The Hong Kong Polytechnic University

Department of Applied Mathematics

Sparse and Dynamic Portfolio Optimization

Qiyu WANG

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

June 2017

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

______________________________(Signed)

Wang Qiyu (Name of student)

Dedicate to my parents.

Abstract

The thesis is concerned with sparse portfolio optimization and dynamic portfolio investment. The sparse portfolios we consider in this thesis are least-0-norm portfolio and least-p-norm with $p \in (0, 1)$ portfolio from the solution set of the Markowitz mean-variance (MV) optimization model. They are both NP hard problems. The dynamic portfolio investment model we consider is the projection of the given returns into a constraint comprised of a time-varying expected return in the form of parameterized ordinary differential equation (ODE) involving the Markowitz model. In this thesis, we resort to the stochastic linear complementarity approach, penalty methods to solve the sparse portfolio optimization problems and numerical methods to discretize and solve the parameter identification problem in the dynamic portfolio investment problem.

The least-0-norm portfolio we consider is in the framework of the classical Markowitz MV model when multiple solutions exist, among which the sparse solution is stable and cost-efficient. We study a two-phase stochastic linear complementarity approach (two-phase approach) to find a sparse solution. This approach stabilizes the optimization problem, finds the sparse asset allocation that saves the transaction cost, and results in the solution set of the Markowitz MV model. We apply the sample average approximation (SAA) method to overcome the randomness in the two-phase approach and give detailed convergence analysis. We implement this methodology on the data sets of Standard and Poor 500 index (S&P 500), real data of Hong Kong

and China market stocks (HKCHN) and Fama & French 48 industry sectors (FF48). With mock investment in the training data, we construct portfolios, test them in the out-of-sample data, find their Sharpe ratios and compare with the ℓ_1 penalty regularized portfolios, ℓ_p penalty regularized portfolios, cardinality constrained portfolios, and $1/N$ investment strategy. The least-0-norm portfolio is naturally sparser than the Markowitz portfolio. Moreover, we show the advantage of our approach in the risk management by using the criteria of standard deviation (STD), Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR).

We also consider the least-p-norm solution, $p \in (0, 1)$ in the Markowitz model solution set. The sparse portfolio model we study is solved by the penalty method. This model finds the least-p-norm sparse asset allocation in the solution set of the Markowitz MV model, saves the transaction cost and stabilizes the optimization problem. We apply the SAA method to overcome the randomness in the least-p-norm sparse portfolio model and give detailed convergence analysis. We implement this penalty method on the data sets of 20 A&H stocks, Fama & French 12 industry sectors (FF12) and Fama & French 25 portfolios formed on size and book-to-market (FF25). Using portfolios constructed in the training sample, we test them in the out-of-sample data, find their Sharpe ratios and compare with the $\|\cdot\|_0$ sparse portfolio, ℓ_1 penalty regularized portfolios, cardinality constrained portfolios, and $1/N$ investment strategy. Theoretically, least-p-norm portfolio would be sparser than the least-0-norm portfolio, but the least-0-norm portfolio might be more robust due to the simple structure of the two-phase approach.

The dynamic investment model that we propose and study is a model with expected return evolution containing unknown parameters. We project the target return to the constraint comprised of the parametric differential equation of the expected return coupled with the Markowitz MV model in every period. We discretize the model by the time-stepping method and use quasi-Newton method to identify the

parameters. Portfolios are then constructed according to the expected return evolution in multiple investment periods. The portfolio is re-balanced at the end of a round of dynamic portfolio investment incorporating the updated parameters. An empirical example using Dow Jones Industrial Average component stocks and index is given, which demonstrates the model.

Acknowledgements

First and foremost, I would like to express my gratitude to my chief supervisor Prof. Chen Xiaojun for her careful advice and help all through my study at the Hong Kong Polytechnic University. She has given me many thoughtful ideas, invaluable discussions and insightful suggestions to this thesis. The every-week seminar and group lunch enrich my horizon, stimulate my research interest, and provide precious chances for me to gain more knowledge.

I also wish to thank Dr. Sun Hailin (Nanjing University of Science and Technology) for his patient guidance during his visit in the Hong Kong Polytechnic University. Moreover, I would like to thank Dr. Cedric Yiu for giving precious comments to my thesis as my co-supervisor, Dr. Wang Zhenyu (Nanjing University), Dr. Li Xun, Dr. Ting Kei Pong, Dr. Xinqiu Zhao for their helpful discussion and encouragement.

Especially, I would like to express my deep thanks to my parents, who give me strong encouragement and full support, and are patient for my study. Thanks for their attention on my research and daily life.

Last but not least, I am grateful for the friendship and help of my classmates and my friends to my life in Hong Kong. I would like to express my special thanks to them.

Contents

Notation

A^T	the transpose of matrix (or vector) A.
A^{-1}	the inverse of matrix A.
$\| \cdot \|$	the Hausdorff distance
m	the number of assets in the investment universe
1_m	all one matrix of dimension $m \times 1$
r_i	the return rate of the asset i
μ_i	the mean of return rate for asset i
C	the covariance matrix for assets
ρ	the expected return rate of the portfolio
$\bar{\rho}(t)$	the target return rate of the portfolio at time period t
$\| \cdot \|_0$	the ℓ_0 norm
$\| \cdot \|_p$	the ℓ_p norm

Chapter 1

Introduction

1.1 Background

The portfolio, most of us construct, is comprised of a wide range of stocks from different sectors, cross several markets and through uncertain time periods. Portfolio selection is concerned with constructing an ensemble of assets that maximize one's return. However, return and risk are a pair of trade-off. One can not pursue high return with low risk. Markowitz (1952) ushered the era of modern portfolio management by his mean-variance model. In this model, he initiated a way for portfolio selection that balances the variance and return by mean-variance framework.

Although Markowitz model is elegantly straightforward, it is not trivial to compute portfolios from the Markowitz MV model. In practice, it is challenging to construct perfect portfolios from the model due to many reasons. We name several of them. Firstly, when the assets in the investment universe are large-scale, even if we allocate the investment into the assets wisely according to MV model, the transaction costs are considerable. Not to say the rapidly rising trading complexity along with the increasing number of assets in the market. Secondly, when the distribution of the stock returns is non-normal, a problem of overfitting arises in the estimation of unknown parameters, see Merton (1980). Due to estimation error, the MV model results are opaque and unstable, see Michaud and Michaud (2008). Last but not least, the Markowitz portfolio problem could be highly unstable. When assets in the investment universe are highly correlated, the smallest singular value of the returns matrix is extremely small, which leads to the phenomenon that small changes in the data would result large changes in solution, as reported in Brodie et al. (2009).

A natural thinking is that if we could find a scientific way to obtain sparse and stable portfolios. Sparsity is researched intensively and has a wide application in compressed sensing, portfolio selection.

There is indeed an ensemble of recently developed sparse Markowitz portfolio theories in the literature. In Brodie et al. (2009), Chen et al. (2013), Chen et al. (2017), DeMiguel et al. (2009a), Xu et al. (2014) and Xue et al. (2012), researchers proposed approaches which use penalty, e.g. ℓ_1 penalty, ℓ_p penalty, for $p \in (0, 1)$ to regularize the Markowitz portfolio optimization. Another stream of research is to add a cardinality constraint in the Markowitz portfolio selection model to select a cardinality number of assets. Many of the researchers investigated the cardinality constrained portfolio selection (CCPS) based on the branch-and-bound method, in Bertsimas and Shioda (2009), Bonami and Lejeune (2009). They developed a branch-and-bound method based on a novel geometric approach in Gao and Li (2013), or constructed a tight semidefinite program (SDP) problem approach in Zheng et al. (2014). Various other methods to study the CCPS include heuristic approaches in Cesarone et al. (2013), Deng et al. (2012), a positive programming approach in Tian et al. (2016), and a nonmonotone projected gradient (NPG) method in Xu et al. (2016).

Although these several works are used to enhance the Markowitz portfolio selection, they have problems, such as the change of the objective function, imposition of additional constraints. Then there is an idea arising from those frameworks. How about finding a sparse solution from the optimal solution set of the classic Markowitz MV model? When the covariance matrix is semi-definite, i.e., when there exist multiple solutions of the Markowitz portfolio optimization problem, we study a novel method to select a sparse portfolio solution of the classic Markowitz MV model.

Our approach is traced back to inspirations from Candes and Tao (2005). Chen and Xiang (2016) gave a way to find sparse solutions of the linear complementarity problem (LCP). By this method, we study a two-phase stochastic linear complementarity approach to find an approximate sparse solution of MV model by solving two simple convex problems in two phases. Note that finding a sparse portfolio over the solution set of MV model is an NP-hard problem, see Natarajan (1995). One of the advantages of the method is that we can approximate the NP-hard problem by solving a convex quadratic optimization problem and a linear optimization problem. Moreover, due to the discontinuity of ℓ_0 norm, it may not be easy to know how to guarantee the convergence when the SAA method is applied to the problem. Therefore, another contribution of this chapter is that we establish the convergence analysis when we apply the SAA method with the two-phase approach to find a sparse portfolio.

We use the two-phase approach to compute the sparse Markowitz portfolio of a randomly generated example, and three empirical examples, which are S&P 500 portfolios, HKCHN cross market portfolios and FF48 portfolios. We test their out-of-sample performance. From the preliminary results, our approach has better performance compared with the reported well-performed ℓ_1 penalty regularized portfolio, ℓ_p penalty regularized portfolio, cardinality constrained portfolio and $1/N$ investment strategy in DeMiguel et al. (2009b).

Our approach outperforms traditional Markowitz portfolios in the sense that: the sparse portfolio saves transaction fees and reduces investment complexity. Normally, the transaction cost is positive linearly related to the number of assets allocated. The sparse solutions from our approach naturally produce a cost efficient investment.

Moreover, our approach outperforms other sparse portfolios in the sense that the sparse solution of the two-phase approach is located in the solution set of the classical Markowitz portfolio optimization where the objective is only the risk measure. The merit of variance minimization of the Markowitz portfolio is inherited by the sparse solution from the two-phase approach in the sense that the sparse solution is exactly a solution of the classic Markowitz portfolio optimization problem.

From the results of our empirical examples, our sparse portfolio could outperform the comparative portfolios from the perspective of Sharpe ratio. Therefore, by our approach, we find a sparse and stable Markowitz portfolio such that the portfolio is well-posed to achieve the goal of classical MV model and preserves the stability and sparsity properties.

It is widely know that ℓ_p norm penalty yields even sparser solution. Based on our previous work in Wang and Sun (2017) where we had studied a novel two-phase approach to find a ℓ_0 norm sparse solution of MV model, we then propose a model by imposing ℓ_p-norm regularization to find a least-p-norm sparse solution over the optimal solution set. Our approach is traced back to inspirations from Chen and Xiang (2016) and Candes and Tao (2005). Note that finding a ℓ_p-norm sparse portfolio over the solution set of MV model is an NP-hard problem, see Ge et al. (2011). One of the challenges is that ℓ_p-norm is non-convex and non-Lipschitz. Moreover, it is hard to guarantee the convergence when the SAA method is applied to the problem. In order to overcome the difficulties, we use the penalty method to solve the problem. We also establish the convergence analysis when we apply the SAA method to handle the stochastic returns.

To test the effectiveness of the penalty method model, we use the nonmonotone proximal gradient (NPG) algorithm to compute the least-p-norm sparse Markowitz portfolio of a randomly generated example, and three empirical examples using data sets of 20 A&H stocks portfolios, FF12 and FF25 portfolios. We construct a series of portfolios and compare their out-of-sample performance. From out-of-sample results, our approach has better performance in Sharpe ratio compared to the reported well-performed ℓ_1 penalty regularized portfolio, cardinality constrained portfolio, least-0-norm sparse portfolio and $1/N$ investment strategy, see DeMiguel et al. (2009b).

This approach outperforms traditional Markowitz portfolios in the sense that: the sparse portfolio saves transaction fees and reduces investment complexity. Normally, the transaction cost increases with the increasing number of selected assets. Thus, the least-p-norm sparse solution of portfolio allocation saves investors transaction cost as well as mental energy to management the individual assets.

Moreover, our approach outperforms other sparse portfolios in the sense that the least-p-norm solution selects a limited number of stocks in the solution set of the MV model. The merit of balanced return-risk tradeoff of the Markowitz portfolio is inherited by the least-p-norm solution in the sense that the least-p-norm solution is exactly a solution of the classic Markowitz portfolio optimization problem.

Therefore, by our approach, we find a sparse and stable Markowitz portfolio such that the portfolio is well-posed to achieve the goal of classical MV model and preserves the stability and sparsity properties.

The Markowitz (1952) MV portfolio selection model is a seminal work in the contemporary finance. The MV model provides the portfolio allocation strategy concerning return and risk tradeoff. However, investors need to consider portfolio construction over a long-time horizon. The dynamic portfolio management thus embarks an ensemble of research works. Li et al. (2002), Li and Ng (2000), Zhou and Li (2000), Cui et al. (2014) and etc. ushered a new norm of self-financing dynamic portfolio policy formed with time evolution. Recently developed modern methods in works such as Liu et al. (2013) enhanced the dynamic portfolio construction for real-time optimization.

We consider the dynamic portfolio management within a certain period to fit in given target return and then make investment at the end of the time period. It is very often that investors pursue for a certain return. Ammann (2003), Holthausen (1981) and Horneff et al. (2010) reported that investors and portfolio managers often pursue for a guaranteed return. However, the setting of return is unexamined. Therefore, we propose a model to fit this gap. We minimize the return dispersion to a specified target or benchmark return with a constraint involving the return and risk dynamics. This modeling is meaningful in the real practice. For some investors, they concern themselves with index tracking, fund tracking or inflation-indexed securities tracking, see Cowell (2013). Then it is natural to have a target return.

The mathematical model could trace back to models discussed in Pang and Stewart (2008), Chen and Wang (2013), Chen and Wang (2014). In our model, the expected return ρ is a state variable and the portfolio position w is a control variable. Expected return at the current period depends on the expected return and positions in the portfolio at the last period with unknown parameters. Investors demand for the minimization of the expected return dispersion to the target return constrained by a parametric differential equation of ρ involving the MV model. The parameters in a round of dynamic portfolio investment are thus identified by optimization. In this way, we also determine the expected return by the parametric differential equation.

We use real data on asset returns. With time elapsing, the information accumulates. The investor could adopt a re-balancing strategy every certain periods of time. On the re-balancing date, the parameters which set the expected return are learnt in the sample by the multi-period investment. In this way, the investors use the parameters learnt in the dynamic investment in the training sample to set the expected return and construct the portfolio for the out-of-sample investment.

It is hard to derive the continuous solution, so we use the simple discrete time model to resolve the continuous optimization problem. We discretize the parametric optimization model with an ODE constraint involving MV model by time-stepping method and solve it. The optimization arguments are the unknown parameters in the state variable dynamics and the state variable ρ_i, for $i = 1, ..., N$. We use them to construct the portfolio and make out-of-sample investment decision.

The contributions of our model are in several aspects. Firstly, we consider parameter identification in the dynamic portfolio using quasi-Newton method. Secondly, it is a brand new model to set the return dynamically with regards to target return. We have constructed the discrete scheme to identify the parameters and proved the solvability and convergence.

We implemented this model in an empirical example. The data sets that we use are Dow Jones Industrial Average (DJIA) component stocks and index returns. Using a reasonable amount of sample data to do multi-period investments, we identify the unknowns of the optimization system. We then determine the dynamic investment strategy by using the predicted expectation return and predicted sample to construct the Markowitz portfolios. We show the out-of-sample portfolio returns, Sharpe ratios and portfolio values on each re-balance date.

1.2 Contributions and organization of the Thesis

In this thesis, we study sparse portfolio selection and dynamic portfolio selection. For the least-0-norm and the least-p-norm sparse portfolio selection, the main difficulty to solve the sparse portfolio selction is nonsmooth nonconvex optimization problem. The sparse solution of linear complementarity problem in Chen and Xiang (2016) provides a way to approximate the least-0-norm portfolio optimization problem by a two-phase method which solves a quadratic programming (QP) problem and a linear programming problem. The penalty method is designed for a class of non-Lipschitz optimization problems in Chen et al. (2016). The least-p-norm portfolio selection is thus constructed via the penalty method. Investors often need to manage portfolio dynamically. We use the quasi-Newton method for parameter identification in a dynamic portfolio selection with minimization of the deviation of the expected return to the target return as the objective and a parametric ODE with a QP problem as the constraint. The model is solved by time-stepping method discretely.

In Chapter 2, we present a brief introduction to the Markowitz MV model, some important sparse portfolio models. Some of the introductory knowledge in the linear complementarity problem (LCP), penalty method which will be used in the following chapters are also given.

In Chapter 3, we study a two-phase approach to find the sparse portfolio via stochastic linear complementarity approach. We formulate the model as a minimization to find the least-0-norm solution in the solution set of Markowitz MV model. We construct the solution set and approximates the least-0-norm solution by the least-1-norm solution. Hence, the nonconvex nonsmooth problem is solved by a QP problem and a linear programming problem. We use the SAA method to deal with the randomness and prove the convergence. Some properties of the solutions are discussed. A randomly generated example and several empirical examples are presented to illustrate the performance of the proposed model.

In Chapter 4, we study a penalty method to find the least-p-norm sparse portfolio. This is to find the least-p-norm solution in the solution set of Markowitz MV model. We use the penalty method which solves a class of non-Lipschitz problem by the nonmonotone proximal gradient (NPG) algorithm to construct the sparse portfolio. We use the SAA method and prove the convergence. Some properties of the solutions are discussed. A randomly generated example and several empirical examples are presented to demonstrate the performance of the proposed model.

In Chapter 5, we study parameter identification in the dynamic portfolio selection which is formed as a problem to minimize the deviation of the expected returns to the target return with a constraint of a parametric differential equation involving the MV model. The expected return dynamics is an ODE function with unknown parameters. At every time step, the investors construct portfolio in the framework of the Markowitz MV model with the expected return given by the ODE. We solve for the unknown parameters to minimize the deviation of the expected return to the target return in a discrete scheme. A numerical example is given to show the performance of the dynamic portfolio.

The whole thesis deals with the sparse portfolio selection via stochastic linear complementarity approach and penalty method; and parameter identification in the dynamic portfolio selection via quasi-Newton method with a constraint of an ODE equation involving a QP problem. Chapter 6 makes the conclusion of the thesis and points out the further work.

Chapter 2

Preliminary

The purpose of this chapter is to review the basic concepts of portfolio selection model and present some lemmas and prerequisite knowledge which will be used in the following chapters.

2.1 Several Models

Specifically, for m securities, the traditional Markowitz portfolio is to find a portfolio that has minimal variance for a given expected return ρ by solving the following problem:

$$\begin{array}{ll} \min & w^T C w \\ \text{s.t.} & w^T \mu = \rho \\ & w^T 1_m = 1, \end{array} \tag{2.1}$$

where $w = (w_1, w_2, ..., w_m)^T$, 1_m is an m-dimensional vector with all entries being one, random variable $r_i : \Xi \to \mathbb{R}$ is the return of the ith security, $\mu_i = \mathbb{E}[r_i(\xi)]$ is its expected return, the return of the securities $r(\xi) = (r_1(\xi), r_2(\xi), ..., r_m(\xi))^T$, the expected return of the securities $\mu = (\mu_1, \mu_2, ..., \mu_m)^T$, the covariance matrix of the returns $C = \mathbb{E}[(r(\xi) - \mu)(r(\xi) - \mu)^T]$. Note that C is an $m \times m$ positive semi-definite matrix. Denote $\mathcal{H}$ as its optimal solution set.

