Titre: Title:	Empirical Study of Out-of-Sample Performance of Sparse Mean-Variance Portfolio Optimization
Auteur: Author:	Shiva Zokaee
Date:	2021
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Zokaee, S. (2021). Empirical Study of Out-of-Sample Performance of Sparse Mean-Variance Portfolio Optimization [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/6274/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/6274/
Directeurs de recherche: Advisors:	Michel Gendreau, & Erick Delage
Programme: Program:	Maîtrise recherche en génie industriel

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Empirical Study of Out-of-Sample Performance of Sparse Mean-Variance Portfolio Optimization

SHIVA ZOKAEE
Département de mathématiques et de génie industriel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie industriel

Avril 2021

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Ce mémoire intitulé :

Empirical Study of Out-of-Sample Performance of Sparse Mean-Variance Portfolio Optimization

présenté par **Shiva ZOKAEE**
en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Andrea LODI, président
Michel GENDREAU, membre et directeur de recherche
Erick DELAGE, membre et codirecteur de recherche
Fabian BASTIN, membre

DEDICATION

To my parents, Roshanak and Abbas,
who are the light of my life,
and to my sister, Mahnaz,
who has a heart of gold.

ACKNOWLEDGEMENTS

Although I was responsible for completion of this thesis, I was fortunate to receive a great deal of support along my studies at Polytechnique Montréal.

I am delighted to have had Prof. Michel Gendreau as my supervisor and I would like to express my deepest gratitude to him for his great guidance, encouragement and believing in me.

It was my honor to work under the supervision of Prof. Erick Delage who generously guided me with his immense knowledge and provided me with the access to facilities that helped me accomplish this thesis. Also, I appreciate his financial support during my studies at Polytechnique Montréal.

Moreover, I would like to thank Prof. Andrea Lodi and Prof. Fabian Bastin for accepting to be part of this thesis committee and for their valuable feedback.

I feel so blessed for having the unconditional love and support of my parents, Abbas Zokaee and Roshanak Jafarzadeh, and my kind-hearted sister, Mahnaz Zokaee. I would like to take this opportunity to thank them sincerely.

I am grateful for having the support of my sharp-minded and hard-working peers : Amir Ardestani-Jaafari, Mehran Poursoltani, Abderrahim Fathan, Saeed Marzban, Sajad Aliakbari Sani and Chun Peng at HEC Montréal who generously shared their knowledge and experience with me.

Special thanks to Sahar Mokhtari, Setareh Farajollahzadeh, Monaz Yasoubi, Vida Fathi, Naghmeh Mahootian, Hamid Shaker, Maryam Shademani, Naeemeh Sohrabi, Sara Adrom, Elham Bagheri, Hoda Faghankhani, Samareh Azimpoor, Nima Panahifar, Eve Lamargot, Tara McGrath, Reem El-Abd, Melissa Vosough Tehrani, Sama Hajabedi, Donya Rahmani, Marjan Jafarzadeh Hafshejani, Hossein Dabiri, Ilnaz Ghahreman Bamdadi, Narges Sereshti and Sevda Mohammadi for their genuine friendship and all their encouragements.

RÉSUMÉ

La théorie moderne du portefeuille développée par Markowitz est considérée comme le fondement de la finance moderne. Cette théorie porte sur la diversification des portefeuilles d'investissement grâce à une logique rationelle et vise simultanément à minimiser le risque et à maximiser le rendement des portefeuilles. D'une part, un problème qui préoccupe les investisseurs est l'augmentation des coûts de surveillance et de gestion lorsque le nombre de titres dans leurs portefeuilles d'investissement est élevé. Par conséquent, ils préfèrent limiter le nombre de titres dans lesquelles ils investissent à un sous-ensemble de tous les titres disponibles. Ces portefeuilles sont appelés portefeuilles parcimonieux. La question qui se pose est de savoir si investir dans des portefeuilles parcimonieux est optimal par rapport aux situations où tous les titres disponibles sont inclus dans les portefeuilles d'investissement. Ce mémoire tente de répondre à cette question.

Deux principaux paramètres d'entrée du modèle de sélection de portefeuille moyenne-variance, à savoir la moyenne et la variance des rendements boursiers, sont estimés par la moyenne et la variance de l'échantillon car les vraies valeurs de ces paramètres ne sont pas connues. Le problème lié à l'utilisation de la moyenne de l'échantillon et de la variance des rendements boursiers est que les portefeuilles optimaux souffrent de l'instabilité résultant des fluctuations des échantillons des rendements boursiers. Cela signifie que le portefeuille optimal pour un échantillon de rendements boursiers spécifique donné pourrait ne pas être optimal pour d'autres échantillons. Par conséquent, les investisseurs devraient rééquilibrer leurs portefeuilles plus souvent et supporter des coûts de transaction en raison des instabilités des portefeuilles dues à l'erreur d'estimation des paramètres d'entrée.

Dans ce mémoire, le modèle de sélection de portefeuille moyenne-variance parcimonieux, qui est dérivé en incluant une restriction sur le nombre de titres dans le portefeuille optimal, est utilisé pour construire les portefeuilles optimaux parcimonieux. En outre, nous examinons si investir dans des portefeuilles parcimonieux est une décision optimale pour les investisseurs neutres et averses au risque.

Afin d'augmenter le pouvoir de généralisation du modèle de sélection de portefeuille moyenne-variance, nous évaluons la performance du modèle non seulement sur la base de la qualité des résultats dérivés des données utilisées pour estimer les paramètres d'entrée, dites données à l'intérieur de l'échantillon, mais également sur la base de la qualité des résultats obtenus à partir du nouvel ensemble de données, dites données hors échantillon. L'étude approfondie de la performance hors échantillon est menée en partant du principe que la vente à découvert

est soit interdite, soit autorisée. Les résultats de l'étude empirique confirment qu'investir dans des portefeuilles parcimonieux est une décision optimale pour les investisseurs neutres et averses au risque. Mais pour les investisseurs très averses au risque, il est optimal d'investir dans toutes les actions disponibles, en particulier lorsque la vente à découvert est autorisée.

ABSTRACT

The portfolio selection theory introduced by Markowitz is referred to as the foundation of modern finance. This theory cares about diversifying investment portfolios by a reasonable logic and aims at minimizing the risk and simultaneously maximizing the return of the portfolios. One issue that investors are concerned with is the increase in monitoring and managerial costs when the number of stocks included in their investment portfolios is high. Therefore, investors prefer limiting the number of stocks they invest in to a subset of all available stocks. Such portfolios are called sparse portfolios. The question that arises is whether investing in sparse portfolios is optimal compared to the situation in which all available stocks are included in the investment portfolio. This research aims at answering this question.

On the other hand, two main input parameters of the mean-variance portfolio selection model, namely the mean and the variance of the stock returns, are estimated by the sample mean and variance since the true values for these parameters are not known. The issue that arises due to employing the sample mean and variance of the stock returns is that the optimal portfolios suffer from instability resulting from the fluctuations in the stock returns' samples. This means that the optimal portfolio for a given specific stock returns' sample might not be optimal for other samples. Hence, the investors should rebalance their portfolios more often and incur transaction costs due to the instabilities of the portfolios resulting from the estimation error of the input parameters.

In this research, the sparse mean-variance portfolio selection model derived by including a restriction on the number of the stocks in the optimal portfolio is employed to construct the sparse optimal portfolios. Moreover, we examine whether investing in sparse portfolios is an optimal decision for the risk-neutral and risk-averse investors.

To increase the generalization power of the sparse mean-variance portfolio selection model under study, the model's performance is evaluated not only based on the quality of the results derived from the data used to estimate the input parameters, called in-sample data, but it is also evaluated based on the quality of the results obtained from the new dataset, called out-of-sample data. A comprehensive study of the out-of-sample performance is conducted under the assumptions that short-selling is prohibited and also short-selling is allowed. The empirical study results substantiate that investing in sparse portfolios is an optimal decision for the risk-neutral and risk-averse investors. But for highly risk-averse investors, it is optimal to invest in all the available stocks especially when short-selling is allowed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ACRONYMS

MIP	Mixed-integer programming
NLP	Non-Linear programming
MIQP	Mixed-integer quadratic programming
SOCP	Second order cone programming
MISOCP	Mixed-integer second order cone problem
LASSO	Least absolute shrinkage and selection operator
SR	Sharp ratio
B&C	Branch & Cut
MLE	Maximum likelihood estimator
EM	Evaluation measure

CHAPIER 1 INTRODUCTION

Every year, several billion dollars are invested in various sectors by individual investors, brokers and large fund management corporations which are the representatives of the mutual funds, pension funds, and other institutions. Hence, proper selection of the assets and securities to invest in plays a crucial role in the financial markets to be able to incur the least losses and maximize the profit to the highest possible level out of the investments [1, 2]. In recent years, more and more investment decisions are made based on the quantitative approaches and such decisions are made less qualitatively. One of the important quantitative approaches is mathematical modeling which is applied to develop mathematical formulations considering the financial environment that the investors want to invest in and the investors' goals that have to be embedded. The next step while applying mathematical modeling for decision making is to optimize the model to achieve the optimal decision [3].

One prominent and widely used investment strategy is not to restrict investment of all the available budget in one asset and invest in different assets or securities which is referred to as creating investment portfolios. Therefore, the key question is how to select the assets that create the optimal investment portfolios which has led to numerous researches conducted in the literature focusing on the portfolio selection problem. Specifically, portfolio selection problem cares about allocation of the limited budget to a finite set of assets or securities [4].

The preliminary versions of portfolio selection models used to focus on maximization of the discounted expected returns of the assets. Besides, it was widely accepted by the finance experts as a rule of thumb that diversified portfolios outperform the portfolios limited to contain one or very few number of assets. In fact, this rule of thumb was interpreted based on the law of the large numbers saying the return of the diversified portfolio focusing on maximization of the expected return is almost the same as its anticipated expected return and such diversification was able to eliminate the variance in the return of the selected portfolio [4, 5].

In 1952, Harry Markowitz criticized the shortcomings of relying on maximization of the discounted expected returns in his prominent work for which he later won the Nobel prize award in economics for his contribution to the quantitative finance. He argued that the securities yields are highly intercorrelated, therefore, diversifying a portfolio following the law of the large numbers will not minimize the variance of the portfolio return. In other words, he elaborated that the portfolio with the highest expected return will not necessarily lead to have the portfolio with the lowest variance [4].

Furthermore, the other main shortcoming he addressed was the fact that merely relying on maximizing the discounted expected returns fails to create diversified portfolios since it leads to investing all the funds in the asset which has the highest discounted return and, when there are extra funds, this approach allocates the rest of the budget on the other asset(s) with the second highest discounted return value and the procedure continues untill there is no budget left to invest. Hence, in this situation, the portfolio will not be diversified and it does not matter how precisely the discount rates are chosen and estimated and how discounted returns are derived [4].

Specifically, the main focus of Harry Markowitz in [4] is to address how to choose the components of an investment portfolio selection model and what logic have to be behind developing such models. He suggests to include the expected return and the variance of the returns as the main components of the portfolio selection model. He also asserts that following such rule, which he refers to it as “the expected return-variance of return rule”, leads to have the right type of diversification. According to Markowitz, this rule is not just about investing the available fund in a lot of securities but actually means that it is crucial to avoid investing in securities that are highly correlated with each other. The reason is that the high correlation between a set of securities implies those securities belong to the same industry and the investor is actually misdiversifying if he/she ignores variance-covariance matrix of the returns while optimizing the portfolio selection model. The idea behind such diversification is to invest in many securities belonging to different industrial segments so that the portfolio profitability will not be affected substantially in case that a specific industrial sector that is invested in be negatively impacted. Therefore, the right diversification occurs when the correlation between the selected securities is low and minimization of the variance of the returns is included in the portfolio selection model. Markowitz’s findings and proposed approach of embedding both the risk and the expected return in the portfolio selection model is referred to as the foundation of the modern finance [1].

The portfolios in stock markets are normally constructed out of 500 to 3200 number of stocks [6]. Besides, applying the original mean-variance portfolio selection model to build portfolios typically results in including very large numbers of the considered stocks in the ultimate optimal portfolio [7]. It is discussed in [8] that the large number of selected stocks in the optimal portfolios concerns both the investors and the investment companies’ managers about the following issues :

1. The number of transactions for rebalancing the portfolio can be high and the transaction costs will consequently be high, which is not desirable.
2. The more stocks are included in a portfolio, the higher the monitoring costs will be.

To tackle these issues, in practice, investors adjust the degree to which they diversify the optimal portfolio and aim to construct sparse portfolios meaning that a limited number of stocks will be held in the portfolio.

To achieve an optimal selection of a subset of stocks out of the total number of available stocks, one approach that was proposed and is criticized in [7] was to consider a postprocessing step to omit the investments that are smaller than a specific threshold. This approach does not work since usually the small investment portions belong to small-market-capitalization stocks and they will proliferate and it is not wise to eliminate them from the potential choices. Besides, hundreds of the stocks in the stock markets belong to small-market-capitalization stocks. Furthermore, from the optimization point of view, the elimination of small investments in small-market-capitalization stocks will not result in optimal portfolios and the clients of the investment institutions expect the investment institutions have full control over the investment market opportunities [7].

With respect to the above-mentioned discussions, it was needed to come up with the mathematical models to enable the decision makers to allocate optimally the available budget to a subset of stocks out of all possible stocks to invest in. Adding a cardinality constraint to the portfolio optimization model is known as the main approach that lets the investor control the size of the subset of stocks that is desired to be included in the optimal portfolio. It is worth mentioning that the models with the ability to determine the size of the portfolio are called sparse portfolio optimization models and the size of the subset to be included in the optimal portfolio is called sparsity degree. Adding a cardinality constraint to the portfolio optimization model makes the model computationally challenging to be solved by exact solution methods. Therefore, various solution methods and reformulations are proposed in order to tackle the computational intractability of the sparse portfolio optimization model for real world applications which will be discussed extensively in Chapter 2 and Chapter 3.

