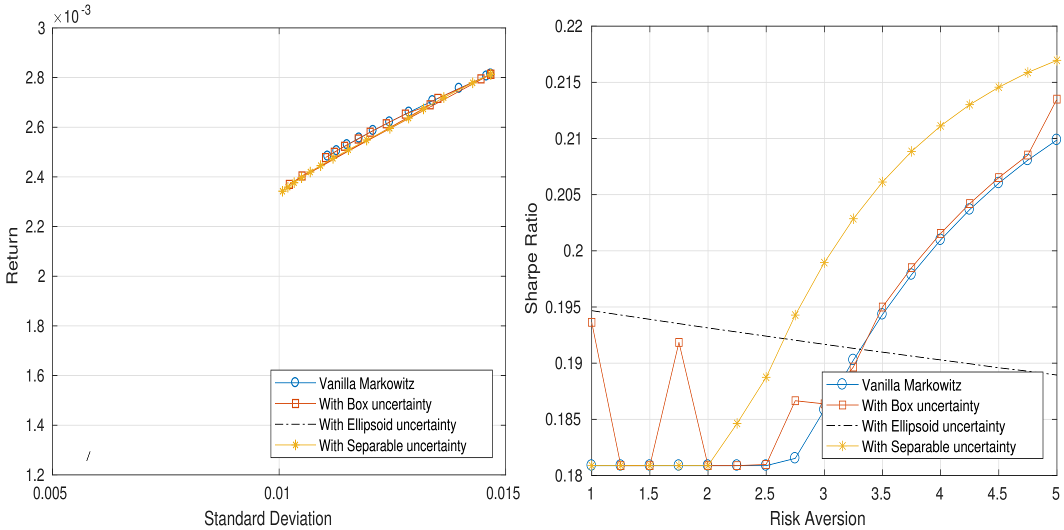

Figure 3.3: Efficient Frontier plot and Sharpe ratio plot for different portfolio optimization models in case of Market Data (31 assets)

A common observation that could be inferred from three cases considered in the scenario involving less number of assets is Sep and Ellip model perform superior or equivalent in comparison to Markowitz model in the ideal range of risk-aversion.

λ	SR_{Mark}	SR_{Box}	SR_{Ellip}	SR_{Sep}
2	0.181	0.181	0.193	0.186
2.5	0.181	0.181	0.192	0.193
3	0.186	0.191	0.192	0.202
3.5	0.194	0.195	0.191	0.209
4	0.201	0.202	0.19	0.213
Avg	0.189	0.19	0.192	0.2

Table 3.3: Comparison of different portfolio optimization models in case of Market Data (31 assets)

3.2 Performance with $N = 98$ assets

On applying robust model along with the Markowitz model on a simulated data having 1000 samples, we observe results similar to the corresponding case in previous scenario when we compare Box model with Markowitz model. This is evident from the coinciding plots of the efficient frontier and the Sharpe ratio for both the models in Figure 3.4. Contrary to the similar case for the scenario involving less number of assets (31 assets), we observe that not only does the Ellip model but also the Sep model outperforms the Mark model taking into consideration the portfolios constructed in the ideal range of risk-aversion. From Table 3.4, we infer that Ellip model performs superior in comparison to Sep model in terms of greater average Sharpe ratio.

λ	SR_{Mark}	SR_{Box}	SR_{Ellip}	SR_{Sep}
2	0.185	0.185	0.245	0.191
2.5	0.185	0.185	0.244	0.202
3	0.19	0.192	0.244	0.213
3.5	0.199	0.199	0.243	0.221
4	0.207	0.207	0.242	0.226
Avg	0.193	0.194	0.244	0.21

Table 3.4: Comparison of different portfolio optimization models in case of Simulated Data with 1000 samples (98 assets)