In recent years, there is an ensemble of recently developed sparse Markowitz portfolio theories in the literature as follows.

1. Regularized method: The approaches use penalty to regularize the Markowitz portfolio optimization, such as ℓ_1 penalty in Brodie et al. (2009),

$$\begin{array}{ll} \min & w^T C w + \lambda \|w\|_1 \\ \text{s.t.} & w^T 1_m = 1, \quad w^T \mu = \rho, \end{array} \tag{2.2}$$

or in Chen et al. (2013), ℓ_p penalty for $p \in (0, 1)$,

$$\begin{array}{ll} \min & w^T C w + \lambda \|w\|_p^p \\ \text{s.t.} & w^T 1_m = 1, \quad w^T \mu = \rho. \end{array} \tag{2.3}$$

2. Cardinality constrained portfolio selection (CCPS): The core of the model is to add a cardinality constraint in the Markowitz portfolio selection model to select a cardinality number of assets in the portfolio,

$$\begin{array}{ll} \min & w^T C w \\ \text{s.t.} & w^T 1_m = 1, \quad w^T \mu = \rho \\ & \|w\|_0 \leqslant k. \end{array}$$

2.2 Some Lemmas

Throughout this thesis, we use the following notation. $x^T y$ denotes the scalar products of two vectors x and y, $\|\cdot\|_p^p$ and $\|\cdot\|_0$ denote the p-norm and 0-norm of a vector respectively. The p-norm is $\|x\|_p := (\Sigma_{i=1}^n |x_i|^p)^{1/p}$. $d(x, S) := \inf_{x' \in S} \|x - x'\|$ denotes the distance from point x to set S. For two sets S_1 and S_2, $\mathbb{D}(S_1, S_2) := \sup_{x \in S_1} d(x, S_2)$ denotes the deviation of set S_1 from set S_2 and

$$\mathbb{H}(S_1, S_2) := \max(\mathbb{D}(S_1, S_2), \mathbb{D}(S_2, S_1))$$

denotes the Hausdorff distance between two sets S_1 and S_2.

We present some auxiliary lemmas on error bound, obtained from Theorem 3.1 in Luo and Luo (1994). Let $A \in R^{n \times n}$ and $b \in R^{m \times 1}$.

Lemma 2.1. *There exists a $C_0 > 0$ so that for all $x \in R^n$, $n = m + 2$, we have*

$$dist(x, S) \leqslant C_0 \|Ax - b\|_1, \tag{2.4}$$

where $S = \{x : \|Ax - b\|_1 = 0\}$.

The constant C_0 is not explicitly computed. In this case, S is compact.

Chapter 3

Sparse portfolio selection via stochastic linear complementarity approach

In this chapter, we study a two-phase approach to find the sparse portfolio via s-tochastic linear complementarity approach. The model is formulated as minimization of the 0-norm of the solutions in the solution set of Markowitz MV model. By the two-phase approach, we turn the nonconvex nonsmooth problem to a QP problem and a linear programming problem which approximates the sparse solution. We first deduce the formulation by sparse solutions of LCP in Chen and Xiang (2016), then we use the SAA method to deal with the randomness and prove the convergence. Some properties of the solutions are derived. We tested the model by a randomly generated example and several empirical examples which demonstrated the performance of the proposed model.

This chapter is organized as follows. In Section 3.2, we introduce the two-phase approach to find a sparse Markowitz portfolio. In Section 3.3, we analyze the convergence of the SAA method. In Section 3.4, we show the numerical and empirical applications of the model. Section 3.5 is a conclusion.

3.1 Model formulation

3.1.1 Model construction and two-phase stochastic linear complementarity approach

We consider the sparse Markowitz portfolio optimization problem as follows:

$$\begin{array}{ll} \min & \| w \|_0 \\ \text{s.t.} & w \in \mathcal{H}, \end{array} \tag{3.1}$$

where $\mathcal{H}$ is the optimal solution set of the Markowitz portfolio optimization problem

$$\begin{array}{ll} \min & w^T C w \\ \text{s.t.} & w^T \mu = \rho \\ & w^T 1_m = 1 \\ & \|w\|_1 \leqslant \eta, \end{array} \tag{3.2}$$

where $\eta \geqslant 1$.

Compared with problem (2.1), problem (3.2) has an additional constraint to limit the arbitrage $\|w\|_1 \leqslant \eta$. Short selling, which is defined as "the sale of a security that the seller does not own or that the seller owns but does not deliver" by the SEC, is considered crucial for effective arbitrage. In fact, short sale is often constrained. In other words, arbitrage is limited, see implications in DeMiguel et al. (2009a), Diamond and Verrecchia (1987), Shleifer and Vishny (1997). Moreover, when $\eta =$ 1, problem (3.2) is equivalent to the classical Markowitz portfolio model without shortselling:

$$\begin{array}{ll} \min & w^T C w \\ \text{s.t.} & w^T \mu = \rho \\ & w^T 1_m = 1, \\ & w \geqslant 0, \end{array} \tag{3.3}$$

and when $\eta = \infty$, problem (3.2) is equivalent to (2.1).

Problem (4.1) is a nonconvex noncontinuous optimization problem and may not be easy to solve. Inspired by the method in Chen and Xiang (2016), we provide the two-phase stochastic linear complementarity approach to approximate problem (4.1) in the following subsection.

3.1.2 Two-phase stochastic linear complementarity approach

The basic idea of this approach is that we reformulate the feasible set of problem (4.1) (solution set of problem (3.2)) as a linear system and find its least ℓ_1 norm to approximate its sparse solution. To this end, we consider the reformulation of problem (3.2)

$$\begin{array}{ll} \min & (w^+ - w^-)^T C (w^+ - w^-) \\ \text{s.t.} & (w^+ - w^-)^T \mu = \rho \\ & (w^+ - w^-)^T 1_m = 1 \\ & w^+ \geqslant 0, w^- \geqslant 0 \\ & 1_{2m}^T (w^+, w^-) \leqslant \eta. \end{array} \tag{3.4}$$

Firstly, we will show the equivalence between problem (3.2) and problem (3.4). Secondly, we will study how to find a sparse solution from the optimal solution set of problem (3.4).

Proposition 3.1. *Problem (3.2) and problem (3.4) are equivalent in the sense that (i) If w^* is an optimal solution of (3.2), then $((w^+)^*, (w^-)^*)$ is a solution of problem (3.4), where $(w^+)^* = \max(w^*, 0)$, $(w^-)^* = \max(-w^*, 0)$; (ii) if $((w^+)^*, (w^-)^*)$ is an optimal solution of (3.4), then $w^* = (w^+)^* - (w^-)^*$ is an optimal solution of problem (3.2); (iii) the least ℓ_1 norm solution and sparse solution of problem (3.4) must be the least ℓ_1 norm solution and sparse solution of problem (3.2) and vice versa.*

Proof. (i) Let w^* be an optimal solution of problem (3.2). Then $(\hat{w}^+, \hat{w}^-)$ is a feasible solution of (3.4), where $\hat{w}^+ = \max(w^*, 0)$ and $\hat{w}^- = \max(-w^*, 0)$. Let us prove that $(\hat{w}^+, \hat{w}^-)$ is an optimal solution of (3.4). Assume for the sake of a contradiction that $(\hat{w}^+, \hat{w}^-)$ is not an optimal solution of problem (3.4), then there exists $(\bar{w}^+, \bar{w}^-)$ such that

$$(\bar{w}^+ - \bar{w}^-)^T C(\bar{w}^+ - \bar{w}^-) < (\hat{w}^+ - \hat{w}^-)^T C(\hat{w}^+ - \hat{w}^-). \tag{3.5}$$

Let $\bar{w} = (\bar{w}^+ - \bar{w}^-)$. Since $\|\bar{w}\|_1 \leqslant \|\bar{w}^+\|_1 + \|\bar{w}^-\|_1 = 1_{2m}^T(\bar{w}^+, \bar{w}^-) \leqslant \eta$, $\bar{w}$ is a feasible solution of problem (3.2), then (3.5) implies

$$\bar{w}^T C\bar{w} < (w^*)^T C w^*,$$

a contradiction that w^* is an optimal solution of problem (3.2).

(ii) Let v_1 and v be the optimal values of problem (3.2) and problem (3.4) respectively. It is obvious that for any feasible solution $\tilde{w}$ of problem (3.2), $(\tilde{w}^+, \tilde{w}^-)$ is a feasible solution of problem (3.4) with the same objective function value, that is

$$(\tilde{w}^+ - \tilde{w}^-)^T C(\tilde{w}^+ - \tilde{w}^-) = \tilde{w}^T C\tilde{w},$$

where $\tilde{w}^+ = \max(\tilde{w}, 0)$ and $\tilde{w}^- = \max(-\tilde{w}, 0)$. Then $v_1 \geqslant v$. Let $((w^+)^*, (w^-)^*)$ be an optimal solution of problem (3.4), $w^* = (w^+)^* - (w^-)^*$. Note that w^* is a feasible solution of problem (3.2). Let w' be an optimal solution of problem (3.2). Then

$$v \leqslant v_1 = w'^T C w' \leqslant (w^*)^T C w^* = ((w^+)^* - (w^-)^*)^T C((w^+)^* - (w^-)^*) = v,$$

which means $(w^*)^T C w^* = w'^T C w'$ and w^* is an optimal solution of problem (3.2).

(iii) By (i) and (ii), we have that for any optimal solution w^* of (3.2), any (w^+, w^-) such that $1_{2m}^T(w^+, w^-) \leqslant \eta$, $w^+ - w^- = w^*$ and $w^+ \geqslant 0$, $w^- \geqslant 0$ is the optimal solution of (3.4). Then let $w_{\ell_1}^*$ be the least ℓ_1 norm solution of (3.2). Then we claim $(w_{\ell_1}^+, w_{\ell_1}^-)$ is a least ℓ_1 norm solution of (3.4) where $w_{\ell_1}^+ = \max\{w_{\ell_1}^*, 0\}$ and $w_{\ell_1}^- = \max\{-w_{\ell_1}^*, 0\}$. Otherwise, there exists an optimal solution (w^+, w^-) of (3.4) such that

$$1_m^T|w^+ - w^-| \leqslant 1_{2m}^T(w^+, w^-) < 1_{2m}^T(w_{\ell_1}^+, w_{\ell_1}^-) = 1_m^T|w_{\ell_1}^*|,$$

which contradicts the fact that $w_{\ell_1}^*$ is the least ℓ_1 norm solution of problem (3.2).

Moreover, let $(w_{\ell_1}^+, w_{\ell_1}^-)$ be the least ℓ_1 norm solution of (3.4), it is obvious that $(w_{\ell_1}^+)^T w_{\ell_1}^- = 0$. Otherwise, $1_{2m}^T(w_{\ell_1}^+, w_{\ell_1}^-) > 1_{2m}^T(\bar{w}_{\ell_1}^+, \bar{w}_{\ell_1}^-)$, where $\bar{w}_{\ell_1}^+ = \max(w_{\ell_1}^+ - w_{\ell_1}^-, 0)$, $\bar{w}_{\ell_1}^- = \max(w_{\ell_1}^- - w_{\ell_1}^+, 0)$ and $(\bar{w}_{\ell_1}^+, \bar{w}_{\ell_1}^-)$ is a feasible solution of (3.4), which contradicts the fact that $(w_{\ell_1}^+, w_{\ell_1}^-)$ is the least ℓ_1 norm solution of problem (3.4).

Moreover, $w_{\ell_1}^* = (w_{\ell_1}^+ - w_{\ell_1}^-)$ is the least ℓ_1 norm solution of (3.2). Otherwise, there exists an optimal solution of $\bar{w}_{\ell_1}$ of (3.2) such that

$$1_{2m}^T(\max(\bar{w}_{\ell_1}^+, 0), \max(-\bar{w}_{\ell_1}^-, 0)) = 1_m^T|\bar{w}_{\ell_1}| < 1_m^T|w_{\ell_1}^*| = 1_{2m}^T(w_{\ell_1}^+, w_{\ell_1}^-),$$

which contradicts the fact that $(w_{\ell_1}^+, w_{\ell_1}^-)$ is the ℓ_1 norm solution of problem (3.4).

The proof for sparse solution is similar to the least ℓ_1 norm solution, we omit the details. □

It is obvious that problem (3.4) can be rewritten as:

$$\begin{array}{ll} \min & ((w^+)^T, (w^-)^T)H((w^+)^T, (w^-)^T)^T \\ \text{s.t.} & G\begin{pmatrix} (w^+) \\ (w^-) \end{pmatrix} \geqslant b \\ & w^+ \geqslant 0, w^- \geqslant 0, \end{array} \tag{3.6}$$

where $H = \begin{pmatrix} C & -C \\ -C & C \end{pmatrix}$, $G = \begin{pmatrix} \mu^T & -\mu^T \\ 1_m^T & -1_m^T \\ -\mu^T & \mu^T \\ -1_m^T & 1_m^T \\ -1_m^T & -1_m^T \end{pmatrix}$, $b = \begin{pmatrix} \rho \\ 1 \\ -\rho \\ -1 \\ -\eta \end{pmatrix}$. Moreover, problem (3.6) can be rewritten as an LCP:

$$x^T(Mx + q) = 0, Mx + q \geqslant 0, x \geqslant 0, \tag{3.7}$$

where $M = \begin{pmatrix} H & -G^T \\ G & O_{5\times 5} \end{pmatrix}$, $q = \begin{pmatrix} O_{2m\times 1} \\ -b \end{pmatrix}$, $x = \begin{pmatrix} w^+ \\ w^- \\ y \end{pmatrix}$, $y = (y_1, y_2, y_3, y_4, y_5)^T$.

We use LCP(q, M) to denote LCP(3.7) with the related q and M.

Hereafter, for simplicity, we use $x = (w^+, w^-, y)$ to denote $x = ((w^+)^T, (w^-)^T, y^T)^T$. Problem (3.6) and problem (3.7) are equivalent in the sense that (i) if $((w^+)^*, (w^-)^*) \in$

R^{2m} is a solution of (3.6) then there is $y^* \in R^4$ such that $((w^+)^*, (w^-)^*, y^*)$ is a solution of (3.7); (ii) if $((w^+)^*, (w^-)^*, y^*)$ is a solution of (3.7) then w^* is a solution of (3.6).

By Theorem 3.1.7 in Cottle et al. (1992), since M is a positive semi-definite matrix, the solution set of the LCP(q, M), denoted by the SOL(q, M) can be written as:

$$\text{SOL}(q, M) = \{x \in R_+^{2m+4} | Mx + q \geqslant 0, H \begin{pmatrix} w^+ \\ w^- \end{pmatrix} = \varsigma, q^T x = \gamma\} \tag{3.8}$$

where $\varsigma = H \begin{pmatrix} (w^+)^* \\ (w^-)^* \end{pmatrix}$, $\gamma = q^T x^*$ and $x^* = ((w^+)^*, (w^-)^*, y^*)$ is an arbitrary solution of the LCP(q, M).

By a given solution x^* of the LCP(q, M), all the solutions (w^+, w^-) of the Markowitz portfolio optimization with corresponding Lagrange multipliers y are given in the SOL(q, M). Then we try to look for a sparse solution in the solution set SOL(q, M). We can find the sparse Markowitz portfolio of problem (4.1) by solving

$$\begin{array}{ll} \min & \|(w^+, w^-)\|_0 \\ \text{s.t.} & Mx + q \geqslant 0 \\ & x \geqslant 0 \\ & H \begin{pmatrix} w^+ \\ w^- \end{pmatrix} = \varsigma \\ & q^T x = \gamma, \end{array} \tag{3.9}$$

where $\varsigma = H \begin{pmatrix} (w^+)^* \\ (w^-)^* \end{pmatrix}$, $\gamma = q^T x^*$ and $x^* = ((w^+)^*, (w^-)^*, y^*)$ is an arbitrary solution of the LCP(q, M).

Note that $\|\cdot\|_1$ is a good approximation of $\|\cdot\|_0$, which has been widely used such as Brodie et al. (2009). Then we can approximate problem (3.9) by the following linear programming:

$$\begin{array}{rl} \min & (w^+, w^-)^T 1_{2m} \\ \text{s.t.} & Mx + q \geqslant 0 \\ & x \geqslant 0 \\ & H \begin{pmatrix} w^+ \\ w^- \end{pmatrix} = \varsigma \\ & q^T x = \gamma, \end{array} \tag{3.10}$$

where $\varsigma = H \begin{pmatrix} (w^+)^* \\ (w^-)^* \end{pmatrix}$, $\gamma = q^T x^*$ and $x^* = ((w^+)^*, (w^-)^*, y^*)$ is a solution of the LCP(q, M).

It will be very interesting to consider the upper bound of the ℓ_0 norm of the classical Markowitz portfolio optimization sparse solutions. We will discuss it in the following proposition.

Proposition 3.2. *Consider problem (4.1), the number of nonzero entries in a solution of problem (4.1) is at most* $\text{rank}(C) + 3$*, where* C *is the covariance matrix.*

Proof. $\mathcal{H}$ denotes the solution set of problem (3.2). Let $\tilde{w}$ and (w^+, w^-) be the sparse solution of problem (3.2) and (3.4). By Proposition 3.1, $\|\tilde{w}\|_0 = \|(w^+, w^-)\|_0$. Note that problem (3.4) can be written as the LCP(q, M). From rank(G)= 3, it is obvious that $\text{rank}(M) \leqslant \text{rank}(H) + 3 = \text{rank}(C) + 3$, the last equation is from the definition of H. Then by (Chen and Xiang, 2016, Theorem 2.1) and Remark 3.1, we have

$$\|\tilde{w}\|_0 = \|(w^+, w^-)\|_0 \leqslant \text{rank}(C) + 3.$$

□

3.1.3 Equivalence between the least ℓ_1 norm solution and the sparse solution

Under some conditions, we can show that a solution of problem (3.9) can be found exactly by solving problem (3.10) if we replace (w^+, w^-) by x in the objective functions for both two problems (In this subsection, we consider problem (3.9) and (3.10) with objective function $\|x\|_0$ and $\|x\|_1$ respectively). To this end, we need the following definition from Candes and Tao (2005). Here $|t|$ is the number of elements of the set t.

Definition 3.1. *An $m \times n$ matrix P is said to satisfy the s-restricted isometry property (RIP) with a restricted isometry constant δ_s, if for every $m \times |t|$ sub-matrix P_t of P and for every vector $z \in R^{|t|}$ with $|t| \leqslant s$,*

$$(1 - \delta_s)\|z\|_2^2 \leqslant \|P_t z\|_2^2 \leqslant (1 + \delta_s)\|z\|_2^2.$$

P is said to satisfy the s, s'-restricted orthogonality (RO) with a restricted orthogonality constant $\theta_{s,s'}$ for $s+s' \leqslant n$ if for all sub-matrices, $P_t \in R^{m\times|t|}$ and $P_{t'} \in R^{m\times|t'|}$ of P, with $|t| \leqslant s$ and $|t'| \leqslant s'$, for all the vectors $z \in R^{|t|}$ and $z' \in R^{|t'|}$,

$$|(P_t z, P_{t'} z')| \leqslant \theta_{s,s'}\|z\|_2\|z'\|_2$$

holds for all the disjoint sets t and t'.