The assumption that the classical mean-variance portfolio selection model is built on is that the stock returns follow a multivariate normal distribution [2]. Under this assumption, the sample mean and sample variance will be the maximum likelihood estimators for the input parameters which are the mean and covariance of the stock returns. The assumption of normal multivariate distribution for stock returns is violated most of the times since the distribution of the stocks tends to have more extreme values and be more fat-tail than the normal distribution. Therefore, the performance of mean-variance portfolio optimization model will be impacted by error in the estimation of the input parameters of the model. In other words, estimation error of input parameters of the model results in unstable investment weights that fluctuate drastically when the mean-variance portfolio selection model is reoptimized. Such instabilities are not desirable since they cause large transaction costs for the investors. Hence, the model performance in the sense of parameters estimation error mitigation and stability of the portfolio weights over time are two important issues to be considered while optimizing a portfolio selection model. The research question that is aimed to be answered in this research study is whether investing in a sparse portfolio is an optimal decision for risk-neutral and risk-averse investors or not. To achieve this aim, a comprehensive empirical out-of-sample performance study is conducted for the sparse mean-variance portfolio optimization model under the assumptions that short-selling is either allowed or prohibited [1]. The sparse mean-variance portfolio optimization model is a mixed integer second order cone program model and the advantage of such reformulation is discussed in Chapter 3. In the empirical study, the evaluation measure which is the linear combination of the risk and the return computed by the sparse mean-variance portfolio optimization model is compared for the case that no sparsity restriction is applied on the model and the case in which different sparsity degrees are imposed on the desired portfolios. In order to tackle the instability of the optimal portfolio weights which is resulted from the error in estimation of the input parameters of the model, the sparse mean-variance portfolio optimization model is regularized by adding the squared $l_2 - norm$ penalty $(||x||_2^2)$ to the objective function of the model. Furthermore, the parameter that determines the right amount of generalization is tuned in a model selection procedure that is discussed in Chapter 3.

The results derived from the out of sample performance study of the sparse mean-variance portfolio optimization model reveal that investing in a subset of all the available stocks results in better portfolios for a risk-neutral and risk-averse investor under the assumption that short-selling is prohibited. The same result is derived for risk-neutral investors when short-selling is allowed, but the optimal portfolio for the investors that are highly risk-averse is built when she includes all the stocks in her investment portfolio by short-selling some of the stocks and adding the fund gained out of short-selling to the available fund that she has and invest the aggregation in other stocks.

This research study is organized as follows. The literature of portfolio optimization and out-of-sample performance study are reviewed in Chapter 2. Chapter 3 discuses the methodology applied to conduct the empirical study of out-of-sample performance of sparse mean-variance portfolio optimization which includes description of the mathematical formulation of sparse portfolio optimization and the model selection procedure. The out-of-sample performance

1. Short-selling is an investment strategy in which the investor borrows the shares of stocks from a broker and sells them. Subsequently, the money which is earned by short-selling can be used to invest in other stocks. It is worth mentioning that the borrower is responsible for buying the same number of the shares of the same stock(s) that was borrowed and returning them to the broker who is also known as the lender of the stocks [9].

study of the sparse mean-variance portfolio optimization model is conducted in Chapter 4. In Chapter 5, the summary of this research study along with its limitations will be discussed. Furthermore, some avenues for future studies will be presented.

CHAPTER 2 LITERATURE REVIEW

In this chapter, after discussing the crucial role of constructing investment portfolios and diversification of the investment portfolios, the basic mean-variance portfolio selection model will be elaborated specifically to emphasize on the revolutionary role of this optimization model in the quantitative finance field. The importance of this model relies on the logic that is behind its formulation which leads to a diversified and efficient portfolio based on appropriately inclusion of the risk aversion in the portfolio selection optimization model. Then, the different reformulations of the objective function of the classic portfolio selection model and the efforts to enhance the applicability of this model with the help of considering various constraints that are defined based on real-world needs will be reviewed. The focus of the rest of this chapter will be on discussing the usefulness of sparse portfolio construction and the models that result in sparse and stable portfolios, the computational challenges for this stream of models and the advances that are achieved. Finally, the most applied out-of-sample measures and the researches conducted in this area will be elaborated.

2.1 Portfolio Selection with Variance Risk Measure

Employing the variance-covariance of the returns as the risk measure along with the expected returns of the securities as proposed by Markowitz's has led to a revolution in the quantitative finance and has had a great influence on the theoretical and practical aspects of portfolio management field from then on [10]. Markowitz defined his portfolio selection problem following his "expected return-variance of return rule" as follows :

$$\begin{aligned} &\max_{x \in \mathbf{R}^n_+} \hat{\mu}^T x \\ &\min_{x \in \mathbf{R}^n_+} x^T \hat{\Sigma} x \\ &\quad s.t. \quad e^T x = 1 \end{aligned} \tag{2.1}$$

here $\hat{\mu} \in \mathbf{R^n}$ is a vector that contains the expected marginal returns of the stocks, $\hat{\Sigma} \in \mathbf{R^{n \times n}}$ represents the variance-covariance matrix of the returns, $e \in \mathbf{R^n_+}$ is a vector with all the elements equal to one and the elements of x are the investment proportions. The first objective defines the expected value of the portfolios returns and the second term refers to the variance of the returns. There is also one sign constraint that means no short-sell is allowed and the investment proportions have to be zero or positive amounts, i.e. $x \in \mathbf{R}_{+}^{n}$, although this constraint can easily be removed. Furthermore, his proposed model contains one hard constraint that implies the summation of the investments must be equal to the total investment budget. This hard constraint imposes that the summation of the investment portions, which each of them can take the value as $0 \leq x \leq 1$, be equal to one. By fixing the expected return to a desired return, his model solves for the associate risk that the decision maker takes to achieve her desired return. This solution approach results to build the efficient frontier which gives the idea to the investor that for each level of profitability what level of risk she has to take. In section 2.2, the other versions of formulating the objective function of this model will be discussed based on the researches that are conducted in the literature of portfolio selection models.

Since 1952, numerous researches in different fields such as finance, mathematics and computer science have been conducted to improve this approach based on faster and more efficient computation and also enhance real-world applicability of this risk-return approach. Besides, numerous variants of this model is solved with various methodologies tested on different data and performance measures (see [1] for a comprehensive review).

2.2 The Objective Function in Markowitz Mean-Variance Model

There are different formulations for the mean-variance portfolio optimization model in terms of defining the objective function. The problem either is a single-objective or a multi-objective model. In case of the single-objective function, it is presumed that the desired return level is known to the decision maker and by including the restricted expected return term by the known return level in the constraints, the objective function of the model minimizes the risk for each level of the desired portfolio return. This type of formulation results in creating the efficient frontier that helps the decision maker to see what is the optimal risk level for every level of the return that an investor prefers to achieve. The other stream that considers the multi-objective case, either transforms it to a single-objective function by giving weights to each of the return and risk terms or applies Pareto/dominance-based approaches [1].

In [1], after conducting a comprehensive review regarding the objective functions that are considered in the literature of the mean-variance portfolio optimization, the authors conclude that multi-objective formulations are preferred to the single-objective formulation since in reality the return level is not always known to the decision maker. Besides, their research reveals that the tendency to apply weighted sum method to convert the multi-objective model to a single objective one was high and Pareto/dominance-based approaches are employed in the recent years.

2.3 Portfolio Selection with Other Risk Measures than Variance

For the original mean–variance portfolio optimization model it is assumed that the assets return follow a normal multivariate distribution. This assumption is discussed and criticized in the literature of portfolio selection model arguing that the distribution of the assets tends to have more extreme values and be more fat-tail than the normal distribution. Therefore, there is the need to consider the higher moments other than the first (expected mean) and the second momemts (the variance) of the securities returns distribution in order to be able to describe the behaviour of the portfolio more precisely [2]. Since our focus in this research study is on the classical Markowitz mean-variance portfolio selection framework, we refer you to the studies done in [11], [12], [13], [14], [15], [16], [17], [18], [19] and [20] where different risk measures than the variance of the returns are investigated.

2.4 Applying Real-World Constraints to Markowitz Mean-Variance Model

Despite the usefulness of the original mean-variance portfolio selection model, there is the need to enhance its advantages by introducing the constraint(s) depending on the desired real-world application of the model. As discussed before, the original mean-variance portfolio model contains one hard constraint which sets the summation of investment amounts (weights) of all chosen assets in the optimal portfolio to be equal to the total available budget (one).

The most applied constraints to the mean-variance portfolio selection model, beside the mentioned main hard constraint, are the following constraints but not limited to them [1] :

The boundary constraint that restricts the value of each stock amount (weight) between a lower and an upper bound.

The cardinality constraint that ensures the number of the assets included in a portfolio to be equal or within a specific range.

The transaction costs constraint which relates to the fee that the investors pay when they sell or buy the stocks in the portfolio.

It is worth mentioning that even adding the linear and simple structured constraints to the original mean-variance portfolio selection model results in a combinatorial optimization problem and the large number of stocks to invest leads it to have a complex search [21]. We will get back to this issue in section 2.6 for adding the cardinality constraint to the model which is a nonlinear and non-convex constraint and makes the model become intractable.

We have to add that numerous researches are conducted to adjust the original mean-variance portfolio selection model and help to enhance its applicability in the real financial world which we refer the reader to [1], [22] and [23] for a thorough investigation of the literature that exists in this regard.

2.5 Portfolio Selection and Naive Diversification

As mentioned before, diversification is an important component of the investment decisions and diversifying the investments is known to be important in terms of reducing the investments risks. One of the investment diversification approaches is the naïve diversification which is the investment strategy that allocates uniformly the available budget to all the securities available for being invested in. There are studies that advocate applicability of naïve diversification as a general heuristics for selection and as an acceptable investment strategy by asserting that behavioral experiments show that people tend to invest equally in different investment opportunities specifically when the uncertainty in the market is high [24]. But, beside such justifications coming from the behavioral experiments, there are studies conducted to see if there is any superiority for applying naïve diversification rather than the optimization models developed for portfolio selections [25]. The answer as discussed in [26] is that, in uncertain situations and considering the distribution of securities returns to be ambiguous, the result derived from the worst-case risk minimizing portfolio models will converge to the equally invested portfolio. But, the objection against such justification is that generalizing decisions made based on worst-case situations is not reasonable in the sense that tendency of investors to take risk is different and it is not always the case that no information or inference about the investment opportunities be available to the investors.

In the rest of the literature review section, another approach for diversifying the allocation of the wealth in the stock market will be examined which aims at restricting the number of chosen stocks in the portfolios.

2.6 Sparse Portfolios

There are two main approaches to develop mathematical models with the ability to select a subset of all available stocks and such models are called sparse models. The first technique is to add a constraint which is referred to as cardinality constraint in the literature of sparse portfolio selection. Cardinality constraint is defined as $\{x \leq z, e^T z \leq k\}$, where we have $x \in \mathbf{R}^{\mathbf{n}}_{+}$, $z \in \{0, 1\}^n$ and k which determines the sparsity degree. As discussed in section 2.1, the classical mean-variance portfolio selection model can be represented as a single-objective mathematical model as :

$$\begin{aligned} \min_{x \in \mathbf{R}^n_+} \quad & x^T \hat{\Sigma} x - \hat{\mu}^T x \\ s.t. \quad & e^T x = 1 \end{aligned} \tag{2.2}$$

The mean-variance portfolio selection model defined in (2.2) is a quadratic model that for simplicity the weights of the return and the risk of the portfolio optimization model are assumed to be equal as one in the objective function. Therefore, adding the cardinality constraint and consequently its binary variable to model (2.2) results in a mixed-integer quadratic programming (MIQP) model. Quadratic models even with significant number of variables can be solved optimally as convex models by commercial solvers such as `CPLEX` in a reasonable amount of time [27]. But, MIQP's are nonlinear and nonconvex which are proved to belong to the NP-hard class of models [28, 29], therefore, such formulations are computationally challenging to be solved [27].

As mentioned before, the number of available stocks in real-world problems vary in the range $[500, 3200]$ and it is also desired to have the portfolios with the size in [20, 50] [6]. The research studies that aimed to solve the mean-variance portfolio selection problems with cardinality constraint as a mixed-integer nonlinear model with exact solution approaches were able to handle about 400 stocks [6]. Since the applied model in this empirical study solves the sparse mean-variance by exact solution methods, a comprehensive review of mean-variance models with cardinality constraints solved with exact solution methods will be presented in this section.

For the first time, Bienstock (1996) [28] employed branch-and-cut algorithm to solve problem (2.2) with cardinality constraint considering $n = 40$ and $20 \leq k \leq 25$. Another branch-and-bound algorithm was suggested to solve the mean-variance portfolio selection problem constrained with a cardinality constraint in [30] which was tested for a dataset that contained 30 stocks. Also, a Lagrangian relaxation procedure was suggested in [29] and a branch-and-bound algorithm to solve the relaxed MIQP formulation of the sparse mean-variance portfolio selection problem. This procedure was able to solve the problem of 100 stocks optimally and the gap between the upper and the lower bound for handling the problems of 200 stocks was reported to be about 0.0005.

Later on, more efficient branch-and-bound algorithms were implemented in [31] and [32] which led to certifiable optimality of handling instances including 200 stocks. Since applying branch-and-bound algorithms to the cardinality constrainted model results in weak relaxations coming from the big-M method [33], the idea of reformulating the sparse mean-variance portfolio selection model as a second order cone programming model was suggested in [34] to achieve tighter relaxations for second order cone constraints and, subsequently, to have tighter optimality bounds. Improving the suggested method in [34], non-linear branch-and-bound methods were proposed in [35] and [36] to handle up to 450 stocks in a sparse portfolio optimization problem and result in tighter second order cone relaxations and, therefore, better certifiable optimality.

Another approach to achieve tighter relaxations for sparse mean-variance portfolio selection model was proposed in [37] by defining the covariance matrix (Σ) as summation of diagonal matrix (D) and a positive semidefinite matrix ($\Sigma - D \succeq 0$). The next step of their proposed procedure was to generate perspective cuts. This method resulted in bound gap of less than one percent for sparse mean-variance portfolio selection problem of size 200 stocks. Afterwards, Frangioni and Gentile in [38] and [39] extracted larger diagonal matrices by solving auxilliary semidefinite optimization problems which led to handle sparse mean-variance portfolio selection problem with 400 stocks. More extensions on perspective reformulations were presented in [40], [41] and [42] which led to derive an equivalent mixed-integer second order cone problem (MISOCP) for the sparse mean-variance portfolio selection problem under the condition of having a positive definite covariance matrix ($\Sigma \succ 0$) and a positive semidefinite matrix as $\Sigma - D \succeq 0$ after extracting the diagonal matrix ($D \succ 0$) from the positive definite covariance matrix. Their proposed perspective reformulations was able to handle up to 400 stocks. The perspective reformulations for sparse mean-variance portfolio selection problem will be extensively discussed in Chapter 3. Besides efforts to solve sparse mean-variance portfolio selection model as a nonlinear mixed-integer problem, there are studies to reformulate this model as a mixed-integer linear model and solve them for problem size of 400 to 475 stocks [43, 44].