Figure 3.5 and Table 3.5 present the results of simulation study with same

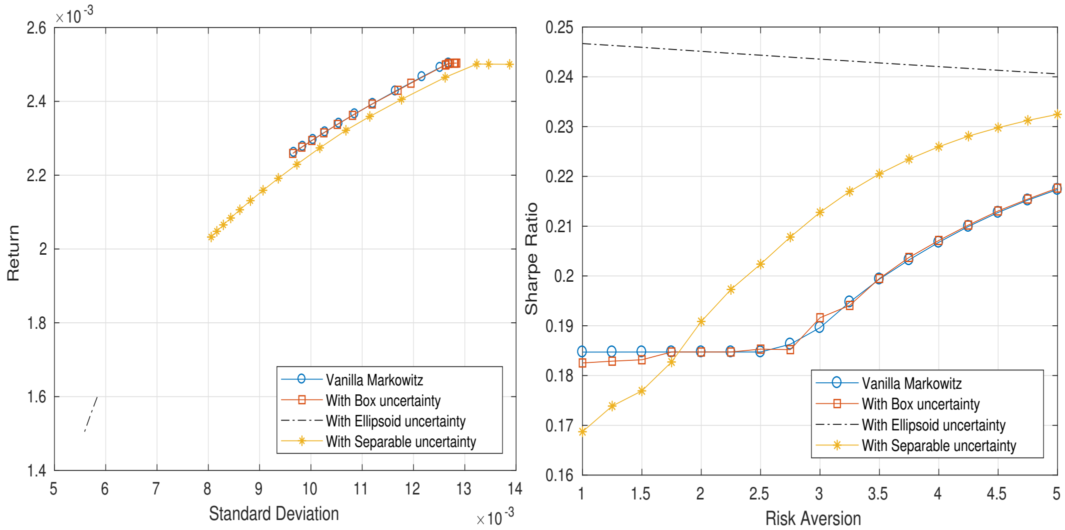

Figure 3.4: Efficient Frontier plot and Sharpe ratio plot for different portfolio optimization models in case of Simulated Data with 1000 samples (98 assets)

number of samples as that of market data (BSE100). Results observed on comparing the Box model with the Mark model are similar to the previous case. In the ideal range of risk aversion, we observe that efficient frontier for Ellip as well as Sep model lie below Mark model. Additionally, both the models perform better than Mark model in terms of the Sharpe Ratio. From the Sharpe ratio plot in Figure 3.5, it is difficult to compare Sep and Ellip model since each outperforms the other in a different sub-interval of risk-aversion. Similar values of the average Sharpe ratio in Table 3.5 supports the claim of equivalent performance of these two models in this case.

The results for the case involving the market data (that contains log-returns of stocks comprising BSE100) are shown in Figure 3.6 and Table 3.6. The efficient frontier plot leads to observations similar to the previous case. However, there is a slight inconsistency in the performance of Box model as observed in the Sharpe ratio plot. Portfolios constructed using Sep and Ellip model outperform the ones constructed using the Mark model in the ideal

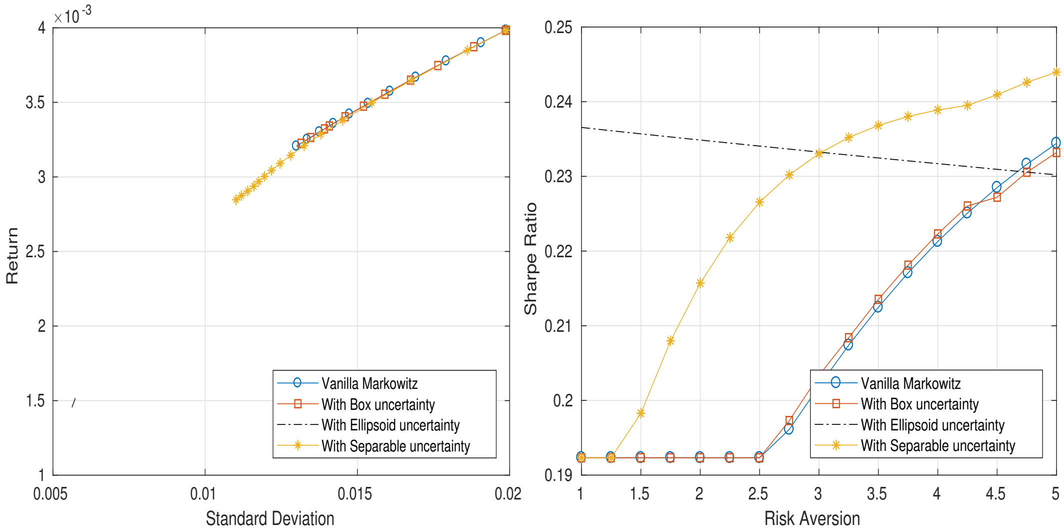

Figure 3.5: Efficient Frontier plot and Sharpe ratio plot for different portfolio optimization models in case of Simulated Data with same number of samples as market data (98 assets)

λ	SR_{Mark}	SR_{Box}	SR_{Ellip}	SR_{Sep}
2	0.192	0.192	0.235	0.216
2.5	0.192	0.192	0.234	0.227
3	0.202	0.203	0.233	0.233
3.5	0.212	0.213	0.232	0.237
4	0.221	0.222	0.232	0.239
Avg	0.204	0.205	0.233	0.23

Table 3.5: Comparison of different portfolio optimization models in case of Simulated Data with same number of samples as market data (98 assets)

range of risk-aversion. Ellip model performs slightly better than the Sep model as evident from the Sharpe Ratio plot. Marginal difference in average Sharpe ratio between these two models supports this inference.

We draw a common inference from the three cases considered in the scenario involving greater number of assets, *i.e.*, Sep and Ellip model outperform the Markowitz model in the ideal range of risk aversion.