The following theorem is from Chen and Xiang (2016).

Theorem 3.1. *Let $\hat{x}$ be the optimal solution of linear programming problem (3.10) with $\|\hat{x}\|_0 \leqslant s$. Then*

1. *if H satisfies the RIP with a restricted isometry constant $\delta_{2s} < 1$, then $\hat{x}$ is the unique sparse solution of the* LCP*(q, M);*

2. *if H satisfies the RIP and RO with*

$$\delta_s + \theta_{s,s'} + \theta_{s,2s'} < 1,$$

then $\hat{x}$ is the unique solution of (3.10) and the unique sparse solution of the LCP*(q, M).*

Theorem 4.1 shows the relationship between problems (3.9) and (3.10).

Therefore, we approach an NP-hard problem by solving two simple problems (3.7) and (3.10) in two phases.

Remark 3.1. *In this subsection, we considers the* $\|x_0\|$ *where* $x = (w^+, w^-, y)$*, but we actually want to consider* $\|w\|_0$*. The reason we consider* $\|x\|_0$ *is that, under the conditions of Theorem 4.1, its solution can be obtained by solving problem (3.10). Moreover, assuming that* w^* *is the optimal solution of problem (4.1) and* $\bar{x} = ((\bar{w})^+, (\bar{w})^-, \bar{y})$ *is the optimal solution of problem (3.9) with objective function* $\|x\|_0$*, it is obvious that*

$$\|((w^+)^*, (w^-)^*, \hat{y})\|_0 \geqslant \|(\bar{w}^+, \bar{w}^-, \bar{y})\|_0 = \|\bar{x}\|_0 \geqslant \|(\bar{w}^+, \bar{w}^-)\|_0 \geqslant \|((w^+)^*, (w^-)^*)\|_0,$$

for any $\hat{y} \in R^5$*. Then* $0 < \|\bar{x}\|_0 - \|((w^+)^*, (w^-)^*)\|_0 \leqslant 5$*. In case that the dimension of* w *is very large, problem (3.9) is a good approximation of problem (4.1).*

But the RIP and RO conditions in Theorem 4.1 are not easy to satisfy. In the rest of chapter, we consider problem (3.9) and (3.10) with objective function $\|(w^+, w^-)\|_0$ *and* $\|(w^+, w^-)\|_1$ *respectively.*

3.2 The SAA method and convergence analysis

In this section, we use the SAA method to model the randomness of the Markowitz portfolio and consider the convergence analysis between the original problem and the approximation problem.

Let $\{r^j = (r_1^j, ..., r_m^j)\}_{j=1}^N$ be the i.i.d. samples of random variable r. Then the SAA mechanism of the LCP sparse Markowitz portfolio optimization is:

$$\begin{array}{rl} \min & (w^+, w^-)^T 1_{2m} \\ \text{s.t.} & M_N x + q \geqslant 0 \\ & x \geqslant 0 \\ & H_N \begin{pmatrix} w^+ \\ w^- \end{pmatrix} = \varsigma_N \\ & q^T x = \gamma_N, \end{array} \tag{3.11}$$

where

$$C_N = \frac{1}{N}\sum_{j=1}^{N}(r^j - \mu_N)(r^j - \mu_N)^T,$$

$\mu_{i,N} = \frac{1}{N}\sum_{j=1}^{N} r_i^j$, $\mu_N = (\mu_{1,N}, \ldots, \mu_{m,N})^T$, $M_N = \begin{pmatrix} H_N & -G_N^T \\ G_N & O_{5\times 5} \end{pmatrix}$, $H_N = \begin{pmatrix} C_N & -C_N \\ -C_N & C_N \end{pmatrix}$,

$G_N = \begin{pmatrix} \mu_N^T & -\mu_N^T \\ 1_m^T & -1_m^T \\ -\mu_N^T & \mu_N^T \\ -1_m^T & 1_m^T \\ -1_m^T & -1_m^T \end{pmatrix}$, $\varsigma_N = H_N((w^+)_N^T, (w^-)_N^T)_N^T$, $y_N = (y_{1,N}, y_{2,N}, y_{3,N}, y_{4,N}, y_{5,N})^T$,

$\gamma_N = q^T x_N$ and $x_N = ((w^+)_N^T, (w^-)_N^T, y_N^T)^T$ is a solution of the following LCP

$$x^T M_N x + q^T x = 0, M_N x + q \geqslant 0, x \geqslant 0. \tag{3.12}$$

Note that the LCP(q, M_N) (3.12) is driven from the first order necessary condition of the following SAA form of problem (3.6):

$$\begin{array}{ll} \min & ((w^+)^T, (w^-)^T) H_N ((w^+)^T, (w^-)^T)^T \\ \text{s.t.} & G_N \begin{pmatrix} w^+ \\ w^- \end{pmatrix} \geqslant b \\ & w^+ \geqslant 0, w^- \geqslant 0. \end{array} \tag{3.13}$$

It follows that we consider the convergence analysis between the original problem (3.10) and its SAA form (4.11). Although problem (3.10) is a stochastic linear problem, some of its parameters are solutions of the LCP(q, M) (3.7). So it is hard to give the convergence analysis between problem (3.10) and its SAA form directly. We also prove it in two steps. In the first step, we prove the convergence analysis of KKT pairs between problem (3.6) and its SAA form (3.13). In the second step, we prove the convergence between the optimal solutions and the optimal values of (3.10) and (4.11). We need the following Slater condition.

Assumption 3.1. *There exists a feasible point* (w_0^+, w_0^-) *such that* $(w_0^+, w_0^-) > 0$, $1_{2m}^T(w_0^+, w_0^-) < \eta$, $(w_0^+ - w_0^-)^T\mu = \rho$ *and* $(w_0^+ - w_0^-)^T 1_m = 1$.

Let $f(w^+, w^-) = ((w^+)^T, (w^-)^T)H((w^+)^T, (w^-)^T)^T$,

$$f_N(w^+, w^-) = ((w^+)^T, (w^-)^T)H_N((w^+)^T, (w^-)^T)^T,$$

$$D = (G^T, I_{2m\times 2m})^T, \text{ and } D_N = (G_N^T, I_{2m\times 2m})^T,$$

and I is an identity matrix. Then

$$g(w^+, w^-) = D\begin{pmatrix} w^+ \\ w^- \end{pmatrix} - \left(b^T, 0_{1\times 2m}\right)^T, \text{ and } g_N(w^+, w^-) = D_N\begin{pmatrix} w^+ \\ w^- \end{pmatrix} - \left(b^T, 0_{1\times 2m}\right)^T,$$

reformulate the constraints $G\begin{pmatrix} w^+ \\ w^- \end{pmatrix} - b$, $G_N\begin{pmatrix} w^+ \\ w^- \end{pmatrix} - b$ and $(w^+, w^-) \geqslant 0$.

We consider the convergence analysis between KKT pairs between problem (3.6) and its SAA form (3.13). The KKT condition of problem (3.6) is:

$$\begin{cases} 0 = 2H\begin{pmatrix} w^+ \\ w^- \end{pmatrix} - D^T\begin{pmatrix} y \\ s \end{pmatrix} \\ 0 = \min\{g(w^+, w^-), \begin{pmatrix} y \\ s \end{pmatrix}\}, \end{cases}$$

and the KKT condition of problem (3.13) is:

$$\begin{cases} 0 = 2H_N\begin{pmatrix} w^+ \\ w^- \end{pmatrix} - D_N^T\begin{pmatrix} y_N \\ s_N \end{pmatrix} \\ 0 = \min\{g_N(w^+, w^-), \begin{pmatrix} y_N \\ s_N \end{pmatrix}\}, \end{cases}$$

where y, s, y_N and s_N are Lagrangian multipliers corresponding to the constraints $G\begin{pmatrix} w^+ \\ w^- \end{pmatrix} \geqslant b$, $(w^+, w^-) \geqslant 0$, $G_N\begin{pmatrix} w^+ \\ w^- \end{pmatrix} \geqslant b_N$ and $(w^+, w^-) \geqslant 0$ respectively.

Proposition 3.3. *Suppose Assumption 3.1 holds. Then there exists a sufficiently large compact set* $\mathcal{C} \subset R^{4m+5}$ *such that (i) the intersection of* $\mathcal{C}$ *and the set of KKT pairs of the true problem (3.6), denoted by* Y^**, is nonempty; (ii) for* N *sufficiently* *large, the intersection of $\mathcal{C}$ and the set of KKT pairs of the SAA problem (3.13), denoted by Y_N, is nonempty; (iii) for every $\epsilon > 0$, there exists $N(\epsilon) > 0$ such that*

$$\mathbb{H}(Y_N, Y^*) \leqslant \epsilon$$

for $N \geqslant N(\epsilon)$.

The result is directly implied by Proposition 2.3 in Xu and Zhang (2012). Note that in Proposition 2.3 of Xu and Zhang (2012), they use the condition named "no nonzero abnormal multipliers constraint qualification (NNAMCQ)" to bound the Lagrangian multipliers. This constraint qualification is well known, see Borwein and Zhu (1999) and Ye (2000).

Here the NNAMCQ may not hold in problem (3.6) since we rewrite every equality constraint in (3.4) as two inequality constraints. Note that problems (3.4) and (3.6) are equivalent and $(y_1 - y_3)$ and $(y_2 - y_4)$ are corresponding Lagrangian multipliers of equality constraints, y_5 and s are corresponding Lagrangian multipliers of inequality constraints in problem (3.4). Moreover, by Assumption 3.1, problem (3.4) satisfies MFCQ, which implies NNAMCQ. Then $(y_1 - y_3, y_2 - y_4, y_5, s)$ and $(y_{1,N} - y_{3,N}, y_{2,N} - y_{4,N}, y_{5,N}, s_N)$ are uniformly bounded for N sufficiently large and there exists sufficiently large compact set $\mathcal{C}$ such that Y^* and Y_N are nonempty.

Note that the KKT pairs of problem (3.6) and problem (3.13) are the solutions of the LCP(q, M) in (3.7) and the LCP(q, M_N) in (3.12), Proposition 3.1 shows the convergence analysis between the first phase problem of our two-phase method and implies $\varsigma_N \to \varsigma$ and $\gamma_N \to \gamma$ as $N \to \infty$ almost surely.

Then we move to the convergence analysis of the second phase problem (3.10) and its SAA problem (4.11). We use the Chapter 6, Proposition 6 and Remark 8 in Bonnans and Shapiro (2000) to show the result. Let v^* and v_N denote the optimal value of the true problem (3.10) and the SAA problem (4.11) respectively.

Proposition 3.4. *Suppose Assumption 3.1 holds. Then there exists a sufficiently large compact set* $\mathcal{K} \in R^{2m+5}$ *such that (i) the intersection of* $\mathcal{K}$ *and the optimal solution set of problem (3.10), denoted by* $\mathcal{S}$*, is nonempty; (ii) for* N *sufficiently large the intersection of* $\mathcal{K}$ *and optimal solution set of problem (4.11), denoted by* $\mathcal{S}_N$*, is nonempty; (iii) we have* $v_N \to v^*$ *and* $\mathbb{D}(\mathcal{S}_N, \mathcal{S}) \to 0$ *w.p.1 as* $N \to \infty$ *w.p.1.*

Proof. Obviously, the optimal solution set of problem (3.10) is nonempty and v^* is finite. Note that the feasible sets of problem (3.6) and problem (3.13) are closed and belong to the compact set $\{w^+, w^- \in R^m : 1_m^T(w^+ + w^-) \leqslant \eta, w^+, w^- \geqslant 0\}$. Moreover, we have uniform boundness of $(y_1 - y_3, y_2 - y_4, y_5)$, and $(y_{1,N} - y_{3,N}, y_{2,N} - y_{4,N}, y_{5,N})$ for N sufficiently large by Proposition 2.2 in Xu and Zhang (2012) and the discussion below Proposition 3.3, we have $\mathcal{S}$ and $\mathcal{S}_N$ are nonempty for N sufficiently large.

Moreover, by Proposition 3.3, we have that $\varsigma_N \to \varsigma$ and $\gamma_N \to \gamma$ as $N \to \infty$ w.p.1. Then by the uniform law of large numbers (Bonnans and Shapiro, 2000, Chapter 6, Proposition 7), the constraint functions of problem (4.11) uniformly converge to the constraint functions of problem (3.10). Note that problem (3.10) and problem (4.11) are linear problems, by Proposition 6 and Remark 8 in Chapter 6 of Bonnans and Shapiro (2000), we have $v_N \to v^*$ and $\mathbb{D}(\mathcal{S}_N, \mathcal{S}) \to 0$ *w.p.1* as $N \to \infty$ w.p.1. □

3.3 Applications

In this section, we demonstrate the two-phase approach by a randomly generated example and three empirical examples. We used Matlab R2014a, in a computer with Intel Core 2 Due CPU E8500 3.16GHz for the randomly generated example portfolio construction, Hong Kong and China cross market portfolio construction and FF48 portfolio construction; Matlab R2015a in the service machine with the Intel Xeon E7-4890v2 processor, 2.8GHz, 37.5M Cache, 15 Cores per CPU, 4 CPU, 60 Core in total for S&P500 portfolio construction.

In the examples, we compute the first phase optimization problem by "quadprog" and the second phase optimization problem by "linprog". For the comparative portfolios, we computed the ℓ_1 norm penalty regularized portfolio by "CVX"[1]. For portfolios with cardinality constraint, we compute with a Matlab tool box Yalmip "optimize"[2]. For the ℓ_p penalty regularized portfolio, we approximate the nonsmooth objective function with a smoothing function which has been extensively investigated in Bian and Chen (2014), Chen (2012) and compute it by "fmincon". The distance of the comparative portfolios to the Markowitz portfolio feasible solution set is also calculated. Together, we also compute and compare VaR and CVaR of these portfolios.

3.3.1 Randomly generated example

We randomly generate sample returns of nine asset and applied the two-phase method to construct sparse portfolios.

Example 3.1. *For nine assets in the investment universe, we generate the multivariate normal random variables with mean μ and covariance Σ as the sample of the asset returns. We set every three assets in a group with the same expected return, the same STD, and fully correlated.*

Explicitly, our assets have been arranged in three groups. Asset 1, asset 2 and asset 3 in Group a; asset 4, asset 5, and asset 6 in Group b; asset 7, asset 8 and asset 9 in Group c. The mean returns of the nine assets are $\mu = [10\ 10\ 10\ 1\ 1\ 1\ 3\ 3\ 3]$, and the variance of the nine assets are $\sigma^2 = [5\ 5\ 5\ 1\ 1\ 1\ 2\ 2\ 2]$. The correlations for assets in the same group is 1 and correlations of assets inter-groups are $\rho_{ab} = 0.5$, $\rho_{ac} = 0.3$ and $\rho_{bc} = 0.8$.

[1] Downloaded from http://cvxr.com/cvx/download/.

[2] Downloaded from http://users.isy.liu.se/johanl/yalmip/. Copyright owned by Johan Lofberg.

We calculate and analyze the true optimal value and the approximate optimal value. The true optimal value and the SAA optimal value (STD) for various sample sizes, are displayed respectively in Table 3.1.

Data missing case: Consider a case where the price data of some trading days is missing for some assets, which is very often encountered in the real practice. Then the covariance matrix of the underlying assets could not be easily constructed. In Qi and Sun (2006), a quadratically convergent Newton method is introduced to compute the nearest covariance matrix. Using the above sample, we randomly delete some data, then apply the method in Qi and Sun (2006) to construct the nearest covariance matrix of the portfolio optimization problem. We then use the two-phase approach to solve the reconstructed problem.

The optimal value (variance) at phase 2 in the case of data missing is 6.23 (the equal STD 2.4960) in Table 3.1 Column 11. The two-phase stochastic approach could be accustomed to the case of asset prices data missing.

Convergence analysis: We also conduct tests for different sample size $N = 100, 500, 1000, \cdots, 10000$. For each sample size, we conduct 100 independent exercises. The true optimal value and approximate optimal values with N increasing are shown in Table 3.1. The convergence of the optimal values with N increasing is shown in Figure 4.1.

N	true	500	1500	3000	4500	6000	7500	9000	10000	dmissing
Val	2.489	2.490	2.487	2.487	2.487	2.489	2.488	2.490	2.486	2.496

Table 3.1: Convergence analysis of SAA sparse portfolio optimal value (STD) for Example 4.1

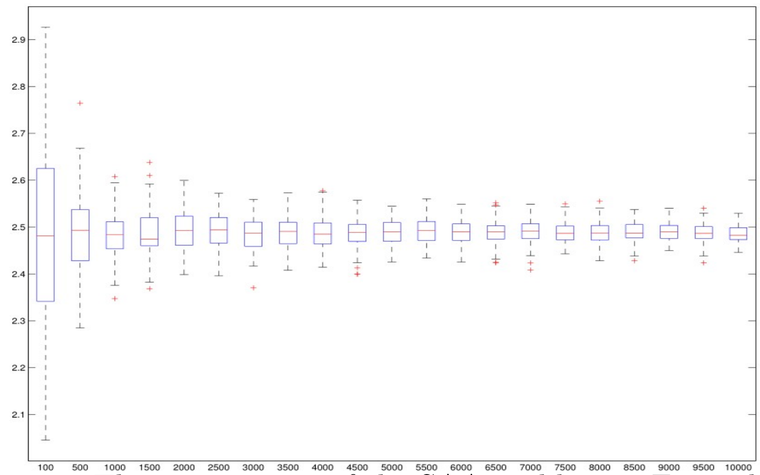


Figure 3.1: The convergence of the SAA problem in Example 4.1.

3.3.2 Empirical applications

In this subsection, we conduct empirical tests. The data sets that we use are returns of S&P 500, 50 Hong Kong and mainland China stocks, and FF48.

We use a rolling window procedure for out-of-sample comparison. T is the total period of returns in the data set; τ is the length of the rolling window; os is the length of the out-of-sample testing window, which is the holding period. The data frequency is daily or monthly, depending on the data set.

Our testing scheme is that we use τ returns as the training data, then use the subsequent os returns as the forecasting data to compute the out-of-sample return, STD, Sharpe ratio and sparsity. We do the exercise from the beginning of our sampling data and roll ahead. For example, our first portfolio selection takes place at the end of the first τ trading days or months. We use the τ historical returns to estimate covariance matrix C and mean μ by SAA method. We then solve the portfolio optimization problem by using the estimated parameters, targeting the required return and compute the weights of optimal solutions. Once a portfolio is thus determined, it is held for the subsequent os trading days or months from $(\tau+1)$ to $(\tau+os)$, and its returns are recorded. We repeat the same process, using returns from 2 to $(\tau+1)$ to construct the portfolio in test 2. The portfolio is observed for the subsequent os trading days or months from $(\tau+2)$ to $(\tau+os+1)$ and their returns are recorded.

To reduce the effects of chance factors, we repeat the same exercise always with a rolling window as described above for 20 rolling times (out-of-sample performance tests). For a given period (whether it is the full period, or the subperiods), all the daily or monthly returns corresponding to this period are used to compute the average return and its STD.