As reviewed in this section, the best exact methods that are applied to solve the mean-variance portfolio selection problem constrained with a cardinality constraint were able to solve the problem size of up to 400 stocks to a certifiable optimality level in a reasonable amount of time. The other stream of the researches conducted to solve such mixed-integer non-linear problem focused on developing heuristics and meta-heuristics algorithms to tackle the problem of large sizes. However, the heuristic and meta-heuristic approaches do not guarantee to result in optimal solutions and, most of the time, are able to find near optimal solutions and, in very rare cases, optimal solutions in a reasonable time. Since the number of such studies is high and a comprehensive review is done in [1], we refer the interested readers to examine this reference. Besides, there are some studies that are more recent and not mentioned in [1] such as [45], [46] and [3]. In [45] and [3], two particle swarm optimization algorithms are employed to solve the mean-variance portfolio selection problem with a cardinality constraint. Furthermore, a hybrid meta-heuristic algorithm [46] consisting of simulated annealing, genetic algorithm, electromagnetism algorithm, genetic network programming and particle swarm optimization is developed to solve the mean-variance portfolio selection problem with cardinality constraint.

The other way a sparse portfolio selection can be obtained is by applying norm penalties to the model. The norm penalty function can be defined as $||x||_\rho$ which can be either added to the objective function as $\lambda||x||_\rho$ or considered as a constraint like $||x||_\rho \leq \lambda$, where λ is a parameter defined as a scalar that controls the magnitude of the penalty. For imposing sparsity to the model, ρ can take the value 0 or 1 ; i.e. $l_0 - norm$, $||x||_0$, and $l_1 - norm$, $||x||_1$. Most of the time, $l_0 - norm$ is applied in the model as the cardinality constraint defined via $||x||_0 = \{x \leq z, e^T z \leq k, z \in \{0,1\}^n\}$ which limitations is discussed before.

Applying $l_1 - norm$, also known as least absolute shrinkage and selection operator (LASSO), results in a convex formulation that is convenient to solve comparing to the non-convex model led by employing $l_0 - norm$ to the model.

Introduction of $l_1 - norm$ was done by Tibshirani [47]. Since a convex model will be the outcome of adding $l_1 - norm$ to the portfolio selection model, the convex optimization solution approaches are applicable to solve the model [48]. Unfortunately, unlike cardinality constraint model which enables the decision maker to choose the desired degree of sparsity, $l_1 - norm$ does not allow to choose the sparsity degree, k, directly. The other shortcoming that has been addressed for $l_1 - norm$ is that it has positive effect in cases that short-sales are allowed and not satisfactory effect in term of sparsifying the model in the cases that short-sales are prohibited [48]. In other words, when short selling is prohibited and the decision variable, x_i, is not allowed to take negative values, $l_1 - norm$ can only take value *one* as the summation of the non-negative investment proportions, x_i's. Therefore, the sparsity cannot be resulted from the portfolio selection model by imposing $l_1 - norm$ in cases that short selling is not allowed.

Before moving to review the literature of the studies aiming at improving the out-of-sample performance of the mean-variance portfolio selection problem, it is worth mentioning that the regularization penalties like $l_1 - norm$, which does not affect the convexity of the problem, are able to handle up to 500 securities and result to optimal and sparse portfolios and the issue in this approach is the inability to control directly and optimally the sparsity degree.

2.7 Portfolio Selection Out-of-Sample Performance Studies

The classical mean-variance portfolio selection model proposed by Markowitz is widely applied to construct optimal portfolios, but it suffers from shortcomings with respect to their unsatisfactory out-of-sample performance. Specifically, the generalization error of the classical mean-variance portfolio selection model is high and affects its performance. The accuracy of the input parameters estimation of the model, which are the mean and the variance, is very important since we are not aware of the actual values of these parameters. Therefore, there is a need to employ appropriate approches for estimating the model parameters. For this purpose, it is widely relied on statistical approaches to make reliable parameter estimations. In statistical experiments, as there is no access to the true value of the parameters and the true distribution of the data under study, the experiments are conducted based on the limited available data/samples. The performance of the statistical experiments are evaluated not only based on the quality of the results derived from the employed data to estimate the parameters, which is called the in-sample-data, but it is also evaluated based on the quality of the results gained by the new set of data called out-of-sample data. In order to do the out-of-sample study evaluation, a measure has to be defined aligned with the objective of the model [49].

In the literature of mean-variance portfolio selection model, the measures that are widely used to evaluate the out-of-sample performance of the model are the *out-of-sample mean* ($\bar{\mu}$), *out-of-sample variance* ($v\bar{a}r$), *Sharpe ratio* ($\hat{SR}$) and *portfolio turnover* [50]. Sharpe ratio is derived through dividing the out-of-sample mean by the out-of-sample variance. In defining portfolio turnover, the aim is to see how much difference is between the investment done in a particular stock comparing to the next period and averaging over the total number of experiments for each stock under study.

The studies regarding the out-of-sample performance of the portfolio selection models can be categorized in three main streams. The first stream is concerned with achieving better out-of-sample performance by applying robust and stochastic optimization approaches. This class of models are extensively reviewed in [1]. The second stream deals with the instability of the input parameters, the estimated mean and covariance matrices, by employing the robust statistical estimators for the mean and the variance [51]. The idea in this approach is to eliminate and shrink the effect of outliers that are the extreme and rarely happening observations in the performance of the portfolio selection model [52]. The third technique is employing regularization terms to achieve more stable and robust out-of-sample performance, which is the focus of this research study and the researches done in this domain will be extensively reviewed in the following paragraphs.

The preliminary goal of applying norms to mean-variance portfolio selection model was to tackle the instability of the optimal portfolio weights caused by the estimation error of the mean and the covariance matrix as the main model parameters. In other words, the optimal investment portions derived from the mean-variance portfolio selection model might be optimal for the sample that the mean and the covariance matrix are calculated from, but due to sample fluctuations such optimal weights are unstable and drastically affect the out-of-sample performance of the model under study (see [53], [54], [55], [56], [57], [58], [59] and [60]). Investigating the portfolio selection problem in the context of statistical learning theory, such instabilities are caused by over-fitting of the model. Over-fitting means the model is suitable for the sample of the data that its parameters are derived from but it fails to estimate the other observations reliably [49]. Before examining the techniques to overcome overfitting, let us explain overfitting problem from a machine learning and statistical learning theory point of view. Since it is common to encounter large estimation errors in the models designed in machine and statistical learning field, the out-of-sample performance will be affected and will not be satisfactory. It means that the model will not be able to reliably preform in the new data sets. Therefore, numerous studies are conducted in the field of machine learning and statistical learning to improve such undesirable performances and overfitting issue and one common technique is to regularize such models [10]. Specifically, in statistical and machine learning, upper bounds that quantify the model expected generalization performance will be achieved on the generalization error by regularization. The error bounds decrease while the capacity of the problem to be fitted to the data set is diminished and this mechanism leads to better generalization of the model, meaning the model can be used by various data sets [61, 62].

The portfolio optimization problem is also known as a special case of regression model [10]. Hence, the norms that are applied in machine and statistical learning models to regularize such models can also be employed in portfolio selection model in order to tackle the overfitting and undesirable out-of-sample performance issue. Two norms that are widely applied to regularize the portfolio selection model are $l_1 - norm$ and $l_2 - norm$ that will be discussed more.

Applying $l_1 - norm$ leads to improve out-of-sample performance of the portfolio selection model and at the same time results in a sparse portfolio [48]. The first attempts to apply $l_1 - norm$ to portfolio selection model as a regularizer was conducted in [57] and [63]. The mean-variance portfolio selection model was reformulated as a least-squares regression model constrained to the budget constraint in [63]. The result of applying the $l_1 - norm$ regularization to the reformulated model was to achieve a sparse portfolio selection model with an improved out-of-sample Sharpe ratio performance.

The portfolio selection model that was investigated in [57] was a minimum-variance portfolio problem. The logic behind omitting the maximization of mean of the stock returns was said to be avoiding the high vulnerability of the mean parameter, its estimation error that commonly occurs and, consequently, its adverse impact on the out-of-sample performance of the portfolio selection model [57]. The improved out-of-sample performance of the regularized minimum-variance portfolio model with respect to variance, Sharpe ratio, and turnover performance measures is also substantiated in [57] comparing to the case of unregularized minimum-variance portfolio model. The other regularized portfolio selection models by l_1-norm aiming at achieving better out-of-sample performances and sparse portfolios are investigated in [64], [65], [66], [67], [68], [69] and [70].

Another interesting research to propose a regularization penalty derived from $l_1 - norm$ was conducted in [48]. This called sorted $l_1 - norm$ penalization (SLOPE), $\rho_\lambda(x)$, is derived as the summation of multiple $l_1 - norm$ regularizers that each $l_1 - norm$ has its own penalty coefficient and it is defined as :

$$\rho_\lambda(x) := \sum_{j=1}^{n} \lambda_j |x_{(j)}| \tag{2.3}$$

where $x_{(j)}$ is the j^{th} largest term of x in term of absolute value and λ_j's follow the order $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$. It is shown in [48] that by choosing different penalty sequences, this regularization approach can result in the whole set of optimal portfolios in terms of risk diversification frontier. Moreover, it is shown that the out-of-sample performance of the model measured in terms of turnover improves comparing to the unregularized model and, also, the case of regularizing by $l_1 - norm$.

For the first time, Andrey N. Tikhonov introduced the squared l_2-norm penalty ($||x||_2^2$) that is also known as ridge regularization [71]. Adding squared $l_2 - norm$ penalty to the portfolio selection model results in a convex model. Consequently, the convex optimization solution approaches are applicable to solve the model [48]. The squared $l_2 - norm$ regularizes the portfolio weights and mitigates the extreme portfolio weights [71], hence it helps to promote the out-of-sample performance of the model. The squared $l_2 - norm$ penalty leads to have diversified and large size portfolios and does not result in any sparsity [10, 71].

Since $l_1 - norm$ tends to set some weights to zero and the number of the zero weights are not controllable and $l_2 - norm$ helps diversification and lowers the extreme weights, both of these two regularizers are recommended to be applied in the portfolio selection models to achieve sparse portfolios and low generalization error [71]. In this direction, $l_1 - norm$ and $l_2 - norm$ regularizers are applied in the minimum variance portfolio selection model in [71] and a mechanism is proposed to calibrate the norm penalty coefficients in order to improve the out-of-sample performance of the model. Another effort was made in [72] to improve the stability of the model by adding $l_2 - norm$ to the minimum variance portfolio selection model and achieving sparse optimal portfolio weights by applying $l_1 - norm$ to the objective function of the model. It is argued in [72] that the out-of-sample performance of the model improved and also comparing to the naïve diversification a more stable and better out-of-sample performance is achieved.

In [73], $l_0 - norm$ and $l_2 - norm$ are applied to the mean-variance portfolio selection model and the model is solved as a sequence of penalized subproblems to overcome the complexity of the resulted non-convex model. It is claimed in [73] that the proposed solution procedure can result in suboptimal solutions that are efficient compared to the local solutions resulting from applying other heuristics. Besides, the improvement of the out-of-sample performance of the mean-variance portfolio selection model is reported.

Another recent mean-variance portfolio selection model employing $l_0 - norm$ and $l_2 - norm$ is proposed in [6]. The ultimate optimization model in [6] is the reformulation of the mean-variance portfolio selection model as a constrained regression problem. Furthermore, the $l_0 - norm$ is defined by substitution of $x_i z_i$ instead of $x_i \leq z_i$ to imply the logical constraint $x_i = 0$ if $z_i = 0$. The reason they substituted a non-convex term instead of a linear constraint is to overcome the issue of weak relaxation while solving the sparsity constraint with the big-M method [74]. A solution procedure called "scalable outer-approximation algorithm" is proposed in [6] to achieve exact solution for the mixed-integer quadratic programming reformulation considering the stock market index $S\&P500$ with the power of resulting to certifiable optimality in less than 10 minutes. Also this powerful solution algorithm is shown to have the ability to handle large size mean-variance portfolio selection problems considering the data of $Wilshire5000$ index, which includes about 3500 stocks in the US stock market, with a solution gap of 2% from optimality. Through the various experiments with different problem sizes done in [6], this suggested reformulation and solution procedure is compared to the MISOCP reformulation of the mean-variance portfolio selection problem proposed in [40] and [41] with the modification to the diagonal matrix ($D \succ 0$) and defining it as $\lambda\mathbf{I}$ which is suggested in [75] and [76]. It is reported that although the MISOCP reformulation of portfolio selection model is powerful in handling the problems with the size of about 500, it fails to compete with the power of mean-variance portfolio selection reformulation and the scalable outer-approximation algorithm in [6] to solve the problems with about 1000 or more stocks within a reasonable amount of time. Specifically, the "scalable outer-approximation algorithm", which is actually an enhanced cutting-plane method, accelerates its convergence by starting the solution procedure from a warm-start value for the integer variable z_i derived by a heuristic algorithm proposed in [77] and improved in [78] that uses the objective function gradient information. Besides, this solution algorithm benefits from employing the `lazy-constraint callbacks` available in `CPLEX`.

The research question that is aimed to be answered in this research study is whether investing in a sparse portfolio is an optimal decision for risk-neutral and risk-averse investors. To the best of our knowledge, there is not study in the literature of sparse portfolio optimization that specifically aims at evaluating the out-of-sample performance of the sparse portfolio selection models with presence of $l_0 - norm$ and squared $l_2 - norm$.

CHAPTER 3 RESEARCH METHODOLOGY

In this chapter, after discussing the mathematical reformulation for sparse mean-variance portfolio selection model, the out-of-sample performance study strategy for evaluating the performance of the sparse mean-variance portfolio selection model will be elaborated. Specifically, the aim in section 3.1 is to present the evolution of the sparse mean-variance portfolio selection model as a mixed integer second order cone program that will be used to conduct the numerical study in chapter 4. Furthermore, the importance and logic behind the out-of-sample performance study approach required for conducting empirical study in chapter 4 is presented.

3.1 Mathematical Formulation of the Sparse Portfolio Optimization

As reviewed in chapter 2, there have been numerous researches conducted to solve sparse portfolio selection model to handle real-world size problems. Also, different approaches to reformulate problem (2.1) to achieve sparse portfolios as a non-convex model by considering $l_0 - norm$ or a convex model by adding $l_1 - norm$ were discussed [1]. The focus in this research is to embed $l_0 - norm$ in the portfolio selection problem in order to have sparse optimal solutions and simultaneously enabling the decision maker to decide the degree to which the sparsity is desired.

As a recall, the sparse portfolio selection model with a single-objective function and $l_0 - norm$ aiming for stable and robust solutions can be defined as the following :

$$\begin{aligned} \min_{x \in \mathbf{R}^n_+} \quad & x^T \hat{\Sigma} x - \hat{\mu}^T x + \gamma ||x||_2^2 \\ s.t. \quad & e^T x = 1 \\ & ||x||_0 \leq k. \end{aligned} \tag{3.1}$$

The squared $l_2 - norm$ is considered in problem (3.1) to regularize the portfolio weights and mitigate the extreme portfolio weights [71], hence it helps to promote the out-of-sample performance of the model. The out-of-sample performance of the model and the impact of $l_2 - norm$ on it will be discussed in section 3.3.