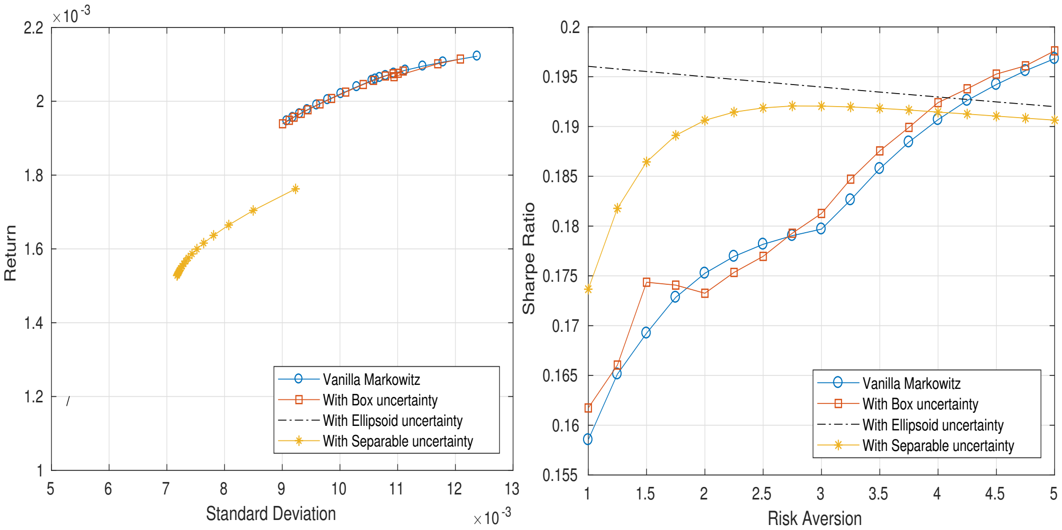

Figure 3.6: Efficient Frontier plot and Sharpe ratio plot for different portfolio optimization models in case of Market Data (98 assets)

λ	SR_{Mark}	SR_{Box}	SR_{Ellip}	SR_{Sep}
2	0.175	0.173	0.195	0.193
2.5	0.178	0.177	0.194	0.193
3	0.18	0.181	0.194	0.193
3.5	0.186	0.188	0.193	0.192
4	0.191	0.192	0.193	0.192
Avg	0.182	0.182	0.194	0.192

Table 3.6: Comparison of different portfolio optimization models in case of Market Data (98 assets)

Chapter 4

Conclusions and Comments

In this chapter, we analyze the trend of the Sharpe ratios in different kind of scenarios. We have taken the "Adjusted Close Price" of S&P BSE30 and S&P BSE100 stocks into consideration. In order to analyze the performance of robust models, we have simulated the data using true mean and covariance matrices. Again, as the number of instances in real data we could get for all the stocks was very less, we simulated two types of environments, one where the number of samples in the simulated data matches the number of instances of real market data available, say ζ and another environment where the number of simulations in simulated data is very large, a constant, irrespective of the number of stocks. The sole motivation behind this kind of setup was to understand if the market data we obtained (which was very less) was able to capture the trends and give the better portfolios.

4.1 From the Standpoint of Number of Stocks

At first, we describe the tabulation in Table 4.1. For a particular row and a column, we picked the maximum possible Sharpe Ratio we could get in that

	#stocks = 31	#stocks = 98
#generated_simulations = 1000	0.2	0.244
#generated_simulations = ζ	0.218	0.233
Market data	0.2	0.194

Table 4.1: The maximum average Sharpe ratio compared by varying the number of stocks in different kinds of scenarios.

particular scenario. For example, in order to fill the entry for S&P BSE100 where we simulated ζ instances using true mean vector and true covariance matrix, we refer to Table 3.5 (which explains the experiment corresponding S&P BSE100 with ζ simulated instances) and take the maximum of last row *i.e.*, maximum of average Sharpe ratios that can attained using the available models.

Accordingly, from Table 4.1, the conclusion that can be drawn is that more the number of stocks, the better is the performance of the portfolios constructed using robust optimization. This claim can be supported via both qualitative and quantitative approaches. Qualitatively, stocks in a portfolio represent its diversification. According to Modern Portfolio Theory (MPT), investors get the benefit of better performance from diversifying their portfolios as it reduces the risk of relying on only one security to generate returns. Value Research Online [1] provides us with the information that on an average basis, large-cap funds hold around 38 shares, mid-cap funds around 50 and 52 for balanced funds in which around 65-70% of their assets are in equity. From the above table, we can quantitatively justify by observing that the Sharpe ratio was more for portfolios with larger number of stocks when compared to portfolios with smaller number of stocks. On the contrary, we believe that the reason for the opposite behaviour when it comes to real market is the insufficient available data, when it comes to larger number of stocks. This will be made more clear in the subsequent sections. One can argue about the second case where the number of simulations is equal to available market instances, in which case our simulated data follows multivariate normal distribution whereas the real market data need not to follow any kind of distribution.