The criteria that we pay attention to are portfolio STD, Sharpe ratio. **Sharpe ratio**, which measures the risk-adjusted return, is a ratio of return and STD in Sharpe (1994),

$$SR = \frac{r_p}{\sigma_p}, \tag{3.14}$$

where r_p is portfolio return; σ_p is portfolio STD. We compute the

$$(\sigma_p)^2 = \frac{1}{20}\frac{1}{os}\sum_{t=\tau}^{19+\tau}\sum_{s=0}^{os-1}(w_t^T r_{t+s+1} - r_p)^2, \tag{3.15}$$

with

$$r_p = \frac{1}{20}\frac{1}{os}\sum_{t=\tau}^{19+\tau}\sum_{s=0}^{os-1} w_t^T r_{t+s+1}. \tag{3.16}$$

It measures the trade-off between returns and volatilities of the portfolios.

We also compute each test Sharpe ratio,

$$SR^t = \frac{r_p^t}{\sigma_p^t}, t = \tau, ..., \tau+19 \tag{3.17}$$

where r_p^t is portfolio return; σ_p^t is portfolio STD at test t. We compute the

$$(\sigma_p^t)^2 = \frac{1}{os} \sum_{s=0}^{os-1} (w_t^T r_{t+s+1} - r_p)^2, \tag{3.18}$$

with

$$r_p^t = \frac{1}{os} \sum_{s=0}^{os-1} w_t^T r_{t+s+1}. \tag{3.19}$$

We compare the value of risk measure between different portfolios by using VaR and CVaR. VaR is a widely used risk measure to evaluate the market risk. The definition of VaR in McNeil et al. (2015) is:

$$VaR_\alpha = \inf\{l \in R : P(L \leqslant l) \geqslant \alpha\}.$$

which says that VaR at confidence level α is given by the smallest number l so that the probability that the loss L exceeds l is at most $(1-\alpha)$ and the definition of CVaR Rockafellar and Uryasev (2000) is:

$$CVaR_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 VaR_\gamma(L) d\gamma.$$

We use the historical simulation method to calculate VaR and CVaR. For the computation of VaR and CVaR, the portfolios are thus held for the subsequent os' days or months after the training period and their daily or monthly out-of-sample returns are observed. The confidence level of the VaR and CVaR is $\alpha = 99\%$.

We also investigate the distance of the ℓ_1 penalty regularized portfolio, cardinality constrained portfolio solution and ℓ_p penalty regularized portfolio to the optimal Markowitz portfolio solution set, which is defined as the shortest distance between the comparative portfolio solution to the optimal solution set:

$$\begin{array}{ll} \min & \| z - w \|_2 \\ \text{s.t.} & z \in \mathcal{H}, \end{array} \tag{3.20}$$

where $\mathcal{H}$ is the solution set of the Markowitz portfolio optimization problem and w is a solution of the comparative portfolio. The optimal value of the optimization problem (3.20) is the distance of the comparative portfolio solution to the optimal solution set.

Example 3.2. *S&P 500. The data set contains the returns of S&P 500 index component stocks of big companies by market capitalization listed on the NYSE or NASDAQ.*

Our sampling data is from Compustat. Sampling data are daily returns of S&P 500 from January 2001 to August 2001. The required return is set as the average of the training sample returns. Our empirical analysis relies on a "rolling-sample" procedure. In this example, $\tau = 100$, $os = 5$ for STD, Sharpe ratio computation, $os' = 20$, $\alpha = 99\%$ for VaR and CVaR computation.

We likewise carry out computation for our comparative portfolios. The empirical results of the forecasting return, STD, Sharpe ratio, sparsity, Var and CVaR are in Table 3.2. The sparsity of the MV model is contained in the bracket of the LCP sparse portfolio column. Abbreviations are LCP sparse portfolio (LCPSP), ℓ_1 penalty regularized portfolio with tuning parameter $\lambda = 0.1$ (L1 0.1), cardinality constrained portfolio with cardinality number 100 (CCPS100) and $1/N$ investment strategy (1/N). The one-by-one test results of Sharpe ratio and sparsity are displayed in Figure 3.2.

S&P 500	LCPSP	ℓ_1 $\lambda = 0.1$	CCPS100	1/N
return	0.001	0.000823	0.0014	-0.00003
STD	0.0024	0.002287	0.0084	0.0074
Sharpe	0.3989	0.359736	0.1694	-0.0045
VaR	0.0041	0.004289	0.0115	0.0131
CVaR	0.0046	0.004292	0.0147	0.0131
sparsity	89(406)	66.25	58.6	500
distance		1.00E-05	3.50E-07	

Table 3.2: S&P 500 Portfolio return, STD, Sharpe Ratio, sparsity, VaR, CVaR and distance.

Example 3.3. *We randomly select 50 stocks which includes 25 Hong Kong stocks and 25 China stocks.*

Data construction: These stocks include 20 stocks in the Shanghai-Hong Kong Stock Connect Scheme and 30 A&H share stocks[3]. We collected the daily adjusted close price from October 2013 to January 2014 of the 50 stocks and calculated the daily returns of stocks constructed from the collected price data.

These stocks include Hong Kong Stock Exchange and Clearing Ltd., China railway group Ltd., GOME Electrical Appliances Holding Ltd., Digital China Holdings Ltd., Brightoil Petroleum Holdings Ltd.,Shandong Weigao Group Medical Polymer Co, Aluminum Corporation of China Ltd., Shanghai Electric Group Co. Ltd., China Everbright Ltd.,Shenzhen International Holdings Ltd. listed in Hong Kong market, and China Merchants Bank, SAIC Motor, Gansu Yasheng Industrial, China Spacesat, Kweichow Moutai, Sichuan Roal and Bridage, Daqin Railway, Huaxia Bank, Industrial and Commercial Bank of China, Bank of China listed in the China market. A and H shares include China Vanke Co Ltd, Ping An insurance, China Pacific Insurance, Huaneng Power International Inc, Anhui Conch Cement, Luoyang Glass,China Minsheng Bank, First Tractor, China CITIC Bank, Jingwei Textile, Jiangsu Express,

[3] A Chinese company could raise capital by issuing A shares and H shares. In other words, the equity structure of a company could be comprised of A share, H share and other shares. H share: shares of company incorporated in mainland China that are traded on the HKEx.

Guangzhou Automobile, Shanghai Pharmaceuticals, Jingcheng Machinery Electric Co. Ltd., and CSSC Offshore & Marine Engineering.

We use the same data rolling scheme. The required return is set as the average of the total returns. The training data window length is $\tau = 25$. The holding period data length is $os = 10$. For VaR and CVaR computation, we use $os' = 20$ for observation and set confidence level $\alpha = 99\%$.

The empirical results of the forecasting return, STD, Sharpe ratio, sparsity, Var and CVaR are in Table 3.3. The sparsity of the MV model is contained in the bracket. The one-by-one rolling results of the tests are displayed in Figures 3.3(a), 3.3(b). Our portfolio outperforms ℓ_1 portfolio (with tuning parameter 0.1), cardinality constrained portfolio (with the cardinality number 20 and 25), ℓ_p portfolio (with p=0.5, tuning parameter 0.015) and $1/N$ investment strategy, denoted by L1 0.1, CCPS 20, CCPS 25, Lp 0.015, and $1/N$ respectively.

After the Shanghai-Hong Kong Stock Connect, there will be soon Shenzhen-Hong Kong Stock Connect[4]. The sparse portfolio allocation research is thus of great importance for cross market investment.

HKCHN	LCPSP	ℓ_1 0.1	CCPS 20	CCPS25	ℓ_p 0.015	1/N
return	0.001296	0.00044	0.001348	0.001583	0.000537	-0.00171
STD	0.007345	0.007115	0.013617	0.012368	0.010248	0.006009
Sharpe	0.1764	0.061903	0.099	0.128	0.0524	-0.2841
VaR	0.013248	0.011514	0.023592	0.024182	0.014756	0.010944
CVaR	0.01408	0.011807	0.026692	0.024382	0.015791	0.010944
sparsity	27(49)	14	19.45	23.1	24.1	
distance		0.0024	0.003406	0.00012	0.144709	

Table 3.3: Hong Kong and Mainland China Cross Market Portfolio return, STD, Sharpe Ratio, sparsity, VaR and CVaR and distance.

Example 3.4. *FF48. It includes 48 industry sector portfolios (abbreviated to FF48).*

[4] Source: Hong Kong Wenhui

Our sampling data is from Fama and French. Sampling data are monthly returns of 48 industry sector value-weighted portfolios in percentage from September 2005 to January 2011.

We use the same data rolling scheme. The required return is set as the average of the total returns on each test. The size of the rolling window is $\tau = 25$. The out-of-sample performance tests use $os = 10$. For VaR and CVaR computation, we use $os' = 20$ for observation and set confidence level $\alpha = 99\%$. Our sample data is rolling ahead monthly. We roll the window and repeat the procedure for 20 tests to generate the average aggregate results.

The empirical results of the forecasting return, STD, Sharpe ratio, sparsity, Var and CVaR are in Table 3.4. The sparsity of the MV model is contained in the bracket. The one-by-one test results of Sharpe ratio and sparsity are displayed in Figures 3.4(a), 3.4(b). Our portfolio outperforms ℓ_1 portfolio (with tuning parameter 0.1), cardinality constrained portfolio (with the cardinality number 18 and 24), ℓ_p portfolio (with p=0.5, tuning parameter 0.015) and $1/N$ investment strategy, denoted by L1 0.1, CCPS 18, CCPS 24, Lp 0.015, and $1/N$ respectively.

FF48	LCPSP	ℓ_1 0.1	CCPS18	CCPS 24	ℓ_p 0.015	1/N
return	-0.1201	-0.13259	-0.6413	-0.3736	-0.3334	-0.7838
STD	5.9265	5.917113	8.5639	7.5703	5.3545	8.002
Sharpe	-0.0203	-0.0224	-0.0749	-0.0493	-0.0623	-0.098
VaR	8.5242	14.59896	12.8905	9.8635	10.0079	7.827
CVaR	10.3835	14.86594	15.514	12.8947	13.5395	10.0896
sparsity	29(48)	24.35	21.45	7.2	31.2	48
distance		0.0044	0.2232	0.5608	0.0938	

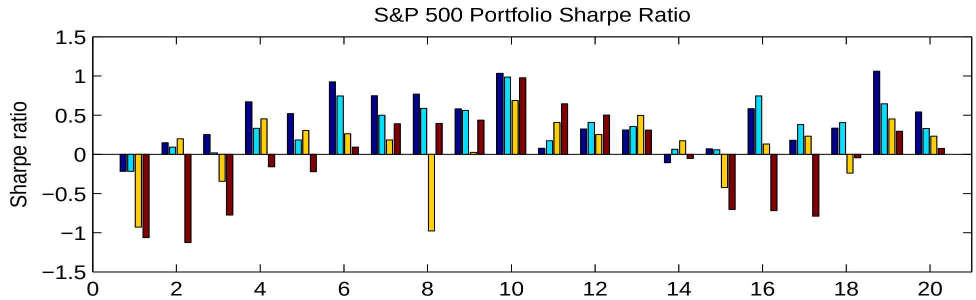
Table 3.4: FF48 Portfolio return, STD, Sharpe Ratio, VaR, CVaR, sparsity and distance.

3.3.3 Discussion of the empirical results

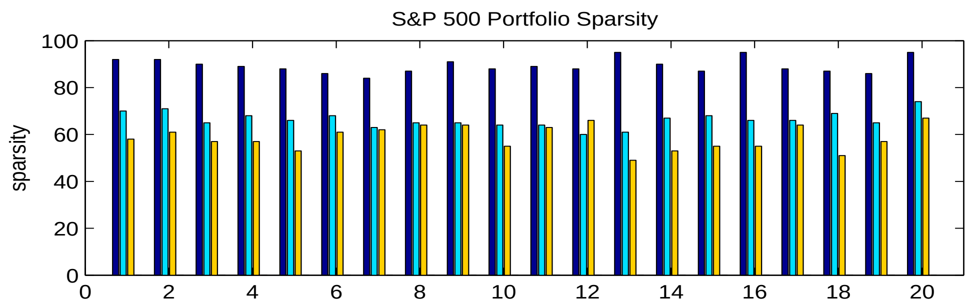
In this subsection, we discuss the out-of-sample performance of our approach and our comparing approaches. Table 3.2, Table 3.3, and Table 3.4 report the out-of-sample performance of the portfolios constructed from the S&P 500, Hong Kong and mainland China cross-market stocks and FF48 data sets. We illustrate each test's Sharpe ratio performance and sparsity by bar graph in Figures 3.2, 3.3 and 3.4.

On one hand, we can observe from the tables and figures in Section 4.2 that the number of assets selected in the our approach are sparser than that in the classical Markowitz portfolios. That means we could construct sparse Markowitz mean-variance portfolios with reduced transaction cost in this way. Note that the simple structure of sparse portfolios enables the portfolio construction to depend less on the human management. Moreover, the sparsity of the sparse Markowitz portfolio solution could provide a reference for the setting of k in the cardinality constrained portfolio selection model.

On the other hand, the empirical results of the examples show that the two-phase optimization scheme could find sparse solutions in the optimal solution set of the Markowitz MV model. The out-of-sample results indicate that the LCP sparse portfolio from our approach can keep the property of the classical Markowitz portfolio very well. That is in all the tests, the STDs of LCP sparse portfolio are reasonably small and in most of the tests, its STDs are the smallest. Moreover, not only the performance of the Sharpe radio, VaR and CVaR are reasonably good, but also the Sharpe radios in all the periods are stabler than the other portfolios in Figures 3.2(a), 3.3(a) and 3.4(a).

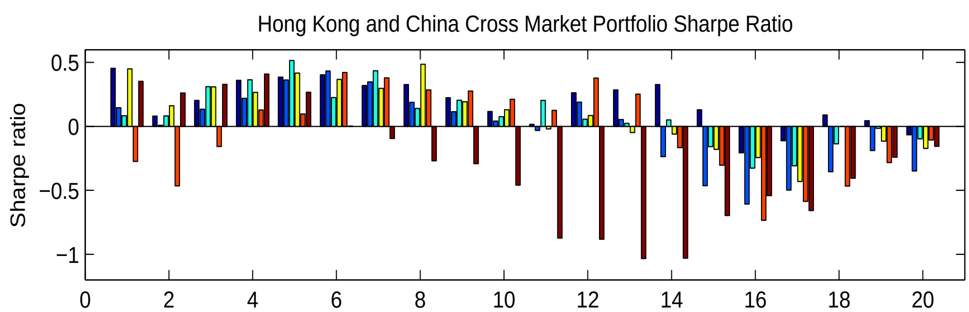


(a)

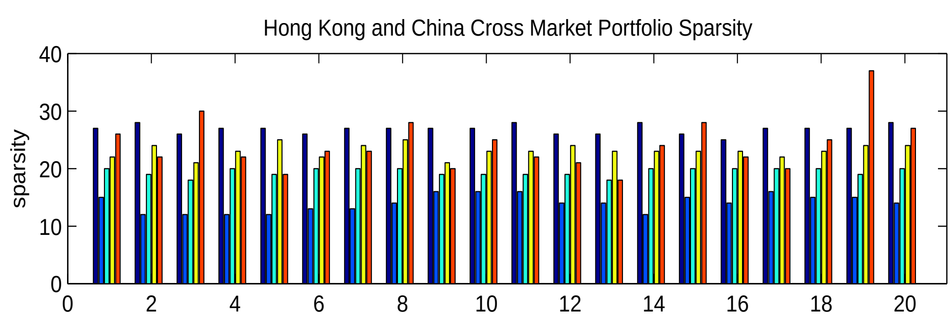


(b)

Figure 3.2: S&P 500 Portfolio (a)Sharpe Ratio, (b)Sparsity. The bar from the left to the right in each test stands for LCP sparse portfolio, ℓ_1 0.1, CCPS100 and 1/N.

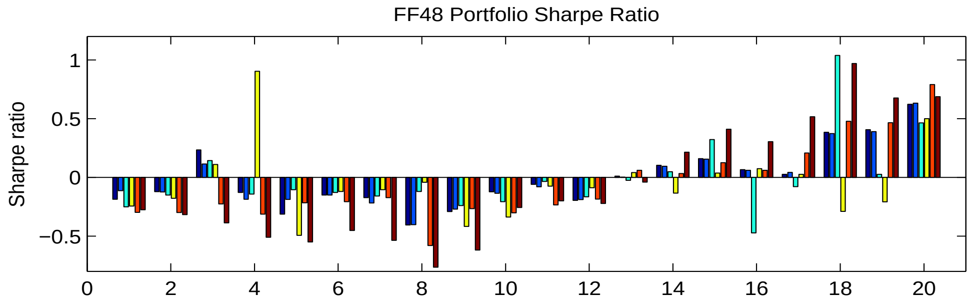


(a)

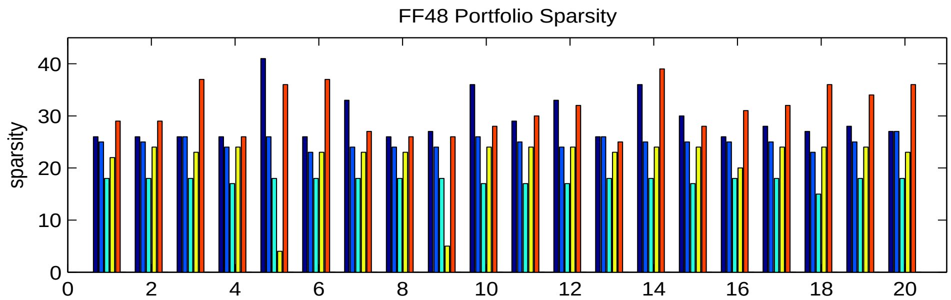


(b)

Figure 3.3: Hong Kong and Mainland China Cross Market Portfolio (a)Sharpe Ratio, (b)Sparsity. The bar from the left to the right in each test stands for LCP sparse portfolio, ℓ_1 0.1, CCPS20, CCPS25, ℓ_p 0.015 and 1/N.



(a)



(b)

Figure 3.4: FF48 Portfolio (a)Sharpe Ratio, (b)Sparsity. The bar from the left to the right in each test stands for LCP sparse portfolio, ℓ_1 0.1, CCPS18, CCPS24, ℓ_p 0.015 and 1/N.

Chapter 4

Sparse Markowitz portfolio selection by penalty methods

In this chapter, we study a penalty method to find the least-p-norm sparse portfolio. The model is formulated to minimize the p-norm of the solutions in the solution set of Markowitz MV model. Using the penalty method solved by the nonmonotone proximal gradient (NPG) algorithm, we construct the sparse portfolio when the optimization problem is nonconvex nonsmooth. We then use the SAA method to handle randomness and prove the convergence. Some properties of the solutions are derived. We tested the model by a randomly generated example and several empirical examples which demonstrated the performance of the proposed model.

This chapter is organized as follows. In Section 4.2, we introduce the penalty method to find a sparse Markowitz portfolio. In Section 4.3, we analyze the convergence of the SAA method. In Section 4.4, we show the numerical and empirical applications of the model. Section 4.5 is a conclusion.