When representing the cardinality constraint $||x||_0 \leq k$ by $\exists z \in \{0,1\}^n, x \leq z, e^T z \leq k$,

1. $l_1 - norm$ does not allow to exactly control the sparsity.

problem (3.1) can be rewritten as follows :

$$\begin{aligned} \min_{x \in \mathbf{R}^n_+, z \in \{0,1\}^n} \quad & x^T \hat{\Sigma} x - \hat{\mu}^T x + \gamma ||x||_2^2 \\ s.t. \quad & e^T x = 1 \\ & x \leq z \\ & e^T z \leq k. \end{aligned} \tag{3.2}$$

As mentioned in chapter 2, branch-and-bound is the preliminary solution method applied to solve the sparse mean-variance portfolio optimization model as a mixed-integer quadratic programming model. Since applying branch-and-bound algorithms to cardinality constrainted models results in weak relaxations coming from the big-M method related to the constraint $x \leq z$ [33], the idea of reformulating the sparse mean-variance portfolio selection model as a second order cone programming model and employing non-linear branch-and-bound algorithm was suggested in [34] to achieve tighter relaxations for second order cone constraints and, subsequently, to have tighter optimality bounds. Improving the suggested method in [34], non-linear branch-and-bound methods were proposed in [35] and [36] to handle up to 450 stocks in a sparse portfolio optimization problem and result in tighter second order cone relaxations and, therefore, better certifiable optimality.

Another approach to achieve tighter relaxations for sparse mean-variance portfolio selection model was proposed in [37] by developing perspective reformulation of the sparse mean-variance portfolio optimization problem and solve it by either generating perspective cuts or reformulating the perspective formulation as a mixed integer second order cone programming (MISOCP) model and solving it by off-the-shelf optimization softwares.

The focus of section 3.2 is on the evolution of perspective reformulation to benefit from the tighter relaxations and good quality solutions resulting from it in the empirical study of the sparse mean-variance portfolio selection problem which will be presented in chapter 4.

3.2 Perspective Reformulation of the Mean-Variance Portfolio Selection Model

It is well-known that when solving the mixed integer problem defined as :

$$\min_{x,z} \quad \{f(x) + c^T z : \;\; Ax + Bz \leq d, z \in \{0,1\}^n, x \in \mathbf{R}^n_+\}, \tag{3.3}$$

with the Branch & Cut (B&C) approach, the continuous relaxation of (3.3), requiring $z \in [0,1]^n$, is used and results in the lower bound of the objective function in (3.3). Then, adding cuts enhances the quality of the lower bound and results in a better determination of the convex hull related to the integer solutions. As mentioned before, using the linear relationship between the continuous and the integer variable ($x \leq z$) and using big-M technique lead to weak relaxations and, depending on the structure of the MIP problem, perspective reformulation for MIP problem and adding perspective cuts was discussed to be a successful approach to improve the quality of the lower bounds of the solution procedure [6]. In order to be able to reformulate a problem in a perspective form, the objective function has to be separable. Although the objective function of the sparse portfolio selection problem is not separable because of the covariance matrix which includes the relations of each stock with itself and the other stocks in the market, there are reformulation techniques to separate the objective function such that at least a part of the objective function can be used to develop the perspective reformulation and perspective cuts [37]. In order to make the objective function of the sparse portfolio selection model separable, it is suggested in [37] to extract a diagonal matrix, $D \succeq 0$, from the covariance matrix, $\hat{\Sigma}$, in the way that $\hat{\Sigma} - D$ be a positive semidefinite matrix, $\hat{\Sigma} - D \succeq 0$. Therefore, the sparse portfolio selection model with the separable objective function will be defined as :

$$\begin{aligned} \min_{x \in \mathbf{R}^n_+, z \in \{0,1\}^n} \quad & x^T(\hat{\Sigma} - D)x + x^T D x - \hat{\mu}^T x \\ s.t. \quad & e^T x = 1 \\ & x \leq z \\ & e^T z \leq k. \end{aligned} \tag{3.4}$$

The abovementioned reformulation facilitates the writing of the perspective reformulation and subsequently applying perspective cuts.

Generally, the perspective function for a function $f : \mathbf{R} \to \mathbf{R}$ is the function $\tilde{f} : \mathbf{R}^2 \to \mathbf{R}$ defined as :

$$\tilde{f}(z, x) = \begin{cases} 0 & if \;\; z = 0 \\ z f(\frac{x}{z}) & if \;\; z \geq 0 \\ \infty & if \;\; otherwise. \end{cases} \tag{3.5}$$

Hence, the perspective reformulation for the separable part of the sparse portfolio selection model defined in (3.4) can be retpresented as follows :

$$\begin{aligned} \min_{x \in \mathbf{R}_+^n, z \in \{0,1\}^n} \quad & x^T(\hat{\Sigma} - D)x + \sum_{i=1}^{n} z_i \left[d_{ii} (\frac{x_i}{z_i})^2 - \hat{\mu}_i (\frac{x_i}{z_i}) \right] \\ s.t. \quad & e^T x = 1 \\ & x \leq z \\ & e^T z \leq k, \end{aligned} \tag{3.6}$$

where $d_{ii} \geq 0$ is the i^{th} element in the diagonal of matrix D [37].

The perspective reformulation described in (3.6) has a highly non-linear objective function which subsequently is costly in terms of computations to be solved. To tackle this issue, there are two types of reformulations as a *Semi-Infinite Non-Linear Program* (Semi-Infinite NLP) which deals with adding perspective cuts or as a *Second Order Cone Program* (SOCP).

Based on theorem 1 in [37], perspective cuts are the set of linear inequalities that represent the epigraphs as :

$$\hat{f}(z_i, x_i) := inf\{v_i | v_i \geq \quad z_i \left[f(\bar{x}_i) - \bar{x}_i s \right] + x_i s \qquad \forall \ s \in \nabla f(\bar{x}_i), \forall \ \bar{x}_i \in [0,1], \ \forall \ i \in [n]\}. \tag{3.7}$$

Therefore, the sparse portfolio optimization model with perspective cuts can be reformulated as follows when considering $f(x_i) = d_i x_i^2 - \hat{\mu}_i x_i$:

$$\begin{aligned} \min_{x \in \mathbf{R}_+^n, z \in \{0,1\}^n, v \in \mathbf{R}^n} \quad & x^T(\hat{\Sigma} - D)x + e^T v \\ s.t. \quad & v_i \geq x_i \left(2 d_i \bar{x}_i - \hat{\mu}_i \right) - z_i \left(d_i {\bar{x}_i}^2 \right) \quad \forall \ \bar{x}_i \in [0,1], \ \forall \ i \in [n] \\ & e^T x = 1 \\ & e^T z \leq k. \end{aligned} \tag{3.8}$$

Since the separable part of the objective function in (3.4) had quadratic form, the reformulation in (3.8) results in a semi-infinite mixed integer non-linear model. The approach to solve problem (3.8) is to keep a small subset of constraints related to perspective cuts and solve the relaxed version of (3.8). Afterwards, the violated perspective cuts (constraints) will be added iteratively to the problem in order to achieve the feasible and optimal solution [39].

Another reformulation to benefit from tight relaxations resulting from perspective formulation in (3.6) is to reformulate it as a Mixed-Integer SOCP. This aim is achievable by refromulating the epigraph of the perspective fucntion in (3.6) as *conic inequalities*. Therefore, the following representation is required [39] :

$$
\begin{aligned}
\min_{x\in\mathbf{R}_+^n, z\in\{0,1\}^n, v\in\mathbf{R}_+^n} \quad & x^T(\hat{\Sigma}-D)x + diag(D)^T v - \hat{\mu}^T x \\
s.t. \quad & v_i \geq \frac{x_i^2}{z_i} \qquad \forall\ i \in [n] \\
& e^T x = 1 \\
& e^T z \leq k.
\end{aligned}
\tag{3.9}
$$

The first constraint in (3.9) can be rewritten as :

$$
x_i^2 - v_i z_i \leq 0 \qquad \forall\ i \in [n]\,, \tag{3.10}
$$

which represents a convex set described by non-convex functions and there is no guarantee to achieve the globally optimal solution while employing non-linear programming (NLP) solvers to solve the relaxation of the model. Instead, it is preferable to convert (3.10) to a convex constraint by the following mathematical manipulation :

$$
\sqrt{4x_i^2 + (v_i - z_i)^2} - (v_i + z_i) \leq 0 \qquad \forall\ i \in [n]\,. \tag{3.11}
$$

Set of constraints defined in (3.11) can also be shown as :

$$
\|\,(2x_i, v_i - z_i)^T\,\|_2 \leq (v_i + z_i) \qquad \forall\ i \in [n]\,. \tag{3.12}
$$

Constraint (3.10) is the set of rotated second order cone constraints and constraint (3.11) represents the set of second order cone constraints which both are equivalent [41]. Now, set of constraints in (3.12) can be substituted in (3.9) which results in :

$$
\begin{aligned}
\min_{x\in\mathbf{R}_+^n, z\in\{0,1\}^n, v\in\mathbf{R}_+^n} \quad & x^T(\hat{\Sigma}-D)x + diag(D)^T v - \hat{\mu}^T x \\
s.t. \quad & \|\,(2x_i, v_i - z_i)^T\,\|_2 \leq (v_i + z_i) \quad \forall\ i \in [n] \\
& e^T x = 1 \\
& e^T z \leq k.
\end{aligned}
\tag{3.13}
$$

The question that arises in this stage is how to determine matrix D as the separable part of the covariance matrix. As a simple approach, matrix D is defined in [37] by calculating the minimum eigenvalue (γ_{min}) of the covariance matrix ($\hat{\Sigma}$) and set $D = \gamma_{min} I$. Another technique applied in [38] and [39] to reflect more of the covariance matrix $\hat{\Sigma}$ in matrix D

to result in tighter relaxations is to solve an auxiliary semidefinite program. Determination of matrix D in terms of extracting a diagonal matrix that leads to a tighter continuous relaxation of the perspective reformulated model is improved in [42] by solving a semidefinite program. Recently, the idea of imposing a ridge regularizer to the objective function of the model and setting matrix D as γI is employed in [75] and [76]. Furthermore, the combination of the ridge regularizer and decomposition of the covariance matrix Σ to the diagonal matrix $D \succ 0$ and positive semidefinite matrix $\hat{\Sigma} - D \succeq 0$ is applied in [6].

In the empirical study conducted in chapter 4, the strategy in [75] and [76] is followed by defining matrix $D = \gamma I$, therefore the ridge regularizer is embeded in the objective function of (3.13). Parameter γ is called the regularizer parameter which its role will be explained in section 3.3.Furthermore, risk aversion parameter represented by α is considered in the objective function of the model to trade-off between the risk and the return of the sparse portfolios. Moreover, the non-separable part of the covariance matrix is factorized as $\hat{\Sigma} - D = X^T X$ to write the quadratic part of the objective function in the form of the squared $l_2 - norm$. Ultimately, the sparse mean-variance portfolio optimization problem under study will be formulated as a MISOCP model as follows :

$$
\begin{aligned}
\min_{x \in \mathbf{R}^n_+, z \in \{0,1\}^n, v \in \mathbf{R}^n_+} \quad & \alpha \|Xx\|_2^2 + \gamma e^T v - \hat{\mu}^T x \\
s.t. \quad & \| (2x_i, v_i - z_i)^T \|_2 \le (v_i + z_i) \quad \forall\, i \in [n] \\
& e^T x = 1 \\
& e^T z \le k.
\end{aligned}
\tag{3.14}
$$

Since x is allowed to take non-negative values, i.e., $x \in \mathbf{R}^n_+$, shortselling is prohibited in (3.14). Hence, in order to study the case that shortselling is allowed, problem (3.14) is modified as follows by setting x as a free variable :

$$
\begin{aligned}
\min_{x \in \mathbf{R}, z \in \{0,1\}^n, v \in \mathbf{R}^n_+} \quad & \alpha \|Xx\|_2^2 + \gamma e^T v - \hat{\mu}^T x \\
s.t. \quad & \| (2x_i, v_i - z_i)^T \|_2 \le (v_i + z_i) \quad \forall\, i \in [n] \\
& e^T x = 1 \\
& e^T z \le k \\
& -z \le x \le z
\end{aligned}
\tag{3.15}
$$

It is worth mentioning that each z_i as a binary decision variable excludes the i^{th} stock from the optimal sparse portfolio when takes value *zero* and includes the i^{th} stock in the optimal sparse portfolio when takes value *one*. Afterwards, if the i^{th} stock is included in the optimal sparse portfolio, $z_i = 1$, the continuous variable x_i can take its optimal weight in the interval $[0, 1]$ for the case that short-selling is prohibited and is able to take its optimal weight in the interval $[-1, 1]$ for the case that short-selling is allowed.

The empirical study of out-of-sample performance of the sparse mean-variance portfolio optimization problem that will be conducted in chapter 4 exploits the models presented in (3.14) and (3.15).

3.3 Discussion on Out-of-Sample Study

In order to employ the sparse mean-variance portfolio selection model to determine the optimal investment portions, determination of the mean and the covariance matrix is required. Since the actual values of the mean and the covariance matrix are not known, these model inputs are estimated by sample mean ($\hat{\mu}$) and sample covariance matrix ($\hat{\Sigma}$) under the assumption that the stock returns are normally distributed. In fact, sample mean and sample covariance matrix are the maximum likelihood estimators (MLE) for mean and covariance parameters calculated out of historical data. Theoretically, the maximum likelihood estimators are the most efficient estimators for the mean and the covariance matrix of the returns if they follow the normal (assumed) distribution. On the other hand, the true distribution of the returns often deviates from normality because of being a heavy-tailed distribution or having jumps. Consequently, plugging the maximum likelihood estimators $\hat{\mu}$ and $\hat{\Sigma}$ in the mean-variance portfolio optimization model affects the mean-variance portfolio optimization performance due to the estimation error in the input parameters. It is worth mentioning that the estimation error of the mean and the covariance matrix results in unstable optimal portfolio weights with values that fluctuate drastically when the investor reoptimizes the portfolio optimization model over time. Such instabilities are not desirable for the investors since they lead to large transaction costs. What is discussed reveals that the model performance in the sense of estimation error mitigation and stability of the portfolio weights over time are two important issues to be considered while optimizing a portfolio selection model.

Investigating the portfolio selection problem in the context of the statistical learning theory, such instabilities are caused by the over-fitting of the model. Over-fitting means that the model works suitably with respect to the sample of the data that its parameters are derived from while it fails to reliably perform and predict for new datasets [49]. Specifically, in statistical learning and machine learning, upper bounds that quantify the model expected generalization performance can be controlled using regularization. The error bounds decrease while the capacity of the problem to be fitted to the dataset is diminished. This mechanism leads to better generalization of the model which means the model can be used by various datasets [61], [62]. As discussed in section 2.6, one way to tackle the over-fitting issue and instability of the optimal portfolio weights caused by the estimation error of the mean and the covariance matrices is regularizing the model by employing $l_2 - norm$.