4.2 From the Standpoint of Number of Samples Generated

In this section, we solely focus on the performance when different number of samples were generated. We tabulated Table 4.2 in the same way as we described in the preceding section. Here, we notice some interesting performance trends. One can observe that in the case of smaller number of stocks, the performance when same number of instances (ζ) are simulated is better than the scenario when large (1000) number of simulations were generated. On the contrary, exactly opposite trend can be observed when higher number of stocks are taken into consideration. We explain this type of behaviour as follow (as mentioned above): In the available real market data, the number of instances available for larger number of stocks is relatively low. So, when more number of samples were generated, we observe higher Sharpe ratios when compared to ζ number of simulations. We are yet to explore the reason behind such type of behaviour when smaller number of stocks are considered.

	#samples = 1000	#samples = ζ
#stocks = 31	0.2	0.218
#stocks = 98	0.244	0.233

Table 4.2: The maximum average Sharpe ratio compared by varying the number of stocks in different kinds of scenarios.

4.3 From the Standpoint of Type of the Data

In this final section, we focus on the type of data which we are dealing with. Again, Table 4.3 is tabulated as explained in the preceding sections. Here the behaviour is straight forward. In both the cases, the performance in the case of simulated data is better than the real market data. This is clear from the fact that the real market data are difficult to model and hardly may follow any distribution, whereas the simulated data simply follows multivariate normal distribution with mean and covariances as the true values obtained from the data.

	Simulated data	Real Market data
#stocks = 31	0.218	0.2
#stocks = 98	0.244	0.194

Table 4.3: The maximum average Sharpe ratio compared by varying the type of the data in different kinds of scenarios.

Bibliography

[1] Portfolios: do number of stocks matter? https://www.valueresearchonline.com/story/h2_storyView.asp?str=200396&fbclid=IwAR3YkUaNLCY3keZGXfooehh0WU90zkZ5s94h0Uazj1-rz0FSy5bw6R1E-MY.

[2] Reserve bank of india - weekly statistical supplement. https://www.rbi.org.in/scripts/WSSViewDetail.aspx?TYPE=Section&PARAM1=4%0A.

[3] Yahoo finance. https://in.finance.yahoo.com/.

[4] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[5] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Oper. Res. Lett.*, 25(1):1–13, August 1999.

[6] M. J. Best and R. R. Grauer. On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *The Review of Financial Studies*, 4(2):315–342, 1991.

[7] M. J. Best and R. R. Grauer. Sensitivity analysis for mean-variance portfolio problems. *Management Science*, 37(8):980–989, 1991.

[8] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming (Springer Series in Operations Research and Financial Engineering)*. Springer, 2011.

[9] F. Black and R. Litterman. Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43, 1992.

[10] M. Broadie. Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, 45(1):21–58, Dec 1993.

[11] S. Ceria and R. A. Stubbs. Incorporating estimation errors into portfolio selection: Robust portfolio construction. *Journal of Asset Management*, 7(2):109–127, Jul 2006.

[12] V. K. Chopra and W. T. Ziemba. The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, 19(2):6–11, 1993.

[13] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3):595–612, 2010.

[14] F. J. Fabozzi, P. N. Kolm, D. Pachamanova, and S. M. Focardi. *Robust Portfolio Optimization and Management*. Wiley, 2007.

[15] L. E. Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications*, 18(4):1035–1064, 1997.

[16] L. E. Ghaoui, F. Oustry, and H. Lebret. Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization*, 9(1):33–52, 1998.

[17] D. Goldfarb and G. Iyengar. Robust convex quadratically constrained programs. *Mathematical Programming*, 97(3):495–515, Aug 2003.

[18] D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Mathematics of operations research*, 28(1):1–38, 2003.

[19] Z. Lu. A new cone programming approach for robust portfolio selection. *Optim. Methods Softw.*, 26(1):89–104, 2006.

[20] H. M. Markowitz. Portfolio selection. *The journal of finance*, 7(1):77–91, 1952.

[21] H. M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press, 1959.

[22] R. O. Michaud. The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45(1):31–42, 1989.

[23] A. Santos. The out-of-sample performance of robust portfolio optimization. *Brazilian Review of Finance*, 8(2):141–166, 2010.

[24] B. Scherer. Can robust portfolio optimisation help to build better portfolios? *Journal of Asset Management*, 7(6):374–387, Mar 2007.

[25] A. L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21(5):1154–1157, 1973.

[26] R. H. Tütüncü and M. Koenig. Robust asset allocation. *Annals of Operations Research*, 132(1-4):157–187, 2004.