4.1 Model construction and stochastic penalty method

We consider the least-p-norm sparse Markowitz portfolio optimization problem as follows:

$$\begin{array}{ll} \min & \| w \|_p^p \\ \text{s.t.} & w \in \mathcal{H}, \end{array} \tag{4.1}$$

where $\mathcal{H}$ is the optimal solution set of the Markowitz portfolio optimization problem (2.1). We call the portfolio computed from the model the least-p-norm sparse portfolio.

Sparse solution of the optimal solution set of the Markowitz portfolio optimization problem (2.1) is defined as:

$$\begin{array}{ll} \min & \| w \|_0 \\ \text{s.t.} & w \in \mathcal{H}. \end{array} \tag{4.2}$$

We show in Theorem 4.1 that under some condition, problem (4.1) and problem (4.2) are equivalent.

Problem (4.1) is a nonconvex non-Lipschitz NP-hard optimization problem, see Chen et al. (2016), Ge et al. (2011) and may not be easy to be solved. Inspired by the method in Chen and Xiang (2016) and Chen et al. (2016), we provide the penalty method to approximate problem (4.1) in the following subsection.

4.1.1 Penalty method for a least-ℓ_p-norm solution of the MV model

The basic idea of this approach is that we reformulate the feasible set of problem (4.1) (solution set of problem (2.1)) as a linear system and find its least-p-norm solution.

We can find the least-p-norm solution from the optimal solution set of the MV model by solving

$$\begin{aligned} \min \quad & \|w\|_p^p \\ s.t. \quad & Cw + y_1\mu + y_2 1_m = 0 \\ & w^T\mu - \rho = 0 \\ & w^T 1_m = 1, \end{aligned}$$

where the feasible set of above problem is the optimal solution set of MV model problem (2.1). It is derived from the first-order Karush-Kuhn-Tucker condition of the MV model with y_1 and y_2 as the Lagranger multipliers of the two constraints of problem (2.1). Define A as a $(m+2)\times(m+2)$ matrix, b as a $m+2$ vector, where $A = \begin{pmatrix} C & \mu & 1_m \\ \mu^T & 0 & 0 \\ 1_m^T & 0 & 0 \end{pmatrix}$, $b = \begin{pmatrix} 0_m \\ \rho \\ 1 \end{pmatrix}$, $x = \begin{pmatrix} w \\ y \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$. The above problem could be approximated by:

$$\begin{aligned} \min \quad & \|w\|_p^p \\ s.t. \quad & \|Ax - b\| \leqslant \sigma, \end{aligned} \tag{4.3}$$

for a small positive number σ.

We denote the solution set of problem (4.3) as $\mathcal{S}$. Note that $\|w\|_p^p$ is not locally Lipschitz continuous at some points. Then we can approach problem (4.3) by the penalty method provided in Chen et al. (2016). We write the penalty form of problem (4.3) as follows:

$$\min \quad \|w\|_p^p + \lambda(\|Ax - b\|^2 - \sigma^2)_+, \tag{4.4}$$

where λ is the penalty parameter. This λ is, indeed, hard to estimate in practise. To circumvent the problem, the penalty method that we used gradually increases the parameter.

The least-p-norm sparse portfolio model has several advantages:

- Sparsity: The sparsity in optimization problems already exists in many application cases in the literature, such as image, compressed sensing, e.g. Chen et al. (2016), Chen et al. (2017), Chen et al. (2013). The non-Lipschitz nonconvex l_p penalty is a widely used technique to find sparse solutions. It is natural to deploy this idea in the portfolio management. It is desirable in the real investment that a limited number of positions are created and managed.

- Stability: It is possible that assets are collinear. It is reported that when the sample size goes to large, the MV model solution performs well in DeMiguel et al. (2009b) and DeMiguel et al. (2009a). We use the least-p-norm optimization to reduce the sensitivity of the portfolio between the assets. It is possible to use limited training sample when the least-p-norm optimization boasts stability property.

- Transaction cost saving: In addition to the wise portfolio selection, investors concern themselves with the practical problem of transaction cost. Transaction cost is a non-trivial issue when investors execute trades. As discussed in Brodie et al. (2009), large investors' transaction cost is linear to the amount of the assets bought; small investors' transaction cost depends on the number of assets chosen. Therefore, reducing number of assets in portfolio benefits both two types of investors.

4.1.2 Some features of the least-p-norm solution

In this subsection, we show some features of the least-p-norm solution.

Lemma 4.1. *All least-p-norm solutions with $p \in (0,1)$ of the $\mathcal{H}$, the optimal solution set of problem (2.1), are extreme points of $\mathcal{H}$.*

The proof is immediately obtained from Theorem 3 in Ge et al. (2011).

It will be very interesting to consider the upper bound of nonzero entries in a least-p-norm solution and relation of the sparse (least-0-norm) and least-p-norm solutions of the classical Markowitz portfolio optimization problem. We will discuss it in the following proposition.

Proposition 4.1. *Let $\tilde{w}$ be a solution to problem (4.1) and $\bar{w}$ be a solution to problem (4.2). Then*

1. *Consider problem (4.1), the number of nonzero entries in a solution of problem (4.1) is at most* $\text{rank}(C) + 2$*, where C is the covariance matrix.*

2. *There is a $\bar{p} \in (0, 1)$ such that $\|(\tilde{w}, \tilde{y})\|_0 = \|(\bar{w}, \bar{y})\|_0$ for all $p \in (0, \bar{p})$, where $(\tilde{w}, \tilde{y})$, $(\bar{w}, \bar{y})$ stands for $(\tilde{w}^T, \tilde{y}^T)^T$, $(\bar{w}^T, \bar{y}^T)^T$, being the least-p-norm and least-0-norm solution in $\mathcal{H}$.*

Proof. 1. Denote the matrix A as two blocks such that $A = \begin{pmatrix} B & D \end{pmatrix}$, where $B = \begin{pmatrix} C \\ \mu^T \\ 1_m^T \end{pmatrix}$ and $D = \begin{pmatrix} \mu & 1_m \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$. Let $G = \begin{pmatrix} \mu^T \\ 1_m^T \end{pmatrix}$ be the matrix contains the coefficients of the constraints. From rank(G)=2, it is obvious that $\text{rank}(B) \leqslant \text{rank}(C) + 2$. For any fixed y, by (Ge et al., 2011, Theorem 3), we have

$$\|\tilde{w}\|_0 \leqslant \text{rank}(B) \leqslant \text{rank}(C) + 2.$$

2. Similar to the second part of Theorem 2.1 in Chen and Xiang (2016), let $\{x^1, x^2, ..., x^m\}$ be the set of extreme points of $\mathcal{H}$. $\tilde{x} = (\tilde{w}, \tilde{y})$, $\bar{x} = (\bar{w}, \bar{y})$. Then we have for all $\bar{x}$,

$$\|\bar{x}\|_p^p \geqslant \min\{\|x^1\|_p^p, \|x^2\|_p^p, ..., \|x^m\|_p^p\} = \|\tilde{x}\|_p^p. \tag{4.5}$$

Then we prove the second part of the proposition by contradiction. If the part 2 of the proposition does not hold, then there exist a sequence $\{p_i\}$, $p_i > 0$, $p_i \to 0$

as $i \to \infty$ and a sequence $\{x^j\}$ of extreme points of $\mathcal{H}$, where x^{j_i} is a least-p-norm solution and we have

$$\|x^{j_i}\|_0 > \|\bar{x}\|_0. \tag{4.6}$$

Since there are finite many i, without loss of generality, we assume $j_i = j$, we have

$$\|\bar{x}\|_0 = \lim_{p_i \downarrow 0} \|\bar{x}\|_{p_i}^{p_i} \geqslant \lim_{p_i \downarrow 0} \|x^j\|_{p_i}^{p_i} = \|x^j\|_0 \tag{4.7}$$

from (4.5). It contradicts to (4.6). Hence we obtain the second part of the proposition. □

Remark 4.1. *From the conclusion of Proposition 4.1 part 2 and* $0 \leqslant \|\tilde{x}\|_0 - \|\tilde{w}\|_0 \leqslant 2$, $0 \leqslant \|\bar{x}\|_0 - \|\bar{w}\|_0 \leqslant 2$, *where the difference is the number of Lagranger multipliers, we have the relationship between problems (4.1) and (4.2) for some* $p \in (0, \bar{p})$ *is* $-2 \leqslant (\|\tilde{w}\|_0 - \|\bar{w}\|_0) = (\|\bar{x}\|_0 - \|\bar{w}\|_0) - (\|\tilde{x}\|_0 - \|\tilde{w}\|_0) \leqslant 2$, *i.e.,* $|\|\tilde{w}\|_0 - \|\bar{w}\|_0| \leqslant 2$.

Corollary 4.1. *There is an extreme point* $\bar{x}$ *of* $\mathcal{S}$ *such that* $\bar{x}$ *is a sparse solution of the* $\mathcal{H}$.

4.1.3 Exact penalty

We apply the Theorem 3.2 in Chen et al. (2016) to demonstrate that the penalty method could find the local minimizer of (4.3). Our problem is a special case of Theorem 3.2 in Chen et al. (2016), where the constraints are equalities. In the paper Chen et al. (2016), a class of non-lipschitz optimization problem is researched with constraints comprised of a simple polyhedron S_1 and $S_2 = \{x : \|Ax - b\| \leqslant \sigma, Bx \leqslant h\}$. In this chapter, we consider a special case where S_1 is the full space and $S_2 = \{x : \|Ax - b\| \leqslant \sigma\}$.

Assumption 4.1. *(blanket assumption on (4.3))* $\|w\|_p^p$ *is a nonnegative continuous function. The feasible set of (4.3) is* $S := \{x : \|Ax - b\| \leqslant \sigma\}$. *Moreover,* A *has full low rank and there exists* $x_0 \in S$ *so that* $\|Ax_0 - b\| < \sigma$.

Theorem 4.1. *Suppose that w^* is a local minimizer of (4.3). Then there exists a $\lambda^* > 0$ such that w^* is a local minimizer of (4.4) whenever $\lambda \geqslant \lambda^*$.*

Proof. The objective function $\Phi(x)$ in the paper Chen et al. (2016) is $\Phi(x) = \| \begin{pmatrix} 1_m & 0_2 \end{pmatrix}^T x\|_p^p = \|w\|_p^p$ in our case, where 1_m is a m-dimensional vector with all entries being 1 and 0_2 is a 2-dimensional vector with all entries being 0. It is a nonnegative continuous function. Moreover, the bridge penalty function $\|w\|_p^p$ is not locally Lipschitz continuous at the point 0. To proceed, it is easy to see that $\|w\|_p^p$ satisfies Assumption 3.1 from Chen et al. (2016).

Applying Theorem 3.2, C_0 defined in Lemma 2.1 and Corollary 3.1 in Chen et al. (2016) with $\Omega_2 = \{x_I : \|A_I x_I - b\| \leqslant \sigma\}$, we immediately obtain the desired results in this theorem. □

We solve a sequence of the smooth problem in the form:

$$\min \quad F_{\lambda,\mu}(w) := \|w\|_p^p + f_{\lambda,\mu}(x), \tag{4.8}$$

for some penalty parameter λ and smoothing parameter μ, where

$$f_{\lambda,\mu}(x) := h_{\lambda,\mu}(\|Ax - b\|^2 - \sigma^2), \tag{4.9}$$

$$h_{\lambda,\mu}(s) := \lambda \max_{0 \leqslant t \leqslant 1} \{st - \frac{\mu}{2}t^2\}. \tag{4.10}$$

We now describe the algorithm to solve the least-p-norm sparse portfolio model (4.8).

Penalty method for problem (4.8)

Let w^{fsbl} be an arbitrary feasible point of problem (4.1). Choose $w^0 \in \mathcal{H}$, $\lambda_0 > 0$, $\mu_0 > 0$, $\epsilon_0 > 0$, $\rho > 1$, and $\theta \in (0, 1)$ arbitrarily. Set $k = 0$ and $x^{0,0} = x^0 \in \mathcal{H}$.

1: If $F_{\lambda_k,\mu_k}(x^{k,0}) > F_{\lambda_k,\mu_k}(x^{fsbl})$, set $x^{k,0} = x^{fsbl}$ and apply NPG method with $x^{k,0}$ as the initial point to find x^k to the problem with $\lambda = \lambda_k$ and $\mu = \mu_k$ satisfying

$$d(0, \nabla f_{\lambda_k,\mu_k}(x^k) + \partial(x^p)) \leqslant \epsilon_k$$

2: Set $\lambda_{k+1} = \rho\lambda_k, \mu_{k+1} = \theta\mu_k, \epsilon_{k+1} = \theta\epsilon_k$, and $x^{k+1,0} = x^k$
3: Set $k \leftarrow k + 1$, go to step 1

4.2 The SAA method and convergence analysis

In this section, we use the SAA method to model the randomness of the Markowitz portfolio and consider the convergence analysis between the original problem and the approximation problem.

Let $\{r^j = (r_1^j, ..., r_m^j)\}_{j=1}^N$ be the i.i.d. samples of random variable r. Then the SAA mechanism of the Markowitz portfolio optimization is:

$$\begin{array}{rl} \min & \|w\|_p^p \\ \text{s.t.} & \|A_N x - b\| \leqslant \sigma, \\ & x \in \mathcal{C} \end{array} \tag{4.11}$$

where

$$C_N = \frac{1}{N}\sum_{j=1}^N (r^j - \mu_N)(r^j - \mu_N)^T,$$

$\mu_{i,N} = \frac{1}{N}\sum_{j=1}^N r_i^j$, $\mu_N = (\mu_{1,N}, \ldots, \mu_{m,N})^T$, $A_N = \begin{pmatrix} C_N & \mu_N & 1_m \\ \mu_N^T & 0 & 0 \\ 1_m^T & 0 & 0 \end{pmatrix}$, and $x_N = (w_N^T, y_N^T)^T$ is a solution of the above problem. Additionally, we assume that the solution belongs to a compact set $x \in \mathcal{C}$. Denote the solution set of problem (4.11) as $\mathcal{S}_N$.

It follows that we consider the convergence analysis between the original problem (4.3) and its SAA form (4.11).

Assumption 4.2. *There exists a compact set $\mathcal{C}$ such that $\mathcal{S} \cap \mathcal{C} \neq \emptyset$, for N sufficiently large, $\mathcal{S}_N \cap \mathcal{C} \neq \emptyset$.*

Assumption 4.2 implies the compactness of $\{w \in R^m : \|A_N(w^T, y^T)^T - b\| \leqslant \sigma\}$. Then we move to the convergence analysis of problem (4.3) and its SAA problem (4.11). We use the Chapter 6, Proposition 6 and Remark 8 in Bonnans and Shapiro (2000) to show the result.

Proposition 4.2. *Suppose Assumption 4.2 holds. Denote the set $\mathcal{S}$ and v^* as the optimal solution set and optimal value of the true problem (4.3) and $\mathcal{S}_N$ and v_N as the optimal solution set and optimal value of the SAA problem (4.11) respectively. Assume $\mathcal{S}$ is nonempty and v^* is finite. Then we have $v_N \to v^*$ and $\mathbb{D}(\mathcal{S}_N, \mathcal{S}) \to 0$ w.p.1 as $N \to \infty$ w.p.1.*

Proof. As we have the optimal solution sets of problem (4.3) and problem (4.11) are contained in a compact set, by the uniform law of large numbers (Bonnans and Shapiro, 2000, Chapter 6, Proposition 7), the constraints function of problem (4.11) uniformly converge to the constraints function of problem (4.3). In other words, the convergence between the the feasible sets of problem (4.3) and its SAA problem (4.11) is almost surely.

Note that problem (4.3) and problem (4.11) are optimization problems with linear constraints, by Proposition 6 and Remark 8 in Chapter 6 of Bonnans and Shapiro (2000), we have $v_N \to v^*$ and $\mathbb{D}(\mathcal{S}_\mathcal{N}, \mathcal{S}) \to 0$ *w.p.1* as $N \to \infty$ w.p.1. □

4.3 Applications

In this section, we demonstrate the least-p-norm sparse portfolio model by a randomly generated example and three empirical examples. We use Matlab R2014a, in a computer with Intel Core 2 Due CPU E8500 3.16GHz.

4.3.1 Randomly generated example

We randomly generate sample returns of nine assets and applied the penalty method to construct least-p-norm sparse portfolios.

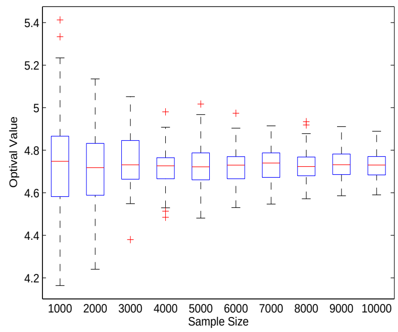
Example 4.1. *For the nine assets in the investment universe, we generate the multivariate normal random variables with mean μ and covariance Σ as the sample of the asset returns. We set every three assets in a group with the same expected return, the same standard deviation, and fully correlated.*

Explicitly, our assets have been arranged in three groups. Asset 1, asset 2 and asset 3 in Group a; asset 4, asset 5, and asset 6 in Group b; asset 7, asset 8 and asset 9 in Group c. The mean returns of the nine assets are $\mu = [10\ 10\ 10\ 1\ 1\ 1\ 3\ 3\ 3]$, and the variance of the nine assets are $\sigma^2 = [5\ 5\ 5\ 1\ 1\ 1\ 2\ 2\ 2]$. The correlations for assets in the same group is 1 and correlations of assets inter-groups are $\rho_{ab} = 0.5$, $\rho_{ac} = 0.3$ and $\rho_{bc} = 0.8$.

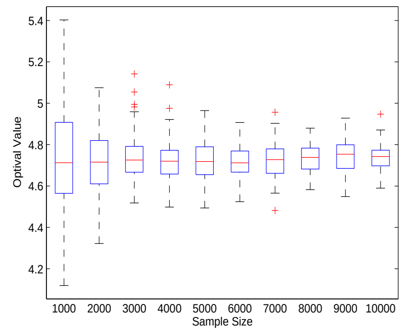
We calculate and analyze the true optimal value and the SAA approximate optimal value. The true optimal value and the SAA optimal value (standard deviation) for various sample sizes, are displayed respectively in Table 4.1. $v^*_{p=0.3}$ stands for the optimal value with $p = 0.3$. Same for $v^*_{p=0.5}$ and $v^*_{p=0.7}$.

Convergence analysis: We conduct tests with different sample size $N = 1000, 2000, \cdots, 10000$. We could observe from the box plot of these results that the optimal values of problem (4.11) converges with N increasing in Figure 4.1. For each sample size, we conduct 100 independent exercises. The true optimal value and approximate optimal values with N increasing are shown in Table 4.1.

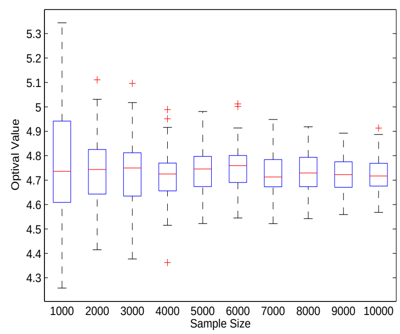
Figure 4.1: $\|\cdot\|_p$, $p = 0.3, 0.5, 0.7$ sparse portfolio SAA convergence



(a)



(b)



(c)

Table 4.1: Convergence results for the $\|\cdot\|_p$ sparse portfolio model

	TRUE	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
$v^*_{p=0.3}$	4.7338	4.7317	4.7276	4.7518	4.7215	4.7266	4.7236	4.7322	4.7292	4.7344	4.7283
$v^*_{p=0.5}$	4.7334	4.7222	4.7216	4.7358	4.7258	4.7214	4.7231	4.7255	4.7380	4.7441	4.7368
$v^*_{p=0.7}$	4.7335	4.7514	4.7367	4.7352	4.7192	4.7384	4.7509	4.7207	4.7312	4.7242	4.7224

4.3.2 Empirical applications

In this subsection, we conduct empirical tests. The data sets that we use are returns of 20 A&H Hong Kong and China bank and insurance stocks, FF12 and FF25 .