Looking into (3.14) and (3.15), the term $e^T v$ which represents $l_2 - norm$ is multiplied by regularization parameter γ which role is to give weight to the $l_2 - norm$ and to help to generalize the model and tackle over-fitting issue. One has to tune parameters like γ in order to obtain a sparse mean-variance model that performs optimally and in a stable manner. In other words, parameter γ can take different values and each of the values may result in a different performance of the model under study. Therefore, it is crucial to determine the value of γ meticulously to obtain a model that does not over-fit or under-fit but generalizes the sparse mean-variance model to achieve stable optimal portfolios over time. The parameters like γ, which values impact the model performance, are called $hyperparameters$. Model selection is the procedure to select one model among a set of candidate models by determining the model's hyperparameter. In model selection, the aim is to achieve optimal performance of the model under study and obtain predictions with minimum error by tuning the model's hyperparameter and as a result to avoid under-fitting or over-fitting issues. Employing model selection on the sparse mean-variance portfolio optimization model defined in (3.14) and (3.15) focuses on minimizing the generalization error of the model and determination of the hyperparameter γ in the way that results in reliable investment predictions for the next period.

In model selection, the performance of the set of candidate models is evaluated not only based on the quality of the results derived from the employed data to estimate the parameters, which is called the in-sample-data, but it is also evaluated based on the quality of the results gained by the new dataset called out-of-sample data. More specifically, out–of–sample performance study is one of the main and commonly applied approaches to do model selection and error estimation. It is worth mentioning that out-of-sample performance study is useful when a large dataset is available and the dataset under study in this research also has this characteristic.

The widely applied out-of-sample performance approaches are leave-one-out cross-validation, k-fold cross-validation, time series cross-validation. The idea behind the out-of-sample techniques is to resample the available dataset to make three independent datasets called *training*, *validation* and *testing* sets. The training set is used to build up the model. The validation set is applied to tune the hyperparameters in the way the generalization error of the model be minimized. Also, the test set is useful for evaluating the model performance in an unbiased manner. Because the training, validation and testing datasets are the independent divisions of the original dataset, they all have the same distribution [79].

Since in this research study the order of the stocks' returns are important and the dataset is a time-related one, the out-of-sample performance study approach suitable to apply is *rolling horizon* cross-validation which is also employed in [57] and [50]. Generally, the rolling horizon cross-validation, also known as time series cross-validation, is used to evaluate the stability of model's parameters over time. Therefore, the rolling horizon cross-validation is an aligned out-of-sample study strategy with the purpose of this research study.

3.3.1 Rolling Horizon Training Procedure of Best Sparsity Degree

The purpose of this section is to explain how the best sparsity degree, k_{best}, can be determined by employing a rolling horizon strategy. In turn, this can allow to find out if investing on sparse optimal portfolios is an optimal investment strategy for a risk-neutral and risk-averse decision maker.

We employ a rolling horizon training procedure that exploits $L-H-1$ experiments where L is the total number of periods in the training data set and H is the size of the history used to estimate sample mean, $\hat{\mu}$, and sample covariance, $\hat{\Sigma}$. Given a fixed level of risk aversion α, one can roll through the $L-H-1$ periods of investments, to monitor the performance of the optimal portfolios produced under different sparsity level k. Specifically, in each period :

1. the mean and the covariance matrix can be estimated from the latest H historical returns ;
2. an optimal portfolio for each sparsity level k can be established ;
3. the realized return of these portfolios can be measured in the following period.

Note that in this procedure, as we roll forward in time, the new portfolios will account for an updated estimate of the mean and the covariance matrix based on the new H most recently observed returns. Finally, given the set of $L-H-1$ produced portfolio returns for each sparsity levels, one can identify the empirically best performing k (for the fixed risk aversion level α) by comparing the different return distribution using a measure that is coherent with the mean-variance minimization objective of the portfolio optimization problems in (3.14) and (3.15). Inspired by the procedure used in [49], we use the trajectory-wise mean-variance trade-off defined as :

$$EM = \alpha \; \bar{\sigma}^2 - \bar{\mu} \tag{3.16}$$

where $\bar{\mu}$, which represents the out-of-sample mean, and $\bar{\sigma}^2$, which defines the out-of-sample variance, are calculated by the following formulations :

$$\begin{aligned} \bar{\mu} =& \frac{1}{L-H} \sum_{t=H}^{L-1} x_t^{*T} r_{t+1} \\ \bar{\sigma}^2 =& \frac{1}{L-H-1} \sum_{t=H}^{L-1} (x_t^{*T} r_{t+1} - \bar{\mu})^2 \end{aligned} \tag{3.17}$$

In (3.17), x_t^* is the optimal portfolio produced at period t and is multiplied by the returns of the next period, r_{t+1}, to see how much is the return of the portfolio for period $t+1$ with respect to the optimal allocation x_t^*. This procedure is done $L-H-1$ times and the resulting portfolio returns are averaged to estimate $\bar{\mu}$. A similar logic is followed to calculate $\bar{\sigma}^2$. Since the aim is to minimize the objective function of the model, the smaller the evaluation measure, the better the performance of the model. Therefore, for each type of decision maker associated to an α value, the best sparsity degree, k_{best}, can be selected as the one that leads to the lowest value of the evaluation measure in (3.16).

As a final remark, it is worth mentioning that EM consists of an unbiased estimator of the instantaneous mean-variance trade-off when the portfolio is kept fixed and the stock return vectors r_t, with $t = H, \ldots, L-1$, are independently and identically distributed. While these conditions do not occur in a real market, it certainly serves as a good motivation for using EM.

3.3.2 Rolling Horizon Testing Procedure of Sparsity Degree

A procedure similar to the one discussed in section 3.3.1 can be followed on the test set to evaluate the out-of-sample performance of the sparse mean-variance portfolio optimization models in (3.14) and (3.15) with the trained sparsity level. In this procedure, L becomes the total available periods in the test set. Moreover, the same evaluation measure as defined in 3.16 will be used in the study of out-of-sample performance of the models in (3.14) and (3.15). The difference is that, for each α, only k_{best} will be employed and the evaluation measure calculated for each risk aversion level will be compared to the evaluation measure calculated for the case that no sparsity is imposed on the model.

CHAPTER 4 NUMERICAL STUDIES

In this chapter, the empirical study of out-of-sample performance is conducted.

4.1 The Data under Study

Data employed in this empirical study is collected from the stock market which consists of 475 companies that compose $S\&P$ index. The time span that the stock prices are collected starts from 1993/09/01 and ends in 2014/08/12 which is equivalent to 1114 weekly prices.

In statistical learning, it is common to handle missing data in the dataset either by deletion of the feature (the stock) that has missing value, imputing the missing value by the mean/median calculated from the prices of that specific stock in other periods or predicting the missing value by statistical techniques like regression. The deletion approach is used to handle the missing data in this empirical study. After deleting the stocks with missing values, there are 334 stocks left.

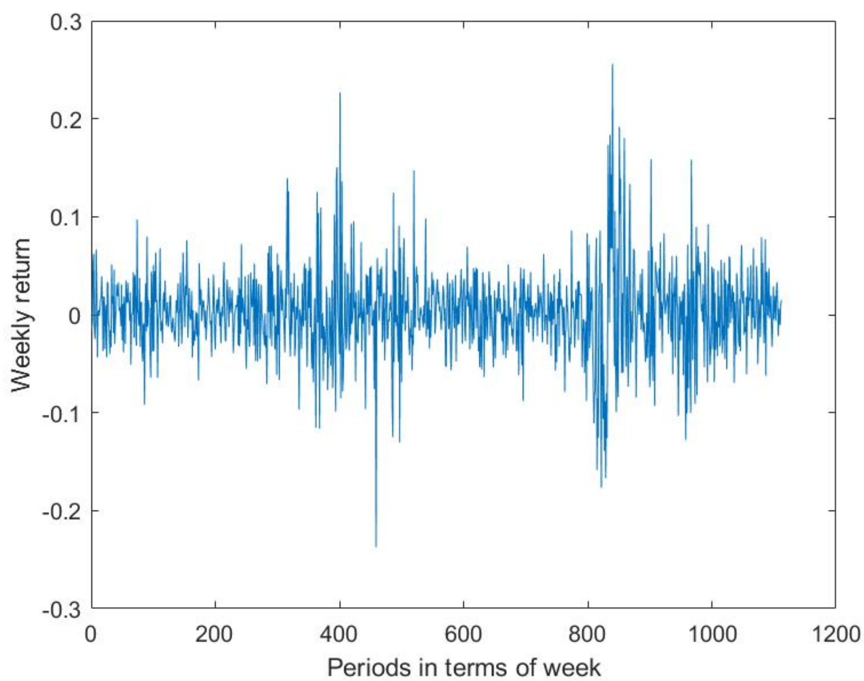

Figure 4.1 The evolution of the weekly return of a randomly chosen stock (1993/3 to 2014/7)

The standard formulation followed to calculate stock returns is $r_t = (p_t - p_{t-1})/p_{t-1}$, where r_t

and p_t denote the return and the price of stock at time t, respectively. The result of converting stock prices over time to stock returns is depicted in Figure 4.1. As mentioned in 3.3, the sample mean and sample variance as maximum likelihood estimators are the estimators designed under the assumption that the data is independent identically distributed (i.i.d). As it is depicted in Figure 4.1, the variance of the returns varies over time which means that the distribution of the returns is not i.i.d and stationary. Also, a short history of the returns will be used for computing $\hat{\mu}$ and $\hat{\Sigma}$ which means the non-stationarity of the stock returns affects the values of these estimators. Therefore, the sample mean and sample variance estimators are naïve estimators in such cases and they could be improved by the techniques that accounts for the non-stationary data [80].

Now that the data is prepared by removing stocks with missing prices and converting the stock prices to stock returns, the next step will be to conduct the model selection.

4.2 Model Selection

As discussed in section 3.3, the sample mean ($\hat{\mu}$) and sample covariance matrix ($\hat{\Sigma}$) as the maximum likelihood estimators (MLE) for the mean and the covariance parameters are calculated out of historical data under normality assumption. Due to deviation of the stock returns distribution from normality, there will be estimation error in determination of the sparse mean-variance portfolio optimization input parameters. One common approach to minimize the effect of estimation errors related to the input parameters is to regularize the model by employing regularization term, which in this research study is done by embedding $l_2 - norm$ in the sparse mean-variance portfolio selection model (3.14) and (3.15). Determination of regularization parameter γ will be discussed in 4.2.1.

4.2.1 Hyperparameter Tuning

To avoid large transaction costs which occur due to instability of optimal portfolio weights over time and, also, to mitigate the effect of error in estimation of the model inputs, the model is regularized by embedding $l_2 - norm$ in its objective function, which also helps with the generalization power of the model. The regularization parameter γ as a hyperparameter needs to be determined with caution to avoid over-fitting and under-fitting. Therefore, different values are associated to regularization parameter, $\gamma \in \{0.01,\ 0.1,\ 1,\ 10,\ 100,\ 1000\}$, to see how the investment proportions will be affected by different values of γ. The outcome of conducting sensitivity analysis on γ under the assumption that short-selling is prohibited is visualized in Figure 4.2. It is worth mentioning that the number of stocks considered to optimize the sparse mean-variance portfolio selection model defined in (3.14) is $n = 100$ and no sparsity restriction ($k = n = 100$) is imposed to determine the optimal investment proportions (x^*) associated to each different γ values.

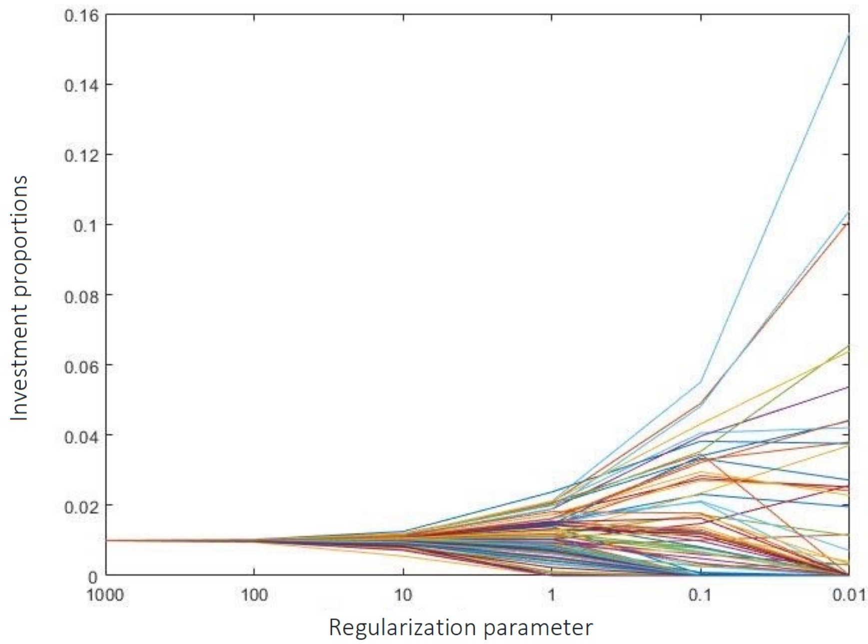

Figure 4.2 The investment proportions versus different values of regularization parameter (γ)

As depicted in Figure 4.2, all n stocks are included in the optimal portfolio in the way that all the investment proportions are equal when γ takes large values like 1000 and 100. Therefore, there is the risk that the sparse mean-variance portfolio optimization models in (3.14) and (3.15) under-fit the training dataset. On the other hand, few stocks are included in the optimal portfolios each with very different weights when γ is set to small values such as 0.01 and 0.1. Hence, there is the chance that the models in (3.14) and (3.15) over-fit the training data. Therefore, it seems reasonable to set γ to values like 1 in order to achieve a reasonable balance of performance for the sparse mean-value portfolio optimization model in this study. Upon conducting the same analysis under the assumption that short-selling is allowed, the regularization parameter takes a different value ($\gamma = 0.2$).

In section 4.2.2, γ is set to the values 1.0 and 0.2 since these values seem reasonable with respect to the distribution of the weights under the assumption that short-selling is prohibited and allowed, respectively. Note that a sensitivity analysis for determining hyperparameter γ could be an interesting additional study to perform, which is out of the scope of this study.

4.2.2 Rolling Horizon Training and Testing Procedures for Sparsity Degree

The performance of the set of candidate models in model selection is evaluated by assessing the quality of the results that are derived from employed data to estimate parameters which is called in-sample study. Also, evaluating the quality of the results of the model by employing new data set, which is called out-of-sample performance, is a crucial step in model selection. As mentioned in section 3.3, the out-of-sample performance study is useful when a large dataset is available and the dataset in this research study has this characteristic. The stock market returns dataset is divided into two subsections as training and testing sets containing 70% and 30% of the periods out of all periods in the original dataset, respectively. In other words, the weekly returns from 1993/09/01 to 2008/02/22 constitute the training set and the weekly returns from 2008/03/04 to 2014/08/12 constitute the the testing set.