We use a rolling window procedure for the out-of-sample comparison. τ is the length of the rolling window; os is the length of the out-of-sample testing window, which is the holding period. The data frequency is daily (20 A & H stocks) or monthly (FF12, FF25).

Our testing scheme is that we use τ returns as the training data to construct the portfolio, then use the subsequent os returns as the out-of-sample data to compute the out-of-sample returns. We do the exercise from the beginning of our sampling data and roll ahead. For example, our first portfolio selection takes place at the end of the first τ trading days or months. We use the τ historical returns to estimate covariance matrix C and mean μ by the SAA method. We solve the portfolio optimization problem by using the estimated C, μ, the required return taken as the average of the μ and compute the weights. Once a portfolio is thus determined, it is held for the subsequent os trading days or months from $(\tau + 1)$ to $(\tau + os)$, and its returns are recorded. We repeat the same process, using returns from 2 to $(\tau + 1)$ to construct the portfolio in test 2. The portfolio is observed for the subsequent os trading days or months from $(\tau + 2)$ to $(\tau + os + 1)$ and their returns are recorded.

To reduce the effects of chance factors, we repeat the same exercise as described above for T rolling times (out-of-sample performance tests). For a given period, all the out-of-sample returns corresponding to this period are used to compute its average return, STD, Sharpe ratio and sparsity.

The comparing portfolios that we consider are the least-0-norm sparse portfolio, modeled by (2.1); regularized method which use penalty to regularize the Markowitz portfolio optimization such as ℓ_1 penalty.

$$\begin{array}{ll} \min & w^T C w + \lambda \|w\|_1 \\ \text{s.t.} & w^T 1_m = 1, \quad w^T \mu = \rho; \end{array} \tag{4.12}$$

cardinality constrained portfolio selection (CCPS) model which adds a cardinality constraint in the Markowitz portfolio selection model to select a cardinality number of assets in the portfolio,

$$
\begin{aligned}
\min \quad & w^T C w \\
\text{s.t.} \quad & w^T 1_m = 1, \quad w^T \mu = \rho \\
& \|w\|_0 \leqslant k,
\end{aligned}
$$

in many investigations.

In the examples, we compute the least-p-norm sparse portfolio model by the penalty method. For the comparative portfolios, we compute the sparse portfolio via a two-phase linear complementarity approach by "quadprog" at the first phase and "linprog" at the second phase; the ℓ_1 norm penalty regularized portfolio by "cvx"; the cardinality constrained portfolio with a Matlab tool box Yalmip "optimize"[1].

The criteria that we pay attention to is portfolio Sharpe ratio. **Sharpe ratio**, which measures the risk-adjusted return, is a ratio of return and STD in Sharpe (1994) and defined in (3.14) to (3.16). It measures the trade-off between returns and volatilities of the portfolios.

We compute the sparsity by taking the average of the number of assets selected in each test during the full period.

Example 4.2. *We select 20 A & H share stocks which includes 10 Hong Kong stocks and 10 China stocks.*

Data construction: These stocks include 10 financial companies, each of them issuing both A and H share stocks. There are 20 A & H share stocks of them[2]. We collect the daily adjusted close price from April 3, 2012 to July 16, 2012 of the 20 stocks and calculate the daily returns of stocks constructed from the collected price data. We repeat the experiments for 60 times.

[1] Downloaded from http://users.isy.liu.se/johanl/yalmip/. Copyright owned by Johan Lofberg.

[2] A Chinese company could raise capital by issuing A shares and H shares. In other words, the equity structure of a company could be comprised of A share, H share and other shares. H share: shares of company incorporated in mainland China that are traded on the HKEx.

In the H share market, there is a term "eight banks five insurances" to indicate the bank and insurance company shares from China. The eight banks (A and H share code in the bracket following each company's name) are Industrial and Commercial Bank of China (ICBC 1398.HK 601398.SS), Bank of China (BOC 3988.HK 601988.SS), China Construction Bank (CCB 0939.HK 601939.SS), Bank of Communications (BC 3328.HK 601328.SS), China Merchants Bank (CMB 3968.HK 600036.SS), China CITIC Bank (CITIC 0998.HK 601998.SS), China Minsheng Bank (MSB 0188.HK 600016.SS) and Agricultural Bank of China (ABC), which joined in 2010. We do not include ABC in our sample, for its IPO date is later than 2009. The five insurance companies are China Life Insurance (2628.HK 601628.SS), Ping An Insurance (2318.HK 601318.SS), China Pacific Insurance (2601.HK 601601.SS), China Taiping Insurance and PICC property and Casualty. The later two companies are only listed in the Hong Kong stock market, so we do not consider.

We use the data rolling scheme as described above. The required return is set as the average of the sample means in each test. The training data window length is $\tau = 15$. The holding period data length is $os = 1$.

The empirical results of the out-of-sample return, STD, Sharpe ratio, sparsity are in Table 4.2. Our portfolio outperforms MV model, least-0-norm sparse portfolio, ℓ_1 portfolio (with tuning parameter 0.1), cardinality constrained portfolio (with the cardinality number 17) and $1/N$ investment strategy, denoted by MV, $\|\cdot\|_0$, ℓ_1 0.1, CCPS 17 and $1/N$ respectively.

The sparse portfolio allocation research is of great importance for cross market investment.

Table 4.2: Empirical results for A&H shares

	$\ell_{0.3}$	$\ell_{0.5}$	$\ell_{0.7}$	MV	$\|\cdot\|_0$	ℓ_1 0.1	CCPS17	1/N
r_p	-0.1535	-0.0794	-0.1007	-0.1577	-0.1537	-0.2276	-0.4200	-0.1653
σ_p	1.1130	1.0262	0.9747	1.1143	1.1091	0.8982	1.2419	0.9643
SR	-0.1379	-0.0774	-0.1033	-0.1415	-0.1386	-0.2534	-0.3382	-0.1714
sparsity	14.50	14.20	15.20	19.90	19.90	15.70	16.70	

Example 4.3. *FF12. 12 industry sector portfolios monthly return data (abbreviated to FF12).*

We use the same data rolling scheme. The required return is set as the average of the return mean in each test. The training data window length is $\tau = 8$. The holding period data length is $os = 1$. Our experiment data are from April 2008 to January 2013. We repeat the experiments for $T = 50$ times.

The empirical results of the out-of-sample return, STD, Sharpe ratio, sparsity are in Table 4.3. Our portfolio outperforms MV model, least-0-norm sparse portfolio, ℓ_1 portfolio (with tuning parameter 0.1), cardinality constrained portfolio (with the cardinality number 9) and $1/N$ investment strategy, denoted by MV, $\|\cdot\|_0$, ℓ_1 0.1, CCPS 9 and $1/N$ respectively.

Table 4.3: Empirical results for FF12

	$\ell_{0.3}$	$\ell_{0.5}$	$\ell_{0.7}$	MV	$\|\cdot\|_0$	ℓ_1 0.1	CCPS9	1/N
r_p	1.3728	1.4090	1.3573	1.3521	1.3456	1.4501	1.1217	-0.0687
σ_p	4.5138	4.4369	4.2259	4.4898	4.4960	4.8551	4.5394	4.1610
SR	0.3041	0.3176	0.3212	0.3011	0.2993	0.2987	0.2471	-0.0165
sparsity	9.98	9.96	10.04	12	12	8.98	8.54	

Example 4.4. *FF25. The data set are 25 portfolios monthly average value weighted returns formed on size and book-to-market* (5×5)*.*

We use the same data rolling scheme and setting for the required return. The training data window length is $\tau = 20$. The holding period data length is $os = 1$.

Our experiment data are from February 1968 to December 1974. We repeat the experiments for $T = 50$ times.

The empirical results of the out-of-sample return, STD, Sharpe ratio, sparsity are in Table 4.4. Our portfolio outperforms MV model, least-0-norm sparse portfolio, ℓ_1 portfolio (with tuning parameter 0.1), cardinality constrained portfolio (with the cardinality number 23) and $1/N$ investment strategy, denoted by MV, $\|\cdot\|_0$, ℓ_1 0.1, CCPS23 and $1/N$ respectively.

Table 4.4: Empirical results for FF25

	$\ell_{0.3}$	$\ell_{0.5}$	$\ell_{0.7}$	MV	$\|\cdot\|_0$	ℓ_1 0.1	CCPS23	1/N
r_p	0.7063	0.7145	0.7043	0.6807	0.7067	0.6860	0.1316	0.2387
σ_p	9.7032	9.6954	9.6928	9.6842	9.7052	10.3757	6.1313	6.7813
SR	0.0728	0.0737	0.0727	0.0703	0.0728	0.0661	0.0215	0.0352
sparsity	24.78	24.84	24.88	25	24.98	20.92	21.24	

4.3.3 Discussion of the empirical results

In this subsection, we discuss the out-of-sample performance of our approach and our comparing approaches. Tables 4.2, 4.3, 4.4 report the out-of-sample performance of the portfolios constructed from the 20 A & H stocks, FF12 and FF25 data sets. In the case when the MV model has multiple solutions, our method combines the advantage of the MV model and its sparsity modification methods (least-0-norm sparse portfolio, ℓ_1 regularization, CCPS and so on).

Sparsity and Sharpe ratio

On the one hand, we can observe from the tables in Section 4.2 that we could select sparser portfolio than portfolio directly constructed from MV model. Ideally, the trading cost is linearly related to the number of assets in the portfolio, that means transaction cost is saved for investors. Moreover the simple structure of the $\|\cdot\|_p$ sparse portfolios enables the portfolio construction independent of parameter setting.

Moreover, the sparsity of the sparse Markowitz portfolio solution could provide a reference for the collinearity of the assets.

On the other hand, the empirical results of the examples show that the least-p-norm sparse portfolio model could find sparse solutions in the optimal solution set of the Markowitz MV model. The out-of-sample results indicate that the least-p-norm sparse portfolio from our approach can keep the return-risk balance very well. That is in all the tests, the sparse portfolio are reasonably well in the SR.

Robustness

We also check the robustness of our method. Observing from Tables 4.2, 4.3, 4.4, we could conclude that for p takes a number of value in (0,1), our method could outperform other portfolios.

For a different sample period, we do the test again. For dataset FF12, if we use the data from August 1991 to January 2013 and $T = 200$ in this case, the least-p-norm sparse portfolio could outperform all the other portfolios except ℓ_1 portfolio. For dataset 20 A&H stocks, if we use the data from February 7, 2012 to July 16, 2012 and $T = 100$ in this case, the $\| \cdot \|_{0.5}$ sparse portfolio could outperform all the other portfolios and the $\| \cdot \|_p$ with $p = 0.3, 0.7$ sparse portfolio could outperform all but 1/N portfolio. For dataset FF25, if we increase the sample period by twice, the sparse portfolio perform well than others except the cardinality portfolio. It shows in most cases, our method could outperform.

Since this penalty method is designed for non-convex, non-Lipschitz optimization problem, it is sensitive to the initial value taken, we use the sparse solution as the initial value and it works better than taking a feasible point by A/b, which might result in extreme results in our examples.

Chapter 5

Dynamic portfolio investment via parameterized expected return evolution

This chapter considers parameter identification in the dynamic portfolio selection by minimizing the deviation of the expected returns to the target returns with a constraint of a parametric ODE involving QP. We assume in this chapter that the dynamics of expected return is a first-order ODE solution with unknown parameters as the coefficients. At every time step, the investors make investment in the framework of the Markowitz MV model with the expected return given by the ODE. To estimate the parameters, it is then using the quasi-Newton method to obtain the unknown parameters in the coefficients of the ODE. We re-balance the portfolio at the end of a round of dynamic investment by using the updated parameter values. Numerical scheme is obtained by using the time-stepping method. Parameters are identified discretely. A numerical example is presented to show the performance of the dynamic portfolio.

The rest of this chapter is organized as follows. In Section 5.2, we present the proposed model, and give detailed analysis of the model. In Section 5.3, we discretize the model by the time-stepping method and provide the way to solve it. In Section

5.4, an empirical example is given. In Section 5.5, we summarize the chapter.

5.1 Model formulation

5.1.1 The proposed model

In this section, we study a parameter identification model with ODE constraint involving MV model applied in the portfolio management. We consider m available securities. At time t, we use $w(t) = (w_1(t), w_2(t), ..., w_m(t))^T$ to be the weight of asset allocation, 1_m to be the m-dimensional vector with all entries equal to one, and $\rho(t)$ to be the expected portfolio return; t in the range $[0, T]$. The Markowitz (1952) MV optimization model portfolio has minimal risk for a given level of expected return by solving the following QP problem at time t:

$$\begin{array}{ll} \min & \frac{1}{2} w(t)^T C(t) w(t) \\ \text{s.t.} & w(t)^T \mu(t) = \rho(t) \\ & w(t)^T 1_m = 1, \end{array} \tag{5.1}$$

where $r(t) = (r_1(t), r_2(t), ..., r_m(t))^T$ is the return of the securities; the expected return $\mu(t) = E[r(t)]$ and the covariance matrix of the returns is

$$C(t) = E[(r(t) - \mu(t))(r(t) - \mu(t))^T].$$

We consider the case where $C(t)$ is positive-definite for any $t \in [0, T]$. Therefore, at each time step, the solution of the MV model is unique. Denote the solution set of problem (5.1) as $\mathcal{S}(t, \rho(t))$.

We define the parametric model generally. We consider $\rho(t)$ as the state variable and $w(t)$ as the control variable. We use p to represent unknown parameters in the state evolution. Although the minimal risk portfolio could be obtained from the MV model for a given expected return level, the expected return level is hard to determine. From the investor's viewpoint, the change of expected return at the current period should depend on her expectation on the return rate and her reflection on the last period's weights. Therefore, a reasonable assumption of the expected return ODE form is:

$$\dot{\rho}(t) = \Theta(t, \rho, w, p) = a(t, p_1)\rho(t) + b(t, p_2)^T w(t) \tag{5.2}$$

where we denote $a(t, p_1) : R^{l_1} \to R$ and $b(t, p_2) : R^{l_2} \to\in R^m$ at time t. l_1 and l_2 are dimensions of the parameters p_1 and p_2. The Θ is an affine function in $(\rho(t), w(t))$. Since $a(t, p_1)$ and $b(t, p_2)$ varies with time, we could assume that $a(t, p_1)$ and $b(t, p_2)$ contain unknown parameters: p_1 in $a(t, p_1)$ and p_2 in $b(t, p_2)$, $p = (p_1, p_2)$.

Coefficient varying with time

We consider the market fluctuation $\dot{\mu}(t)$ to reflect last period's weights on the expected return setting. Moreover, we assume $\dot{\mu}(t)$ is Lipschitz continuous. Then denote $b = \mathrm{b}\dot{\mu}(t)$, $\mathrm{b} \in R$. Moreover, we assume a is linearly varying with time,

$$\Theta(t, \rho, w, p) = (c_0 + c_1 t)\rho(t) + \mathrm{b}\dot{\mu}^T(t)w(t), \tag{5.3}$$

which is $a(t, p_1) = c_0 + c_1 t$, $b(t, p_2) = \mathrm{b}\dot{\mu}(t)$, $p_1 = (c_0, c_1), p_2 = (\mathrm{b})$, $p = (p_1, p_2)$.

From dynamic investment in the training sample, we could evaluate the values of the parameters c_0, c_1, and b. c_0, c_1 reflect the investment tendency and b reflects the feedback to the market fluctuation.

The model with ODE constraint involving MV is to find the optimal parameter p:

$$\begin{array}{ll} \min\limits_{p} & f = \int\limits_0^T (\bar{\rho}(t) - \rho(t))^2 dt \\ \text{s.t.} & \dot{\rho}(t) = \Theta(t, \rho(t), w(t), p) \\ & \rho(0) = \rho^0, t \in [0, T], \end{array} \tag{5.4}$$

where $w(t)$ is the solution to problem (5.1), ρ^0, the initial value, and T, the terminal time.

We denote $\bar{\rho}(t) \in C^1$ as a given target return series. By the objective function, it identifies the parameters in the ODE to make the expected return mostly fit to the target return. In the constraint, the parametric ODE combining with the MV model reflects the investor's viewpoints on the expected return setting.

In this problem, we project the target return to the area where expected return and portfolio weights are defined by the constraint. The MV model boasts the solution, which turns out to be a linear relationship between $w(t)$ and $\rho(t)$. We discuss it in subsection 5.1.2.

5.1.2 MV model

Solution of the MV model

In this subsection, we discuss the solution of MV model (5.1) and gives out an analytical relation between $\rho(t)$ and $w(t)$. The explicit solution to MV model (5.1) is researched in Chapter 3, Evstigneev et al. (2015), and Part III, Markowitz et al. (2000). We start analyzing the problem by introducing several basic assumptions.

Assumption 5.1. *Let $\underline{\lambda}(t)$ be the smallest eigenvalues of $C(t)$ and $\bar{\lambda}(t)$ be the largest eigenvalues of $C(t)$. Suppose*

$$\underline{\lambda} = \min_{t\in[0,T]} \underline{\lambda}(t) > 0, \tag{5.5}$$

and

$$\overline{\lambda} = \max_{t\in[0,T]} \overline{\lambda}(t) < +\infty. \tag{5.6}$$

If Assumption 5.1 is satisfied, then MV model (5.1) is strictly convex and the solution exists and is unique. If Assumption 5.1 holds, then the inverse of the covariance matrix $C(t)^{-1}$ is positive definite.

Assumption 5.2. *$\mu(t)$ and 1_m are linearly independent for any $t \in [0, T]$.*

Proposition 5.1. *When Assumptions 5.1 and 5.2 are satisfied, we obtain that:*

1. *All the following constants,*

$$z_1(t) = 1_m^T C(t)^{-1} 1_m > 0, z_2(t) = \mu(t)^T C(t)^{-1} 1_m,$$

$$z_3(t) = \mu(t)^T C(t)^{-1} \mu(t) > 0, z_4(t) = z_1(t) z_3(t) - z_2(t)^2 > 0$$

are bounded.

2. *The Markowitz portfolio selection problem (5.1) has a unique solution $\bar{w}(t)$ for $\forall t \in [0, T]$ given by*

$$\bar{w}(t) = \bar{y}_1(t) C(t)^{-1} 1_m + \bar{y}_2(t) C(t)^{-1} \mu(t), \tag{5.7}$$

where $\bar{y}_1(t) = \dfrac{z_3(t) - \rho(t) z_2(t)}{z_4(t)}$, $\bar{y}_2(t) = \dfrac{\rho(t) z_1(t) - z_2(t)}{z_4(t)}$.