The rolling horizon training and testing procedures explained in 3.3.1 and 3.3.2, are applied to assess the stability of the optimal portfolios that result from the sparse mean-variance portfolio selection model over time.The size of the history used to estimate sample mean, $\hat{\mu}$, and sample covariance, $\hat{\Sigma}$, is set to $H = 50$ weeks, which is almost equivalent to one year. Experiments are done by selecting randomly $n = 100$ stocks out of the total stocks in the dataset, $N = 334$. Also, this set of 100 randomly chosen stocks is kept unchanged while conducting training and testing procedures.

Furthermore, α as the coefficient applied to trade-off between the risk and the return of the sparse portfolios takes value $\alpha \in \{0.1, 1, 10, 100\}$ to reflect the degree to which the investors might be risk-averse. Since α is multiplied to the term that represents risk in the objective functions of (3.14) and (3.15), the value such as $\alpha = 0.1$ reflects that the investor is nearly risk-neutral since the impact of risk in choosing the optimal sparse portfolio is reduced by multiplying a constant that is less than 1. Following the same logic, $\alpha \in \{10, 100\}$ is considered for risk-averse investors.

The key parameter that is considered for conducting sensitivity analysis is the sparsity parameter k which takes 6 logarithmically spaced values between 1 and 100, $k \in \{1, 3, 6, 16, 40, 100\}$.

4.2.3 In-Sample Performance Study

Now that the values of the parameters are set and explained, the next step will be to describe how the experiments are done based on rolling horizon testing strategy. After choosing $n = 100$ stocks randomly, the sample mean $\hat{\mu}$ and $\hat{\Sigma}$ are calculated using a time window of 50 weeks, $H = 50$. Furthermore, the sample covariance matrix is factorized as $\hat{\Sigma} = X^T X$ in order to be fed to the objective function of the MISOCP models described in (3.14) and (3.15). Per each value that the risk-aversion parameter takes ($\alpha \in \{0.1,\ 1,\ 10,\ 100\}$), the chosen data will be used to solve the sparse mean-variance portfolio selection models in (3.14) and (3.15) while sparsity parameter takes its values as $k \in \{1,\ 3,\ 6,\ 16,\ 40,\ 100\}$. The resulting optimal selected stocks (z^*) and the optimal portfolio weights (x^*) given to each stock that is included in the optimal portfolio will be available upon optimizing models in (3.14) and (3.15). The optimal portfolio weights, x^*, will be multiplied by the stocks returns of the next period of the chosen time window and the outcome represents the *realized portfolio return* for that period. The next step will be to update the data for the next experiment by eliminating the first week of the $|n| \times |H|$ matrix and include the next week stock returns. This procedure has been followed 50 times for each combination of $\gamma = 1$, $\alpha \in \{0.1,\ 1,\ 10,\ 100\}$ and $k \in \{1,\ 3,\ 6,\ 16,\ 40,\ 100\}$ for the case that short-selling is prohibited. For the case that short-selling is allowed, the regularization parameter takes a different value ($\gamma = 0.2$), but the rest of the parameters take the same values as mentioned in case that short-selling is not allowed. Then, the evaluation measure values are calculated out of estimated returns for next periods which are resulted from the experiments for each parameter combination by computing the mean ($\bar{\mu}$) and the variance ($\bar{\sigma}$) and substituting in (3.16). Since the models in (3.14) and (3.15) aim to minimize their objective functions, the sparsity parameter that results in the lowest evaluation measure value for each combination of α's and k's will be considered as the best sparsity value, k_{best}.

Figure 4.3, Figure 4.4, Figure 4.5 and Figure 4.6 show the change in the evaluation measure value defined in (3.16) when the sparse portfolio selection model in (3.14) for the case that short-selling is prohibited is restricted to different values of sparsities k's, and optimized for investors with different levels of risk-aversion. The trends in these figures indicate that the evaluation measure takes its minimum value when the sparsity parameter k, takes a value less than the total number of stocks used in each experiment, $n = 100$.

Table 4.1 Validation experiment results ($\gamma = 1$)

α	*Best Sparsity* (k_{best})
0.1	3
1	6
10	6
100	40

The results of the rolling horizon training procedure for the case that short-selling is prohibited are reported in Table 4.1. The best sparsity degrees, k_{best}'s, that are resulted per different degree of risk-aversion show that the optimal portfolios resulted from the sparse mean-variance portfolio selection model defined in (3.14) contains less number of stocks comparing to the total stocks available in each experiment, $n = 100$. More precisely, it is optimal for the risk-neutral decision maker, who reduces the magnitude of the risk in the portfolio selection model by setting $\alpha = 0.1$, to build a portfolio of 3 stocks to achieve the minimum risk and maximum return. As reported in Table 4.1, the number of stocks in the optimal sparse portfolio increases when the risk-aversion parameter increases. In other words, the sparse portfolio selection model in (3.14) results in the optimal portfolios that include more stocks when the investors are more risk-averse. This result which is gained from the sparse portfolio selection model in (3.14) is aligned with the fact that the risk-averse investors tend to invest in portfolios that include more stocks to reduce the risk of their investments.

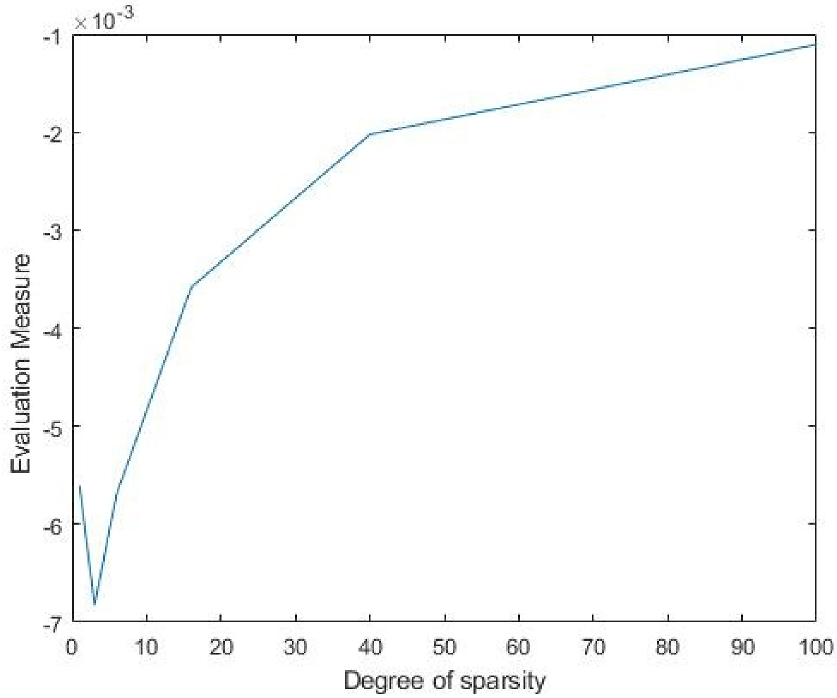

Figure 4.3 Training and validation result for α=0.1 and γ=1 (short-selling is prohibited)

For the case that short-selling is allowed, the variability of evaluation measure with respect to different sparsity degree values are depicted in Figure 4.7, Figure 4.8, Figure 4.9 and Figure 4.10. The trends in these figures show that the evaluation measure takes its minimum value when the sparsity parameter k, takes the value less than the total number of stocks used in each experiment, $n = 100$, except for the case that the risk-aversion coefficient is high, $\alpha = 100$. Also, Table 4.2 summarizes the best sparsity values associated with each α value. In case that short-selling is allowed, optimizing the sparse mean-variance portfolio selection

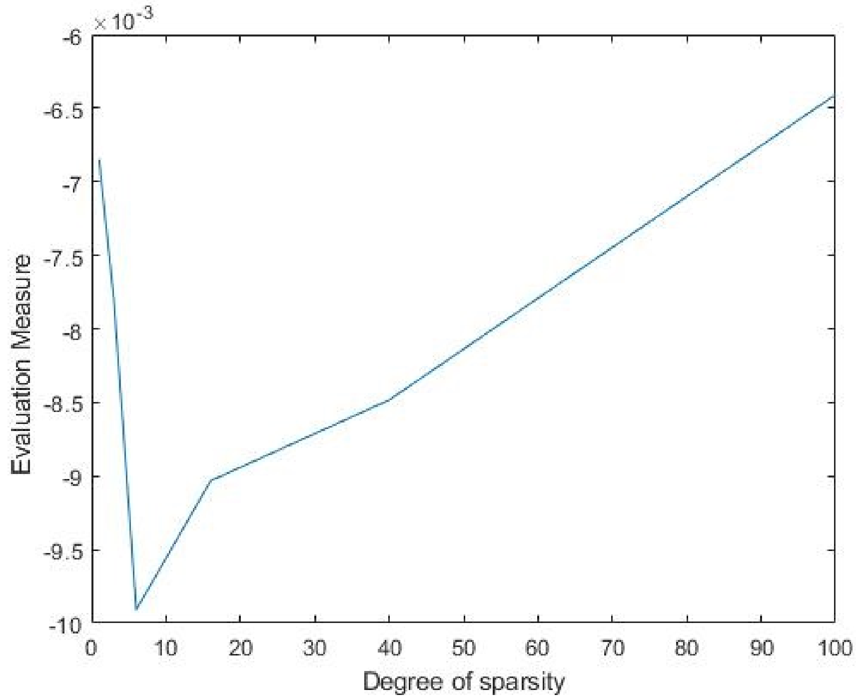

Figure 4.4 Training and validation result for α=1 andγ =1 (short-selling is prohibited)

model defined in (3.15) results in the minimum value for the evaluation measure when the number of stocks included in the optimal portfolio is less than the total number of stocks in each experiment, n=100, except for the case that the risk-aversion degree is set to 100. This means that when the investor is highly risk averse, she prefers to include all the stocks in her portfolio by short-selling some of the stocks and adding the fund gained out of short-selling to the available fund that she has and invest the aggregation in other stocks. Therefore, all stocks with negative and positive proportions will be included in the optimal portfolio of the investor with risk-aversion of 100. It is worth mentioning that Table 4.2 indicates that k_{best} increases as the investors get more risk averse.

The next step is to confirm whether the model is well generalized for the datasets that are new to the model or not.

Table 4.2 Validation experiment results ($\gamma = 0.2$)

α	*Best Sparsity* (k_{best})
0.1	3
1	3
10	6
100	100

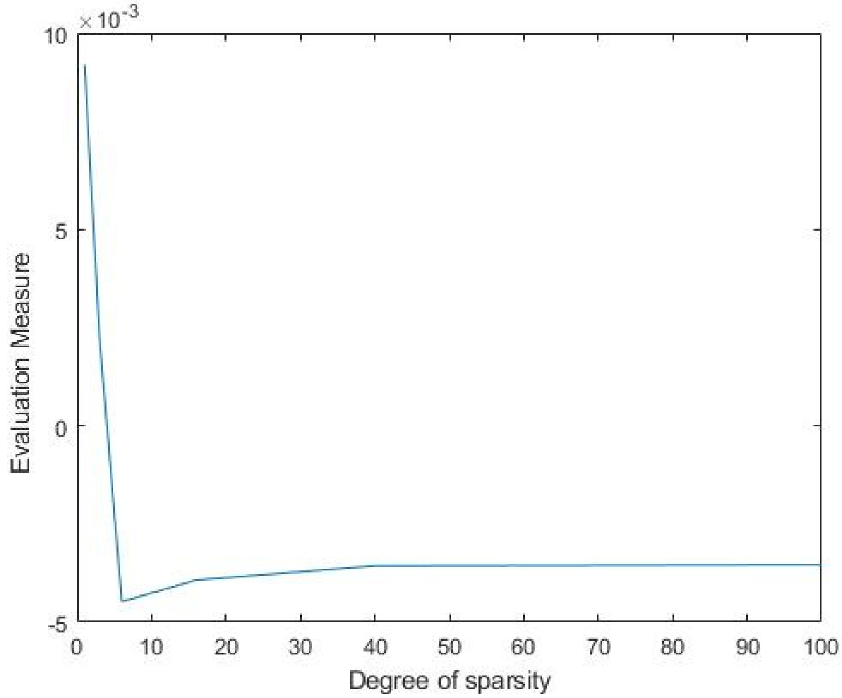

Figure 4.5 Training and validation result for α=10 and γ=1 (short-selling is prohibited)

4.2.4 Out-of-Sample Performance Study

Now that the results from the training step are in hand, the testing dataset will be used to calculate the evaluation measure for the experiments that will be done in the study of out-of-sample performance. It is worth reminding that the aim of out-of-sample performance study is to confirm if the models in (3.14) and (3.15) are well generalized and do not over-fit with respect to the γ values that are selected for them. In the out-of-sample performance study, the evaluation measure calculated per experiment that are done for each combination of α's and k_{best}'s is compared to the evaluation measure calculated under the assumption that the sparsity restrictions in (3.14) and (3.15) are relaxed which means setting $k = n = 100$.

As discussed in section 4.2.3, the training experiments determined the sparsity degrees that lead to the minimum risk and maximum return for each investors with different levels of risk-aversion with respect to the tuned hyperparameter of the model γ. The next step is to see whether the sparse mean-variance portfolio selection models in (3.14) and (3.15) can generalize the results in training step for a new dataset like the testing dataset or not.

In the out-of-sample performance study step, the input parameters $\hat{\mu}$ and $\hat{\Sigma}$ are calculated using the historical data in the testing dataset and have been fed to the models in (3.14) and (3.15). Also, γ, H, k's and α's have the same value(s) employed to conduct the trai-

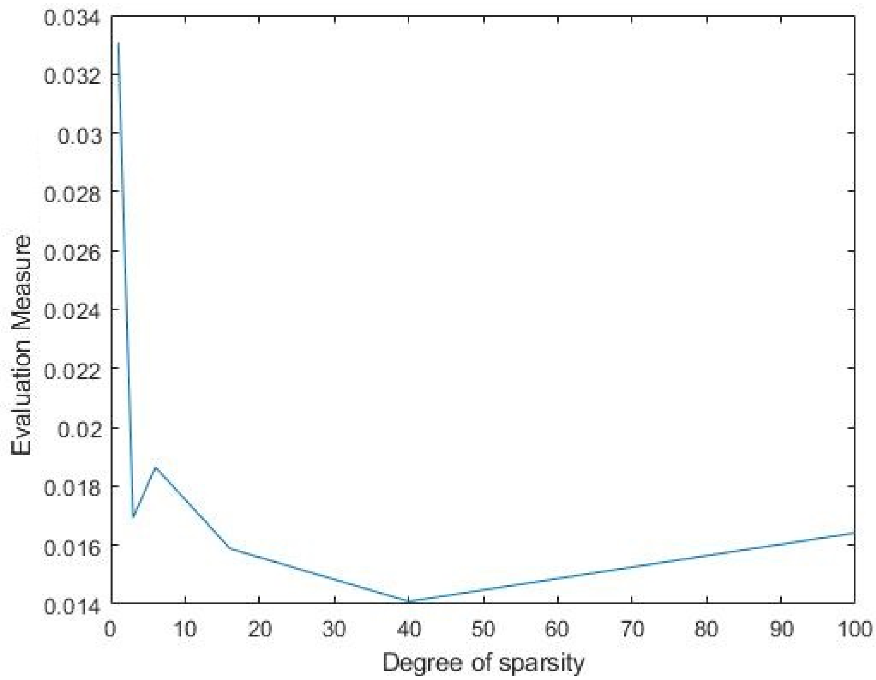

Figure 4.6 Training and validation result for α=100 and γ=1 (short-selling is prohibited)

ning experiments. In out-of-sample study, the evaluation measure is calculated using (3.16) for best sparsity degrees that are determined in training step and also for k=n=100. The results are summarized in Tables 4.1 and 4.2. The aim is to see if the results of the testing step substantiates that the k_{best}'s leads achieving a smaller evaluation measure value than the evaluation measure that is calculated under no sparsity restriction, k=n=100. For this purpose, the rolling-horizon out-of-sample study strategy as explained in 3.3.2 is followed.

The results of the out-of-sample performance study are summarized in Table 4.3, for the case that short-selling is prohibited, and Table 4.4, for the case that short-selling is allowed.

As reported in Table 4.3 and Table 4.4, the evaluation measure takes smaller values for different types of investors comparing to the case that there is no sparsity restriction, k=100. This confirms that the sparse mean-variance portfolio selection models in (3.14) and (3.15) are well generalized by setting hyperparameter γ=1, for the case that short-selling is prohibited, and $\gamma = 0.2$, for the case that short-selling is allowed. As the result, the models do not over-fit the datasets that the models are trained with, which is what was desired to achieve in the study of the out-of-sample performance of the models (3.14) and (3.15).

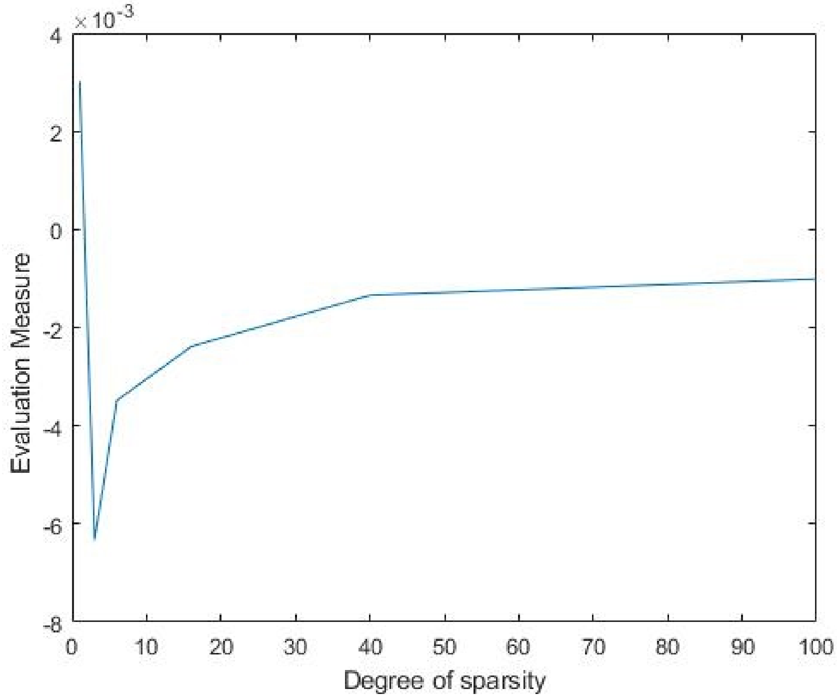

Figure 4.7 Training and validation result for α=0.1 and γ=0.2 (short-selling is allowed)

Table 4.3 Performance on test set, no short-sell

	$\alpha = 0.1$	$\alpha = 1$	$\alpha = 10$	$\alpha = 100$
EM for Sparse Portfolio Selection Model	-0.0011	-0.0242	0.0035	0.0305
EM for Non-sparse Portfolio Selection Model	-0.0005	0.0002	0.0060	0.0350

4.2.5 Discussion on Model Selection Procedure Implementation

The model selection procedure and out-of-sample performance evaluation procedure are implemented in MATLAB R2017b employing the YALMIP toolbox. Moreover, CPLEX 12.8.0 is used as the solver for optimizing the MISOCP models defined in (3.14) and (3.15). Using default settings of CPLEX 12.8.0, optimality and feasibility tolerance of 10^{-6} and integrality tolerance of 10^{-5} were considered for the solution schemes. The longest and shortest computational time for solving MISCOCP model in (3.14) were 4.6 and 113.7 seconds, respectively.

Table 4.4 Performance on test set, with short-sell

	$\alpha = 0.1$	$\alpha = 1$	$\alpha = 10$	$\alpha = 100$
EM for Sparse Portfolio Selection Model	-0.0043	-0.0026	0.0030	0.0198
EM for Non-sparse Portfolio Selection Model	-0.0011	-0.0002	0.0044	0.0198

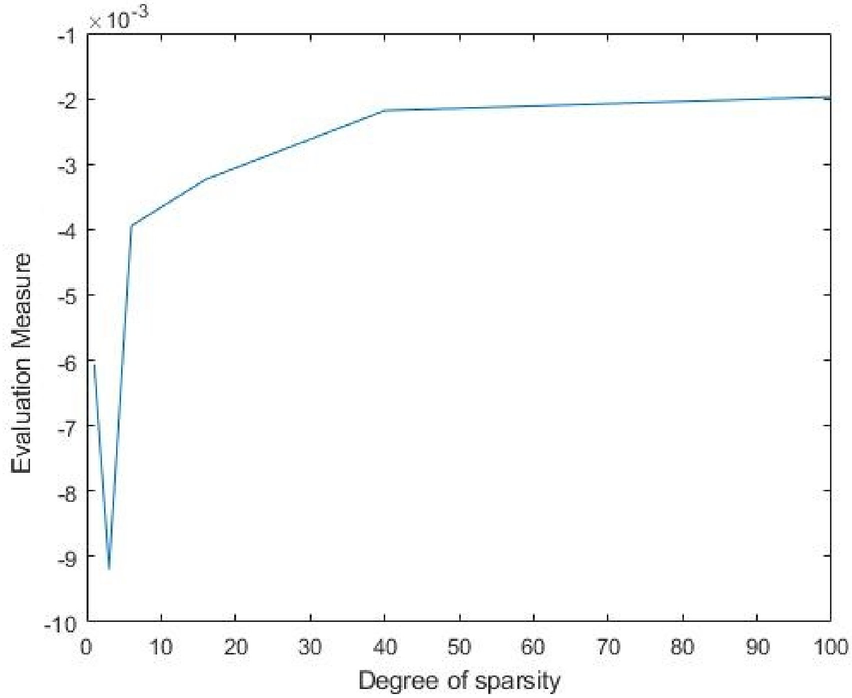

Figure 4.8 Training and validation result for α=1 and γ=0.2 (short-selling is allowed)

Also, for solving MISCOCP model in (3.15), the longest and shortest computational time for were 5.4 and 118.9 seconds, respectively. It is observed that the computational time increases as the sparsity degree k decreases which means the computational time is the shortest for the non-sparse portfolio optimization model. The reason is that the binary decision variable z is redundant when $k = n$ due to inclusion of all stocks in the optimal non-sparse portfolio. It can be inferred that solving the MISOCP models in (3.14) and (3.15) is computationally costly because of the binary decision variable z.

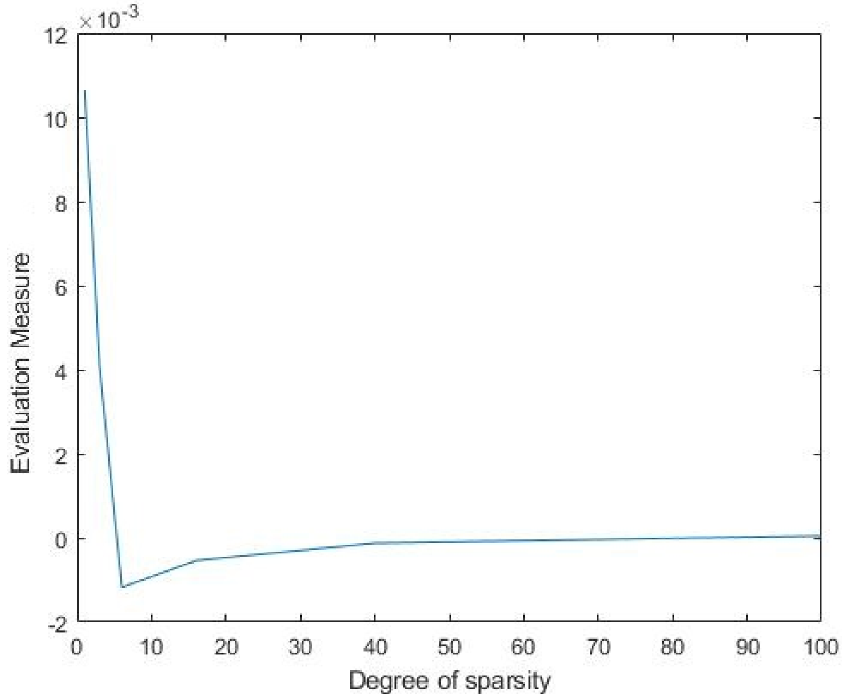

Figure 4.9 Training and validation result for α=10 and γ=0.2 (short-selling is allowed)

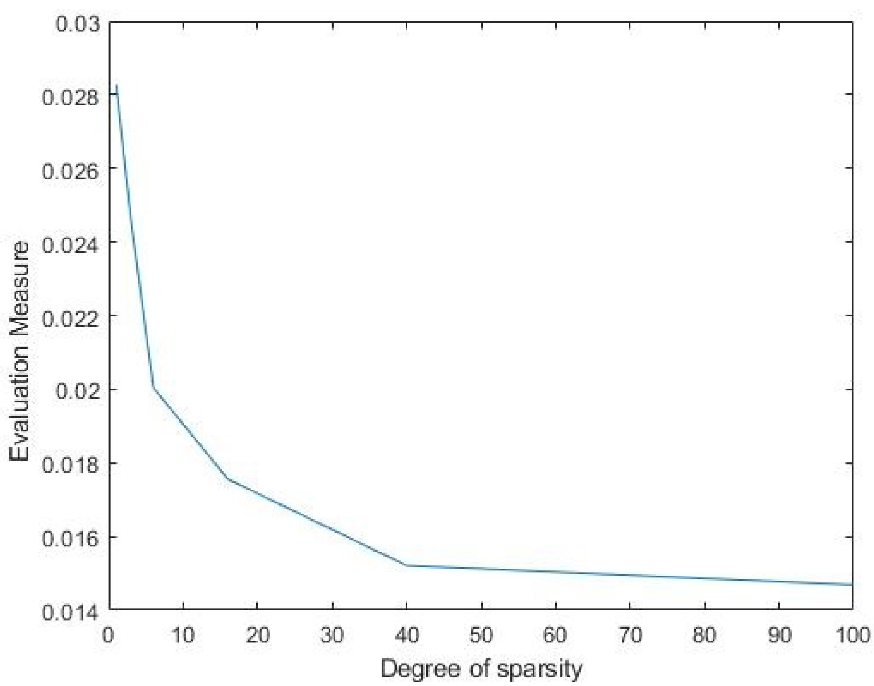

Figure 4.10 Training and validation result for α=100 and γ=0.2 (short-selling is allowed)

CHAPTER 5 CONCLUSION AND RECOMMENDATIONS

In this chapter a brief summary of what was done in this research study will be presented. Furthermore, the limitations of our study will be discussed and some avenues for future studies will be suggested.

5.1 Summary of Works

We tried to answer the question whether investing in sparse portfolios is an optimal investment decision compared to the situation that all available stocks are included in the investment portfolio. To achieve our goal, risk-neutral and risk-averse investors were considered in this research study since it is more likely that these types of investors tend to minimize the risk of their investments by including all available stocks in their investment portfolios.

In order to avoid model overfitting and underfitting issues and improving the generalization power of the sparse portfolio optimization model under study, squared $l_2 - norm$ was considered in the objective function of the model as regularizer and the coefficient that regulates the magnitude of its impact was tuned by following model selection procedure. Moreover, the dataset that contained the stock market weekly prices for 1114 weeks were divided into three independent datasets called training, validation and testing datasets. The training dataset was used to build up the model and the validation dataset was used to tune the sparse portfolio selection models' hyperparameters so that generalization error of the model be minimized. Furthermore, testing dataset helped to evaluate the models' out-of-sample performance in an unbiased manner. It is worth mentioning that rolling-horizon cross-validation approach was applied to evaluate the stability of the optimal sparse portfolios over time since we had a time-related dataset.

The results of our study confirmed that the sparse portfolios were optimal investment decisions for risk-neutral and risk-averse decision makers when short-selling was not allowed. In the scenario that short-selling is allowed, risk-neural investors' optimal decision was to invest in sparse portfolios. But the optimal investment decision for highly risk-averse investors was to short-sell some of the stocks and add the fund gained out of short-selling to the available fund that such investors has and invest the aggregation in other stocks.

5.2 Limitations of the Study

We conducted the empirical study for 100 stocks out of all 334 available stocks that we had in our dataset. The reason was that conducting the experiments for all of these stocks was time consuming due to the nature of integer programming models. One might be able to enhance the exact solution procedure of the sparse model or reformulate it in a way that makes her able to use large datasets and achieve the optimal results in a reasonable amount of time.

5.3 Future Research

For future researches, other risk measures such as conditional value-at-risk, entropic value-at-risk, etc can be substituted with the variance of the stock returns to see if investing in a sparse portfolio derived by employing other risk measures is an optimal investment decision for the investors. Furthermore, stocks' transaction costs can be embedded in the sparse model. Also, boundary constraints which restrict the investment proportion of each stock can be considered in the sparse portfolio selection model.

REFERENCES

[1] C. B. Kalayci, O. Ertenlice et M. A. Akbay, “A comprehensive review of deterministic models and applications for mean-variance portfolio optimization,” *Expert Systems with Applications*, 2019.

[2] T.-J. Chang *et al.*, “Heuristics for cardinality constrained portfolio optimisation,” *Computers & Operations Research*, vol. 27, nº. 13, p. 1271–1302, 2000.