3. *For any $t \in [0, T]$, the Lagranger multipliers $\bar{y}_1(t)$ and $\bar{y}_2(t)$, and $\bar{w}(t)$ in 2 are bounded.*

Proof:

1. By Assumption 5.1, $C(t)$ is symmetric positive definite, so $C^{-1}(t)$ is symmetric positive definite. Therefore,

$$\langle \mu^T(t), C^{-1}(t) 1_m \rangle = \langle C^{-1}(t) \mu^T(t), 1_m \rangle. \tag{5.8}$$

By Assumption 5.2, $\mu(t) \neq \eta 1_m$ for some number η. Using Cauchy-Schwartz inequality, we could obtain that,

$$|\langle \mu^T(t), C^{-1}(t) 1_m \rangle|^2 < \langle \mu^T(t), C^{-1}(t) \mu(t) \rangle \langle 1_m^T, C^{-1}(t) 1_m \rangle, \tag{5.9}$$

which says $z_4(t) = z_1(t) z_3(t) - z_2(t)^2 > 0$. Therefore, $\gamma / \overline{\lambda}^{2m} < z_4(t) < \gamma / \underline{\lambda}^{2m}$ for some number γ.

Clearly, $1/\overline{\lambda}^m < z_1(t) = \langle 1_m^T, C^{-1}(t)1_m \rangle < 1/\underline{\lambda}^m$. $z_3(t) = \langle \mu^T(t), C^{-1}(t)\mu(t) \rangle >$ 0, because $C^{-1}(t)$ is positive definite. As $\mu(t)$ is bounded, $z_2(t)$, $z_3(t)$ are bounded and $\inf_{t \in [0,T]} z_4(t) = \gamma/\overline{\lambda}^{2m}$.

2. Applying the Lagrangian method, we can find the Lagranger multipliers y_1, y_2 such that,

$$L(w(t), y_1(t), y_2(t)) = 0.5w(t)^T C(t)w(t) - y_1(t)(1_m^T w(t) - 1) - y_2(t)(\mu^T(t)w(t) - \rho(t)). \tag{5.10}$$

The first order optimality condition of the MV model is,

$$\begin{cases} \dfrac{\partial L}{\partial w} = C(t)w(t) - y_1(t)1_m - y_2(t)\mu(t) = 0 \\ \dfrac{\partial L}{\partial y_1} = 1 - 1_m^T w(t) = 0 \\ \dfrac{\partial L}{\partial y_2} = \rho(t) - \mu^T(t)w(t) = 0. \end{cases} \tag{5.11}$$

By Assumption 5.1, we solve $\dfrac{\delta L}{\delta w} = 0$, which yields,

$$\bar{w}(t) = y_1(t)C^{-1}(t)1_m + y_2(t)C^{-1}(t)\mu(t). \tag{5.12}$$

Expanding equation (5.12) by timing both sides $\mu^T(t)$ for a particular $\rho(t)$ and 1_m^T to solve for $\bar{y}_1(t)$, $\bar{y}_2(t)$,

$$\begin{cases} \rho(t) = \bar{y}_1(t)\mu^T(t)C^{-1}(t)1_m + \bar{y}_2(t)\mu^T(t)C^{-1}(t)\mu(t) \\ 1 = \bar{y}_1(t)1_m^T C^{-1}(t)1_m + \bar{y}_2(t)1_m^T C^{-1}(t)\mu(t). \end{cases} \tag{5.13}$$

Using $z_1(t)$, $z_2(t)$, $z_3(t)$, $z_4(t)$, we could rewrite (5.13) as,

$$\begin{pmatrix} z_1(t) & z_2(t) \\ z_2(t) & z_3(t) \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} = \begin{pmatrix} 1 \\ \rho(t) \end{pmatrix}. \tag{5.14}$$

Then, we can obtain,

$$\begin{cases} \bar{w}(t) = \bar{y}_1(t)C(t)^{-1}1_m + \bar{y}_2(t)C(t)^{-1}\mu(t) \\ \bar{y}_1(t) = \dfrac{z_3(t) - \rho(t)z_2(t)}{z_4(t)} \\ \bar{y}_2(t) = \dfrac{\rho(t)z_1(t) - z_2(t)}{z_4(t)}. \end{cases} \tag{5.15}$$

3. When the first and second part of the proposition are proved, the Lagranger multipliers $\bar{y}_1(t)$ and $\bar{y}_2(t)$, $\bar{w}(t)$ from (5.15) are bounded for any $t \in [0, T]$.

This completes the proof. □

By solving (5.15), we obtain,

$$\bar{w}(t) = (\frac{z_3(t)}{z_4(t)}C(t)^{-1}1_m - \frac{z_2(t)}{z_4(t)}C(t)^{-1}\mu(t)) + \rho(t)(\frac{z_1(t)}{z_4(t)}C(t)^{-1}\mu(t) - \frac{z_2(t)}{z_4(t)}C(t)^{-1}1_m). \tag{5.16}$$

Let

$$X(t) = \frac{z_3(t)}{z_4(t)}C(t)^{-1}1_m - \frac{z_2(t)}{z_4(t)}C(t)^{-1}\mu(t), \tag{5.17}$$

and

$$Y(t) = \frac{z_1(t)}{z_4(t)}C(t)^{-1}\mu(t) - \frac{z_2(t)}{z_4(t)}C(t)^{-1}1_m, \tag{5.18}$$

in formula (5.16). We know that the weights at time period t,

$$\bar{w}(t) = X(t) + \rho(t)Y(t). \tag{5.19}$$

It is the solution of the MV model. Since we do not know the time distribution of $r(t)$, we may estimate the sample covariance $C(t)$ and mean $\mu(t)$ by using the sample average approximation (SAA) method. Therefore, $X(t)$ and $Y(t)$ could be computed from the sample data at each time step t.

Two-Fund Theorem

Theorem 5.1. *The solution to the QP:*

$$\begin{array}{ll} \min & \frac{1}{2}w(t)^T C(t)w(t) - \lambda\mu(t)^T w(t) \\ s.t. & w(t)^T 1_m = 1, \end{array} \tag{5.20}$$

with $\lambda > 0$ *can be represented as*

$$\tilde{w}(t) = (1-\alpha)w_{min}(t) + \alpha w_{mk}(t), \tag{5.21}$$

where $w(t)_{min} = C(t)^{-1}e/z_1(t)$ *is the minimum variance portfolio*[1], *and* $w(t)_{mk} = C(t)^{-1}\mu(t)/z_2(t)$ *is the market portfolio*[2] *and* $\alpha = \lambda z_2(t)$.

Using Theorem 5.1, we could allocate the portfolio to be the linear combination of two portfolios. The return associated with the portfolio is,

$$\begin{aligned} \rho(t) =&(1-\alpha)\mu(t)w_{min}(t) + \alpha\mu(t)w_{mk}(t) \\ =&(1-\alpha)z_2(t)/z_1(t) + \alpha z_3(t)/z_2(t). \end{aligned} \tag{5.23}$$

We could set target return as given by Equation (5.23).

5.1.3 ODE constraint involving MV model

Combined with MV model, it is a dynamic system which we take as the constraint:

$$\begin{cases} \dot{\rho}(t) = \Theta(t, \rho(t), w(t), p) \\ \rho(0) = \rho^0, t \in [0, T], \end{cases} \tag{5.24}$$

where $w(t) \in \mathcal{S}(t, \rho(t))$, ρ^0 is the initial value of the expected return ODE. We could estimate the variance $C(t)$ and the mean $\mu(t)$ at time t by sample average approximation (SAA).

[1] It is the solution to the optimization problem:

$$\begin{array}{ll} \min & \frac{1}{2}w(t)^T C(t)w(t) \\ \text{s.t.} & w(t)^T 1_m = 1. \end{array} \tag{5.22}$$

[2] It is labeled as market portfolio as it reflects all the market information on the assets.

5.2 Discretization and solving the problem

5.2.1 The proposed model

In this section, we introduce the numerical method to solve the optimization problem with the ODE constraint involving MV in (5.4). We use the time-stepping method, a numerical scheme to discretize the problem. The idea of time-stepping is to use a finite-difference quotient to replace the time derivative $\dot{\rho}$. We divide the time interval $[0, T]$ into N subintervals, with step size h. Therefore, $hN = T$, $t_i^h = t_{i-1}^h + h$ and we have,

$$0 = t_0^h < t_1^h < \ldots < t_N^h = T. \tag{5.25}$$

Let $\rho_i^h \approx \rho(t_i^h)$. The discrete form of the model (5.4) is,

$$\begin{array}{ll} \min\limits_{p} & f = \sum\limits_{i=1}^{N} (\bar{\rho}_i^h - \rho_i^h)^2 \\ \text{s.t.} & \rho_i^h = \rho_{i-1}^h + h\Theta(t_i^h, \gamma\rho_{i-1}^h + (1-\gamma)\rho_i^h, \gamma w_{i-1}^h + (1-\gamma)w_i^h, p) \\ & \rho_0^h = \rho^0 \\ & i = 1, ..., N. \end{array} \tag{5.26}$$

where $w_i^h = X_i^h + Y_i^h \rho_i^h$, $X_i^h = X(t_i^h)$, $Y_i^h = Y(t_i^h)$, $\rho_i^h = \rho(t_i^h)$, $w_i^h = w(t_i^h)$ and $\gamma \in (0, 1)$.

5.2.2 ODE constraint involving MV model

The constraint in our model is explicitly in this section a MV model with an initial value ODE problem of ρ dynamics.

Given a starting point ρ^0, it computes $w_0^h \in \mathcal{S}(0, \rho_0^h)$, and two finite series containing parameter p,

$$\{\rho_1^h, \rho_2^h, \ldots, \rho_N^h\} \subset R \tag{5.27}$$

and

$$\{w_1^h, w_2^h, \ldots, w_N^h\} \subset R^m \tag{5.28}$$

by recursive iteration: for $i = 1, \ldots, N$,

$$\begin{array}{l} \rho_i^h = \rho_{i-1}^h + h\Theta(t_i^h, \gamma\rho_{i-1}^h + (1-\gamma)\rho_i^h, \gamma w_{i-1}^h + (1-\gamma)w_i^h, p) \\ w_i^h = X_i^h + Y_i^h \rho_i^h, \end{array} \tag{5.29}$$

where h is the step size and γ is a scalar. When $\gamma = 0$, it is an implicit numerical scheme; $\gamma = 1$, an explicit scheme, or $\gamma \in (0,1)$, a semi-implicit scheme. Implicit Euler scheme is usually more stable, see Chen and Wang (2014) and Chen and Wang (2013), therefore, we use $\gamma = 0$ in the Euler scheme (5.29).

The implicit Euler method for the ODE

Using the implicit Euler method, we discretize the ODE of ρ. Denote $\rho_i^h \approx \rho(t_i^h)$ at time step i. Let $a_i^h(p_1) = a(t_i^h, p_1) = c_0 + c_1 t_i^h$ and $b_i^h(p_2) = b(t_i^h, p_2) = \mathrm{b}\triangle\mu_i^h$, $\triangle\mu_i^h = \mu(t_i^h) - \mu(t_{i-1}^h)$ for $i = 1, 2, ..., N$. We have,

$$\frac{\rho_i^h - \rho_{i-1}^h}{h} = a_i^h(p_1)\rho_i^h + (b_i^h(p_2))^T w_i^h, \tag{5.30}$$

which implies

$$\rho_i^h = \rho_{i-1}^h + h(a_i^h(p_1)\rho_i^h + (b_i^h(p_2))^T w_i^h). \tag{5.31}$$

The MV model at each time step has a solution in the form of (5.19), therefore, equation (5.31) could be written as

$$\rho_i^h = \rho_{i-1}^h + h(a_i^h(p_1)\rho_i^h + (b_i^h(p_2))^T (X_i^h + Y_i^h \rho_i^h)), \tag{5.32}$$

which is

$$(1 - a_i^h(p_1)h - (b_i^h(p_2))^T Y_i^h h)\rho_i^h = \rho_{i-1}^h + hb_i^h(p_2)^T X_i^h, \ i = 1, ..., N. \tag{5.33}$$

Let $\alpha_i^h(p_1, p_2) := 1 - a_i^h(p_1)h - (b_i^h(p_2))^T Y_i^h h$, $\gamma_i^h(p_2) = h(b_i^h(p_2))^T X_i^h$. Assume $\alpha_i^h(p_1, p_2) \neq 0$, $\gamma_i^h(p_2) \neq 0$. Equations (5.33) are

$$\alpha_i^h(p_1, p_2)\rho_i^h = \rho_{i-1}^h + \gamma_i^h(p_2), \ i = 1, ..., N, \tag{5.34}$$

or

$$\rho_i^h = \frac{1}{\alpha_i^h(p_1,p_2)}\rho_{i-1}^h + \frac{\gamma_i^h(p_2)}{\alpha_i^h(p_1,p_2)},\ i = 1, ..., N. \tag{5.35}$$

With equation (5.34) as a constraint, problem (5.26) is stated as:

$$\begin{aligned}
\min_{p_1,p_2,\rho^h} \quad & f = \|\rho^h - \bar{\rho}^h\|_2^2 \\
\text{s.t.} \quad & \begin{pmatrix} \alpha_1^h(p_1,p_2) & \cdots & \cdots & 0 \\ -1 & \alpha_2^h(p_1,p_2) & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ & & -1 & \alpha_N^h(p_1,p_2) \end{pmatrix} \begin{pmatrix} \rho_1^h \\ \rho_2^h \\ \vdots \\ \rho_N^h \end{pmatrix} = \begin{pmatrix} \gamma_1^h(p_2) + \rho_0^h \\ \gamma_2^h(p_2) \\ \vdots \\ \gamma_N^h(p_2) \end{pmatrix},
\end{aligned} \tag{5.36}$$

where $p_1 = (c_0, c_1)$, $p_2 = b$, $\rho^h = (\rho_1^h, ..., \rho_N^h)$ and $\bar{\rho}^h = (\bar{\rho}_1^h, ..., \bar{\rho}_N^h)$.

In problem (5.36), we reformulate problem (5.26) into a nonlinear optimization problem with nonlinear constraints and solve it by "fmincon" using quasi-Newton method in Matlab. The optimization arguments are p_1, p_2 and ρ together.

5.3 Solvability

In this section, we first show the solution of problem (5.36) exists, then further prove that the solution is the KKT point. Before we show the existence of the solution, we first deduce the formula of ρ_i^h by the implicit Euler method. With $a_i^h(p_1)$, $b_i^h(p_2)$, $\alpha_i^h(p_1,p_2)$, $\gamma_i^h(p_2)$ defined in the previous section, hereafter we use a_i^h, b_i^h, α_i^h, γ_i^h for simplicity.

Suppose $\Theta = a_i^h \rho_i^h + (b_i^h)^T w_i^h$, for a given initial value ρ^0, the expected return ρ_i^h at time step t_i^h satisfies the system

$$\rho_i^h = \begin{cases} \dfrac{\rho_0^h}{\prod_{j=1}^{i} \alpha_j^h} + \sum_{j=1}^{i} \dfrac{\gamma_j^h}{\prod_{j_1=j}^{i} \alpha_{j_1}^h}, & i = 1, ..., N, \\ \rho^0, & i = 0. \end{cases} \tag{5.37}$$

Proof of Formula (5.37): We show that formula (5.37) could be proved by mathematical induction applying equation (5.33).

Suppose at t_i^h for $i = 2...N-1$, we have Formula (5.37)

$$\rho_i^h = \frac{\rho_0^h}{\prod_{j=1}^{i} \alpha_j^h} + \sum_{j=1}^{i} \frac{\gamma_j^h}{\prod_{j_1=j}^{i} \alpha_{j_1}},$$

then at t_{i+1}^h, according to formula (5.33),

$$\begin{aligned} \rho_{i+1}^h =& \frac{1}{\alpha_{i+1}^h}\rho_i^h + \frac{\gamma_{i+1}^h}{\alpha_{i+1}^h} = \frac{\frac{\rho_0^h}{\prod_{j=1}^{i} \alpha_j^h} + \sum_{j=1}^{i} \frac{\gamma_j^h}{\prod_{j_1=j}^{i} \alpha_{j_1}}}{\alpha_{i+1}^h} + \frac{\gamma_{i+1}^h}{\alpha_{i+1}^h} \\ =& \frac{\rho_0^h}{\prod_{j=1}^{i+1} \alpha_j^h} + \sum_{j=1}^{i} \frac{\gamma_j^h}{\prod_{j_1=j}^{i+1} \alpha_{j_1}} + \frac{\gamma_{i+1}^h}{\alpha_{i+1}^h} \\ =& \frac{\rho_0^h}{\prod_{j=1}^{i+1} \alpha_j^h} + \sum_{j=1}^{i+1} \frac{\gamma_j^h}{\prod_{j_1=j}^{i+1} \alpha_{j_1}}. \end{aligned} \tag{5.38}$$

□

Assumption 5.3. *Assume that $\dot{\mu}(t)$ is bounded for any $t \in [0, T]$.*

Theorem 5.2. *If the parameters c_0, c_1, b are box constrained and h small enough, then problem (5.36) has a solution.*

Proof: Firstly, we show that the optimization problem is solvable. The solution set is non-empty. By Proposition 5.1, we could observe that the solution to the MV model is non-empty for each given $\rho(t)$.