[3] Y. L. T. Silva, A. B. Herthel et A. Subramanian, “A multi-objective evolutionary algorithm for a class of mean-variance portfolio selection problems,” *Expert Systems with Applications*, vol. 133, p. 225–241, 2019.

[4] H. Markowitz, “Portfolio selection j. finance,” 1952.

[5] M. H. Moo, “Portfolio selection,” *J. Finance*, vol. 7, nº. 1, p. 77–91, 1952.

[6] D. Bertsimas et R. Cory-Wright, “A scalable algorithm for sparse portfolio selection,” *arXiv preprint arXiv :1811.00138*, 2018.

[7] D. Bertsimas, C. Darnell et R. Soucy, “Portfolio construction through mixed-integer programming at grantham, mayo, van otterloo and company,” *Interfaces*, vol. 29, nº. 1, p. 49–66, 1999.

[8] D. Bertsimas *et al.*, “Portfolio construction through mixed integer programming,” 1998.

[9] M. G. Young, *The complete guide to selling stocks short : Everything you need to know explained simply.* Atlantic Publishing Company, 2010.

[10] S. Still et I. Kondor, “Regularizing portfolio optimization,” *New Journal of Physics*, vol. 12, nº. 7, p. 075034, 2010.

[11] V. Ranković *et al.*, “The mean-value at risk static portfolio optimization using genetic algorithm,” *Computer Science and Information Systems*, vol. 11, nº. 1, p. 89–109, 2014.

[12] I. Kondor, S. Pafka et G. Nagy, “Noise sensitivity of portfolio selection under various risk measures,” *Journal of Banking & Finance*, vol. 31, nº. 5, p. 1545–1573, 2007.

[13] T.-J. Chang, S.-C. Yang et K.-J. Chang, “Portfolio optimization problems in different risk measures using genetic algorithm,” *Expert Systems with applications*, vol. 36, nº. 7, p. 10 529–10 537, 2009.

[14] D. Jadhav et T. Ramanathan, “Portfolio optimization based on modified expected shortfall,” *Studies in Economics and Finance*, 2019.

[15] A. Goel et A. Sharma, “Mixed value-at-risk and its numerical investigation,” *Physica A : Statistical Mechanics and its Applications*, vol. 541, p. 123524, 2020.

[16] L. W. Hoe *et al.*, “An empirical comparison of different risk measures in portfolio optimization,” *Business and Economic Horizons*, vol. 1, nº. 1, p. 39–45, 2010.

[17] R. Mansini, W. Ogryczak et M. G. Speranza, “Twenty years of linear programming based portfolio optimization,” *European Journal of Operational Research*, vol. 234, nº. 2, p. 518–535, 2014.

[18] B. C. Barroso *et al.*, “Composition of investment portfolios through a combinatorial multiobjective optimization model using cvar,” dans *2017 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2017, p. 1795–1802.

[19] H. Babazadeh et A. Esfahanipour, “A novel multi period mean-var portfolio optimization model considering practical constraints and transaction cost,” *Journal of Computational and Applied Mathematics*, vol. 361, p. 313–342, 2019.

[20] R. Sun *et al.*, “Improved covariance matrix estimation for portfolio risk measurement : A review,” *Journal of Risk and Financial Management*, vol. 12, nº. 1, p. 48, 2019.

[21] K. Liagkouras et K. Metaxiotis, “A new efficiently encoded multiobjective algorithm for the solution of the cardinality constrained portfolio optimization problem,” *Annals of Operations Research*, vol. 267, nº. 1-2, p. 281–319, 2018.

[22] Y. Jin, R. Qu et J. Atkin, “Constrained portfolio optimisation : the state-of-the-art markowitz models,” 2016.

[23] L. Mencarelli et C. d’Ambrosio, “Complex portfolio selection via convex mixed-integer quadratic programming : a survey,” *International Transactions in Operational Research*, vol. 26, nº. 2, p. 389–414, 2019.

[24] S. Benartzi et R. H. Thaler, “Naive diversification strategies in defined contribution saving plans,” *American economic review*, vol. 91, nº. 1, p. 79–98, 2001.

[25] V. DeMiguel, L. Garlappi et R. Uppal, “Optimal versus naive diversification : How inefficient is the 1/n portfolio strategy ?” *The review of Financial studies*, vol. 22, nº. 5, p. 1915–1953, 2009.

[26] G. C. Pflug, A. Pichler et D. Wozabal, “The 1/n investment strategy is optimal under high model ambiguity,” *Journal of Banking & Finance*, vol. 36, nº. 2, p. 410–417, 2012.

[27] K. P. Anagnostopoulos et G. Mamanis, “The mean–variance cardinality constrained portfolio optimization problem : An experimental evaluation of five multiobjective evolutionary algorithms,” *Expert Systems with Applications*, vol. 38, nº. 11, p. 14 208–14 217, 2011.

[28] D. Bienstock, “Computational study of a family of mixed-integer quadratic programming problems,” *Mathematical programming*, vol. 74, nº. 2, p. 121–140, 1996.

[29] D. X. Shaw, S. Liu et L. Kopman, “Lagrangian relaxation procedure for cardinality-constrained portfolio optimization,” *Optimisation Methods & Software*, vol. 23, n°. 3, p. 411–420, 2008.

[30] N. J. Jobst *et al.*, “Computational aspects of alternative portfolio selection models in the presence of discrete asset choice constraints,” *Quantitative finance*, vol. 1, n°. 5, p. 489–501, 2001.

[31] D. Bertsimas et R. Shioda, “Algorithm for cardinality-constrained quadratic optimization,” *Computational Optimization and Applications*, vol. 43, n°. 1, p. 1–22, 2009.

[32] P. Bonami et M. A. Lejeune, “An exact solution approach for portfolio optimization problems under stochastic and integer constraints,” *Operations research*, vol. 57, n°. 3, **p.** 650–670, 2009.

[33] D. Bienstock, “Eigenvalue techniques for proving bounds for convex objective, nonconvex programs,” *Integer Programming and Combinatorial Optimization, Lecture Notes in Computer Science*, vol. 6080, p. 29–42, 2010.

[34] J. P. Vielma, S. Ahmed et G. L. Nemhauser, “A lifted linear programming branch-and-bound algorithm for mixed-integer conic quadratic programs,” *INFORMS Journal on Computing*, vol. 20, n°. 3, p. 438–450, 2008.

[35] J. Gao et D. Li, “Optimal cardinality constrained portfolio selection,” *Operations research*, vol. 61, n°. 3, p. 745–761, 2013.

[36] X. Cui *et al.*, “Convex relaxations and miqcqp reformulations for a class of cardinality-constrained portfolio selection problems,” *Journal of Global Optimization*, vol. 56, n°. 4, **p.** 1409–1423, 2013.

[37] A. Frangioni et C. Gentile, “Perspective cuts for a class of convex 0–1 mixed integer programs,” *Mathematical Programming*, vol. 106, n°. 2, p. 225–236, 2006.

[38] A. Frangioni et C. Gentile, “Sdp diagonalizations and perspective cuts for a class of nonseparable miqp,” *Operations Research Letters*, vol. 35, n°. 2, p. 181–185, 2007.

[39] A. Frangioni et C. Gentile, “A computational comparison of reformulations of the perspective relaxation : Socp vs. cutting planes,” *Operations Research Letters*, vol. 37, n°. 3, p. 206–210, 2009.

[40] M. S. AktüRk, A. AtamtüRk et S. GüRel, “A strong conic quadratic reformulation for machine-job assignment with controllable processing times,” *Operations Research Letters*, vol. 37, n°. 3, p. 187–191, 2009.

[41] O. Günlük et J. Linderoth, “Perspective reformulation and applications,” dans *Mixed Integer Nonlinear Programming.* Springer, 2012, p. 61–89.

[42] X. Zheng, X. Sun et D. Li, “Improving the performance of miqp solvers for quadratic programs with cardinality and minimum threshold constraints : A semidefinite program approach,” *INFORMS Journal on Computing*, vol. 26, n°. 4, p. 690–703, 2014.

[43] R. H. Kwon et S. J. Stoyan, “Mean–absolute deviation portfolio models with discrete choice constraints,” *Algorithmic Operations Research*, vol. 6, n°. 2, p. 118–134, 2011.

[44] N. D. Hardoroudi *et al.*, “Solving cardinality constrained mean-variance portfolio problems via milp,” *Annals of Operations Research*, vol. 254, n°. 1-2, p. 47–59, 2017.

[45] K. Boudt et C. Wan, “The effect of velocity sparsity on the performance of cardinality constrained particle swarm optimization,” *Optimization Letters*, p. 1–12, 2019.

[46] I. B. Salehpoor et S. Molla-Alizadeh-Zavardehi, “A constrained portfolio selection model at considering risk-adjusted measure by using hybrid meta-heuristic algorithms,” *Applied Soft Computing*, vol. 75, p. 233–253, 2019.

[47] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society : Series B (Methodological)*, vol. 58, n°. 1, p. 267–288, 1996.

[48] P.J. Kremer et al., “Sparse portfolio selection via the sorted ℓ1-Norm,” *Journal of Ban & Finance*, vol. 110, p.105687, 2020.

[49] G. Shmueli *et al.*, *Data mining for business analytics : concepts, techniques, and applications in R.* John Wiley & Sons, 2017.

[50] A. A. P. Santos, “The out-of-sample performance of robust portfolio optimization,” *Brazilian Review of Finance*, vol. 8, n°. 2, p. 141–166, 2010.

[51] R. E. Welsch et X. Zhou, “Application of robust statistics to asset allocation models,” *REVSTAT–Statistical Journal*, vol. 5, n°. 1, p. 97–114, 2007.

[52] E. D. Supandi, D. Rosadi *et al.*, “An empirical comparison between robust estimation and robust optimization to mean-variance portfolio,” *Journal of Modern Applied Statistical Methods*, vol. 16, n°. 1, p. 32, 2017.

[53] J. D. Jobson et B. Korkie, “Estimation for markowitz efficient portfolios,” *Journal of the American Statistical Association*, vol. 75, n°. 371, p. 544–554, 1980.

[54] M. J. Best et 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*, vol. 4, n°. 2, p. 315–342, 1991.

[55] M. Broadie, “Computing efficient frontiers using estimated parameters,” *Annals of Operations Research*, vol. 45, n°. 1, p. 21–58, 1993.

[56] M. Britten-Jones, “The sampling error in estimates of mean-variance efficient portfolio weights,” *The Journal of Finance*, vol. 54, n°. 2, p. 655–671, 1999.

[57] V. DeMiguel *et al.*, “A generalized approach to portfolio optimization : Improving performance by constraining portfolio norms,” *Management science*, vol. 55, nº. 5, p. 798–812, 2009.

[58] G. M. Frankfurter, H. E. Phillips et J. P. Seagle, “Portfolio selection : the effects of uncertain means, variances, and covariances,” *Journal of Financial and Quantitative Analysis*, vol. 6, nº. 5, p. 1251–1262, 1971.

[59] J. P. Dickinson, “The reliability of estimation procedures in portfolio analysis,” *Journal of Financial and Quantitative Analysis*, vol. 9, nº. 3, p. 447–462, 1974.

[60] P. A. Frost et J. E. Savarino, “For better performance : Constrain portfolio weights,” *Journal of Portfolio Management*, vol. 15, nº. 1, p. 29, 1988.

[61] V. Vapnik et V. Vapnik, “Statistical learning theory wiley,” *New York*, vol. 1, 1998.

[62] V. Vapnik, *The nature of statistical learning theory.* Springer science & business media, 2013.

[63] J. Brodie *et al.*, “Sparse and stable markowitz portfolios,” *Proceedings of the National Academy of Sciences*, vol. 106, nº. 30, p. 12 267–12 272, 2009.

[64] X. Xing, J. Hu et Y. Yang, “Robust minimum variance portfolio with l-infinity constraints,” *Journal of Banking & Finance*, vol. 46, p. 107–117, 2014.

[65] J. Fan, J. Zhang et K. Yu, “Vast portfolio selection with gross-exposure constraints,” *Journal of the American Statistical Association*, vol. 107, nº. 498, p. 592–606, 2012.

[66] B. Fastrich, S. Paterlini et P. Winker, “Constructing optimal sparse portfolios using regularization methods,” *Computational Management Science*, vol. 12, nº. 3, p. 417–434, 2015.

[67] D. Puelz, P. R. Hahn et C. M. Carvalho, “Sparse mean-variance portfolios : A penalized utility approach,” *arXiv preprint arXiv :1512.02310*, 2015.

[68] Z.-R. Lai *et al.*, “Short-term sparse portfolio optimization based on alternating direction method of multipliers,” *The Journal of Machine Learning Research*, vol. 19, nº. 1, p. 2547–2574, 2018.

[69] Z. Dai et F. Wang, “Sparse and robust mean–variance portfolio optimization problems,” *Physica A : Statistical Mechanics and its Applications*, vol. 523, p. 1371–1378, 2019.

[70] Z. Dai, “A closer look at the minimum-variance portfolio optimization model,” *Mathematical Problems in Engineering*, vol. 2019, 2019.

[71] Y.-M. Yen, “Sparse weighted-norm minimum variance portfolios,” *Review of Finance*, vol. 20, nº. 3, p. 1259–1287, 2016.

[72] Z. Dai et F. Wen, “A generalized approach to sparse and stable portfolio optimization problem,” *Journal of Industrial & Management Optimization*, vol. 14, n°. 4, p. 1651–1666, 2018.

[73] Y. Teng *et al.*, “A penalty palm method for sparse portfolio selection problems,” *Optimization Methods and Software*, vol. 32, n°. 1, p. 126–147, 2017.

[74] D. Bertsimas, R. Cory-Wright et J. Pauphilet, “A unified approach to mixed-integer optimization : Nonlinear formulations and scalable algorithms,” *arXiv preprint arXiv :1907.02109*, 2019.

[75] D. Bertsimas, J. Pauphilet et B. Van Parys, “Sparse regression : Scalable algorithms and empirical performance,” *arXiv preprint arXiv :1902.06547*, 2019.

[76] D. Bertsimas, B. Van Parys *et al.*, “Sparse high-dimensional regression : Exact scalable algorithms and phase transitions,” *The Annals of Statistics*, vol. 48, n°. 1, p. 300–323, 2020.

[77] D. P. Bertsekas, “Nonlinear programming. athena scientific,” *Belmont, MA*, 1999.

[78] D. Bertsimas, A. King et R. Mazumder, “Best subset selection via a modern optimization lens,” *The annals of statistics*, p. 813–852, 2016.

[79] D. Anguita *et al.*, “In-sample and out-of-sample model selection and error estimation for support vector machines,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, n°. 9, p. 1390–1406, 2012.

[80] A. A. Gaivoronski et F. Stella, “Stochastic nonstationary optimization for finding universal portfolios,” *Annals of Operations Research*, vol. 100, n°. 1, p. 165–188, 2000.