Secondly, we show the feasible set is bounded. Because of the randomness of $\triangle\mu_j^h$, X_j^h and Y_j^h, suppose that

$$\bar{X} = \max\{\triangle\mu_j^h X_j^h\}_{j=1}^N, \tag{5.39}$$

$$\bar{Y} = \max\{\triangle\mu_j^h Y_j^h\}_{j=1}^N. \tag{5.40}$$

We show that the ρ_i^h for $i = 1, ..., N$ is bounded by the implicit Euler scheme,

$$\rho_i^h = \frac{\rho_0^h}{\prod_{j=1}^{i}(1 - a_j^h h - {b_j^h}^T Y_{j,h} h)} + \sum_{j=1}^{i} \frac{{b_j^h}^T X_j^h h}{\prod_{j_1=j}^{i}(1 - a_{j_1}^h h - {b_{j_1}^h}^T Y_{j_1}^h h)}. \tag{5.41}$$

We take h small enough such that $1 - a_j^h h - \mathrm{b}\triangle\mu_j^h Y_j^h h > 0$ for all $j = 1, ..., N$. When write $b_j^h = \mathrm{b}\triangle\mu_j^h$, use (5.39) and (5.40), write $a_j^h = c_0 + c_1 t_j^h$, and $\rho_0^h > 0$, we have

$$|\rho_i^h| = \left|\frac{\rho_0^h}{\prod_{j=1}^{i}(1 - a_j^h h - \mathrm{b}\triangle\mu_j^h Y_j^h h)} + \sum_{j=1}^{i} \frac{\mathrm{b}\triangle\mu_j^h X_j^h h}{\prod_{j_1=j}^{i}(1 - a_{j_1}^h h - \mathrm{b}\triangle\mu_j^h Y_{j_1}^h h)}\right| \tag{5.42}$$

$$\leqslant \frac{\rho_0^h}{\prod_{j=1}^{i}(1 - a_j^h h - \mathrm{b}\bar{Y} h)} + \sum_{j=1}^{i} \frac{|\mathrm{b}\bar{X} h|}{\prod_{j_1=j}^{i}(1 - a_{j_1}^h h - \mathrm{b}\bar{Y} h)}$$

$$= \frac{\rho_0^h}{\prod_{j=1}^{i}(1 - (c_0 + c_1 t_j^h) h - \mathrm{b}\bar{Y} h)} + \sum_{j=1}^{i} \frac{|\mathrm{b}\bar{X} h|}{\prod_{j_1=j}^{i}(1 - (c_0 + c_1 t_{j_1}^h) h - \mathrm{b}\bar{Y} h)}.$$

Use the fact that $t_i^h \leqslant T$ for $i = 1, ..., N$, the summation formula of geometric progression, and reformulate to use exponential equation, we have

$$|\rho_i^h| \leqslant \frac{\rho_0^h}{(1 - (c_0 + c_1 T) h - \mathrm{b}\bar{Y} h)^i} + \sum_{j=1}^{i} \frac{|\mathrm{b}\bar{X} h|}{\prod_{j_1=j}^{i}(1 - (c_0 + c_1 T) h - \mathrm{b}\bar{Y} h)} \tag{5.43}$$

$$= \frac{\rho_0^h}{(1 - (c_0 + c_1 T + \mathrm{b}\bar{Y}) h)^i} + \frac{|\mathrm{b}\bar{X} h|(1 - \frac{1}{(1 - (c_0 + c_1 T) h - \mathrm{b}\bar{Y} h)^i})}{1 - \frac{1}{(1 - (c_0 + c_1 T) h - \mathrm{b}\bar{Y} h)}}. \tag{5.44}$$

Let $\eta = 1 - (c_0 + c_1 T) h - \mathrm{b}\bar{Y} h$, $\tau = c_0 + c_1 T + \mathrm{b}\bar{Y}$, then $\eta = 1 - \tau h$. Since $h = T/N$, it is easy to see that $0 < \eta \leqslant 1 + |\tau| T$. For an arbitrary i, N sufficiently large, let $i = \theta N$, $\theta \in [0, 1]$, then

$$\eta^i = (1 - \tau h)^i = (1 - \tau T/N)^{N\theta}. \tag{5.45}$$

The limitation of equation (5.45), $\lim_{N\to\infty}(1 - \tau T/N)^{N\theta} = e^{-\tau T\theta}$. If $\tau > 0$, then $\eta < 1$, $e^{-\tau T} < \eta^N < \eta^i < 1$; if $\tau < 0$, then $\eta > 1$, $1 < \eta^i < \eta^N < e^{-\tau T}$. Then we have: $\eta^i \in [\min(1, e^{-\tau T}) - \delta, \max(1, e^{-\tau T}) + \delta]$ for a small number $\delta \in R$. Hence η^i is bounded. The righthand side of (5.44) could be written as

$$\begin{aligned}\frac{\rho_0^h}{\eta^i} + \frac{|\mathrm{b}\bar{X}h|(1 - \frac{1}{\eta^i})}{1 - \frac{1}{\eta}} =& \frac{\rho_0^h}{\eta^i} + \frac{|\mathrm{b}\bar{X}h|\eta(\eta^i - 1)}{\eta^i(\eta - 1)} \\ \leqslant \frac{\rho_0^h}{\eta^i} + \frac{|\mathrm{b}\bar{X}|T\eta(\eta^i - 1)}{\eta^i(\eta - 1)}.\end{aligned} \tag{5.46}$$

If we denote the righthand side term of (5.46) as l, then for sure $\rho_i^h \in [-2l, 2l]$ for $i = 1, ..., N$. c_0, c_1 and b are also bounded.

The objective function is continuous. The feasible set is non-empty and bounded. It is directly obtained that problem (5.36) has a solution.□

We then show that the local minimizer is the KKT point. Denote the constraint of problem (5.36) as $g(\rho^h, c_0, c_1, b) = 0$.

Definition 5.1. *Let $\mathcal{F}$ be the feasible region of problem (5.36). We say that a point $(\rho^{h^*}, c_0^*, c_1^*, b^*) \in R^{N+3}$ is a KKT point of problem (5.36) provided that $(\rho^{h^*}, c_0^*, c_1^*, b^*) \in \mathcal{F}$ and there exists $\lambda \in R^{N+3}$ such that*

$$0 \in \nabla f(\rho^{h^*}, c_0^*, c_1^*, b^*) + \nabla g(\rho^{h^*}, c_0^*, c_1^*, b^*)\lambda.$$

Definition 5.2 (Andreani et al. (2012), Definition 4)**.** *Let $(\rho^{h^*}, c_0^*, c_1^*, b^*) \in \mathcal{F}$ and let $\mathcal{I} \subseteq \{1, ..., N\}$ be such that $\{\nabla g_i(\rho^{h^*}, c_0^*, c_1^*, b^*)|i \in \mathcal{I}\}$ is a basis for s-pan $\{\nabla g_i(\rho^{h^*}, c_0^*, c_1^*, b^*)|i = 1, ..., N\}$. We say that relaxed constant positive linear dependence (RCPLD) holds for the system $g(\rho, c_0, c_1, b) = 0$ at $(\rho^{h^*}, c_0^*, c_1^*, b^*)$*

if there exists $\delta > 0$ *such that* $\{\nabla g_i(x)|i = 1, ..., N\}$ *has the same rank for each* $(\rho^{h^*}, c_0^*, c_1^*, b^*) \in \mathcal{B}_\delta(\rho^{h^*}, c_0^*, c_1^*, b^*)$.

Theorem 5.3. *Suppose the objective function is bounded. Let* $(\rho^{h^*}, c_0^*, c_1^*, b^*)$ *be a local minimizer of problem (5.36), if the RCPLD holds at* $(\rho^{h^*}, c_0^*, c_1^*, b^*)$*, then* $(\rho^{h^*}, c_0^*, c_1^*, b^*)$ *is a KKT point of problem (5.36).*

Proof: It is easy to see that there exists $\delta > 0$ such that $\text{Rank}(\nabla g_i(\rho^h, c_0, c_1, b)) = 4$ for $i = 1, ..., N$. Therefore $\{\nabla g_i(\rho^h, c_0, c_1, b)|i = 1, ..., N\}$ has the same rank for each $x \in \mathcal{B}_\delta(\rho^{h^*}, c_0^*, c_1^*, b^*)$. Thus, the RCPLD condition by Andreani et al. (2012) is satisfied. Moreover, the objective function is bounded. By Theorem 2.1 in Chen et al. (2017), Theorem 7.1.10 in Sun et al. (2004), we could obtain the conclusion.

5.4 Applications

In this section, we demonstrate the model by an empirical example. The Matlab version that we use for the numerical example is Matlab R2014a, in a computer with Intel Core 2 Due CPU E8500 3.16GHz. The data sets that we used are DJIA index and component stocks. The DJIA stocks and index data are daily stock prices downloaded from Yahoo Finance.

The parameter identification process to find the parameters and the investment strategy in each portfolio re-balancing (investment round) is illustrated in Table 5.1. The portfolio weight w_0^h is from the Markowitz MV model given the initial value of expected return ρ^0. Then at the first time step, the expected return depends on the expected return ρ_0^h and the portfolio allocation w_0^h at time step 0. The following $\rho_i^h(p)$ and $w_i^h(p)$, $i = 1, \ldots, N$ are evaluated by the process illustrated in (5.29). We use dynamic investment in every N time steps as the training data to find the parameters. Then put the parameters into the ρ ODE, we obtain the expected return to construct the portfolio and make the out-of-sample investment subsequently after parameter identification. To compute the covariance and expected returns at each time step, we use 40 daily returns, i.e. nearly two trading months as sampling data, at each time step.

time step	0	1	2	...	N	out-of-sample invest
target $\bar{\rho}_i^h$	$\bar{\rho}_0^h$	$\bar{\rho}_1^h$	$\bar{\rho}_2^h$	...	$\bar{\rho}_N^h$	
expected ρ_i^h	$\rho_0^h = \rho^0$	$\rho_1^h(p)$	$\rho_2^h(p)$	...	$\rho_N^h(p)$	ρ_N^h
$\bar{w}_i^h = \mathcal{S}(\bar{\rho}_i^h)$	$\mathcal{S}(\bar{\rho}_0^h)$	$\mathcal{S}(\bar{\rho}_1^h)$	$\mathcal{S}(\bar{\rho}_2^h)$	...	$\mathcal{S}(\bar{\rho}_N^h)$	$\bar{w}_N = \mathcal{S}(\bar{\rho}_N^h)$
$w_i^h = \mathcal{S}(\rho_i^h)$	$\mathcal{S}(\rho_0^h)$	$\mathcal{S}(\rho_1^h(p))$	$\mathcal{S}(\rho_2^h(p))$	...	$\mathcal{S}(\rho_N^h(p))$	$w_N^h = \mathcal{S}(\rho_N^h)$; $w_{N+1,\widetilde{os}} = \mathcal{S}(\rho_{N+1}^h, \widetilde{os})$
Objective	$p = \arg\min \sum_{i=1}^{N} (\bar{\rho}_i^h - \rho_i^h)^2$					

Table 5.1: The parameter identification and investment process in a round of dynamic investment

Example 5.1. *DJIA data set. The Dow Jones Industrial Average Index is a representative index in the American capital market to measure the overall market performance. The components of the index are 30 blue chip stocks.*

One target return is the DJIA index. For each step's portfolio return setting, we simply compute the corresponding average DJIA index price return from the index data. We call the portfolio formed by this setting index tracking portfolio.

The other target return is based on the two-fund theorem (call it two-fund tracking strategy), see Chapter 6, Luenberger (1997). We take $\alpha = 0.5$ in (5.23) for this example. We call the portfolio formed by this setting the two-fund tracking portfolio.

Data Collection: We collected the daily price and computed the returns of the 30 stocks of the DJIA composition. The 30 stocks are Apple, American Express, Boeing, Caterpillar, Cisco Systems, Chevron, E.I.du Pont de Nemours, Walt Disney, General Electric, Goldman Sachs, Home Depot, IBM, Intel, Johnson & Johnson, JPMorgan, Coca-Cola, McDonald, 3M, Merck, Microsoft, NIKE, Pfizer, Procter & Gamble, Travelers, UnitedHealth Group, United Technologies, Visa, Verizon, Wal-Mart and Exxon Mobile, which covers a wide range of industries. The sampling period is from March 18, 2008 to December 31, 2015, which covers 1962 trading days. We construct the portfolio selected from the DJIA component stocks by tracking the DJIA index returns or tracking the target return of two funds.

For the empirical example: the historical data of the first 40×36 transaction days from March 20, 2008 to March 3, 2009 (the number of steps in one round of dynamic investment is $N = 36$; and each step or period is 40 days) are used to build the initial portfolio. The investor makes the first re-balance after 40×36 ($N = 36$) transaction days. The portfolio is re-balanced for every 40 transaction days. The update of stock weights for every 40 transaction days depends on the 40×36 rolling historical data. The example assumes an initial investment of unit 1. From March 18, 2008 to December 31, 2015, re-balancing takes place on 10 occasions. We show the rebalancing performance over time horizon in the empirical example. All the test information is listed in Table 5.2.

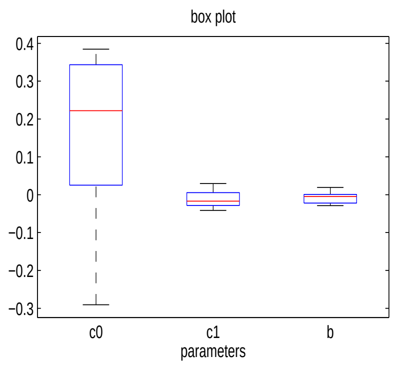
continuous function		discrete scheme	
$\bar{\rho}(t)$:	$\bar{\rho}(t) = r_{DJIA}(t)$	data set:	DJIA daily returns
$\bar{w}(t)$:	$\mathcal{S}(\bar{\rho}(t))$	sampling period:	3/18/2008-12/31/2015
Θ:	$a\rho + b^T w$	the duration of 1 time step:	40 days
τ:	1	N steps to build ODE:	36
		initial value of unknowns:	0_{3+N}
		ρ^0	0.05%

Table 5.2: The given information of the empirical example of DJIA data set

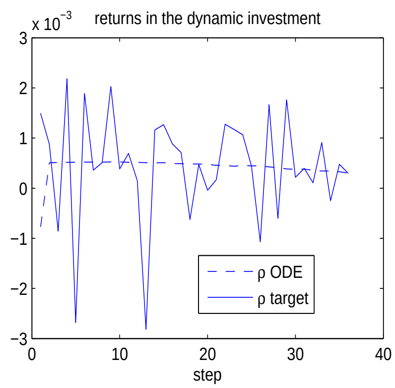
The identified parameters are put in Table 5.3. Due to the limitation of space, we show the mean and standard deviation of the identified parameters. We could obtain the sequence of ρ_i^h for $i = 1, ..., N$ after the parameters are learnt.

DJIA dataset, over 10 rebalances			
index tracking			
parameter	c_0	c_1	b
mean	0.1642	-0.0130	-0.0073
std	0.2130	0.0216	0.0148
two fund tracking			
parameter	c_0	c_1	b
mean	1.3224	-0.1021	0.0668
std	0.9832	0.0606	0.2447

Table 5.3: The identified parameters

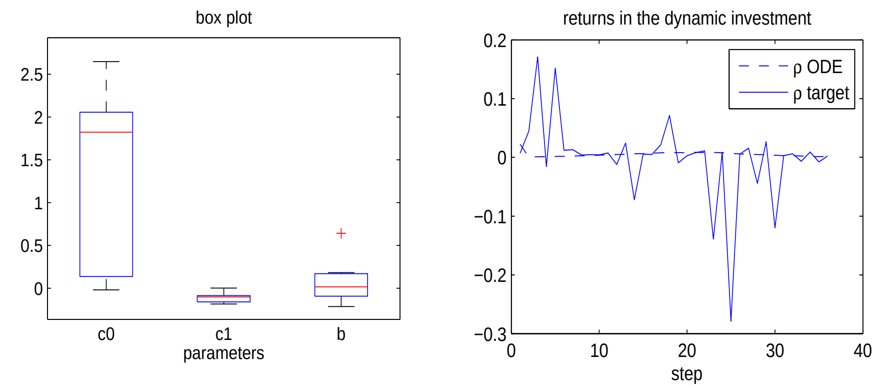


(d)



(e)

Figure 5.1: DJIA index tracking portfolio (a)box plot of parameters; (b)ρ, $\bar{\rho}$ over steps in the last rebalance



(a) (b)

Figure 5.2: DJIA two-fund tracking portfolio (a)box plot of parameters; (b)ρ, $\bar{\rho}$ over steps in the last rebalance

We compare the out-of-sample performance of several portfolios. Namely, the portfolio allocation w_N^h constructed using the final step data by setting expected return as ρ_N^h; the portfolio $\bar{w}_N^h$ constructed at the final step by setting expected return as $\bar{\rho}_N^h$; the portfolio $w_{N+1,\widetilde{os}}$ constructed using out-of-sample data denoted as os in step $N+1$ by setting the expected return as ρ_{N+1}^h; the SPDR Dow Jones Industrial Average ETF Trust (DIA)[3] performance $r_{N+1,DIA}$ in the corresponding $N+1$ step period. We display the first two as a pair as they are kind of static portfolio with portfolio allocation determined by the in-sample data. The last two portfolios are in a pair. We compare the performance of the four portfolios.

5.4.1 Discussion of the results

We show the portfolio return and Sharpe ratio (SR)[4] of the index tracking portfolio and two fund tracking portfolio over 10 re-balancing occasions from March 3, 2009 to December 30, 2015 respectively in Table 5.4 and Table 5.5. When parameters are determined for a round of dynamic investment, we construct out-of-sample portfolios

[3] This fund is Exchange Traded Fund (ETF). Its investment results before expenses corresponds to the proce and yield performance of the Dow Jones Industrial Average. The ticker symbol is DIA.

[4] Sharpe ratio is the ratio of $\frac{r_p}{\sigma_p}$.r_p is portfolio return; σ_p is portfolio standard deviation. It measures the trade-off between returns and volatility of the portfolios.

by setting different returns. "r_{fs}" stands for portfolio return using ρ_N^h at the final step. "$r_{targetfs}$" stands for portfolio return using the target $\bar{\rho}_N^h$ at the final step. "r_{os}" stands for portfolio return where allocation determined by the return ρ_{N+1}^h and the out-of-sample sample. "r_{DIA}" stands for DIA portfolio return. Over 10 occasions, the average return and Sharpe ratio of the $w_{N+1,\tilde{os}}$ portfolio is better than the other portfolios, which means the return-risk tradeoff is well balanced.

For index tracking portfolio, Figure 5.1(d) is the box plot of the identified parameters c_0, c_1 and b. It shows the distribution of the identified parameters over 10 rounds of dynamic investments. We could observe that all the identified parameters are included in the box. There are no scatters. Figure 5.1(e) is ρ_i^h when known parameters substituted into the ODE; and $\bar{\rho}_i^h$ computed from index data set over steps for $i = 1, ..., N$ in the last round of investment. For the two fund tracking portfolio, we illustrate the parameters in box plot in Figure 5.2(a), the expected return setting over steps in the last round of investment in Figure 5.2(b).

After 10 re-balancing which lasts for almost 92 months, the portfolio value $w_{N+1,\widetilde{os}}$ is highest among the portfolios in both cases of expected return setting strategy. In Figure 5.3 and Figure 5.4, we show the portfolio values tracking index or tracking two funds over the time. The parameters learnt in the dynamic investment in the training period well tune the investors' expectation on the target return, due to the reason that the model captures investor's expectation change based on its last period return and the market movement in between two steps. During the overall investment period from 2008 to 2015, it is bull market most of the time. We could learn from the results of the empirical example that the our strategy is even better than the passive strategy of DIA portfolio.

$\bar{\rho}$ set as the index return

round	$r_{w_{N,h}}$	$r_{\bar{w}_{N,h}}$	$r_{w_{N+1,\widetilde{os}}}$	r_{DIV}	$SR_{w_{N,h}}$	$SR_{\bar{w}_{N,h}}$	$SR_{w_{N+1,\widetilde{os}}}$	SR_{DIV}
1	0.02%	-0.02%	-0.04%	-0.06%	0.0131	-0.0135	-0.1384	-0.0709
2	0.04%	0.00%	0.15%	0.19%	0.0326	0.0021	0.1246	0.3129
3	0.36%	0.34%	0.00%	0.03%	0.2832	0.2743	-0.001	0.0506
4	-0.07%	-0.07%	-0.24%	0.04%	-0.0604	-0.0637	-0.1637	0.1071
5	0.01%	0.00%	0.09%	0.02%	0.0063	0.0037	0.0919	0.0415
6	0.04%	0.04%	0.45%	0.10%	0.0229	0.024	0.2379	0.108
7	-0.11%	-0.15%	0.39%	-0.02%	-0.0651	-0.0898	0.3121	-0.0213
8	-0.03%	-0.03%	0.44%	0.06%	-0.0174	-0.017	0.1934	0.0686
9	0.23%	0.23%	-0.15%	0.04%	0.1725	0.1761	-0.0862	0.0513
10	0.05%	0.05%	0.18%	-0.03%	0.0431	0.0432	0.1389	-0.0488
average	0.05%	0.04%	0.13%	0.13%	0.0431	0.0339	0.071	0.0599

Table 5.4: index tracking portfolio 10 times rebalancing performance

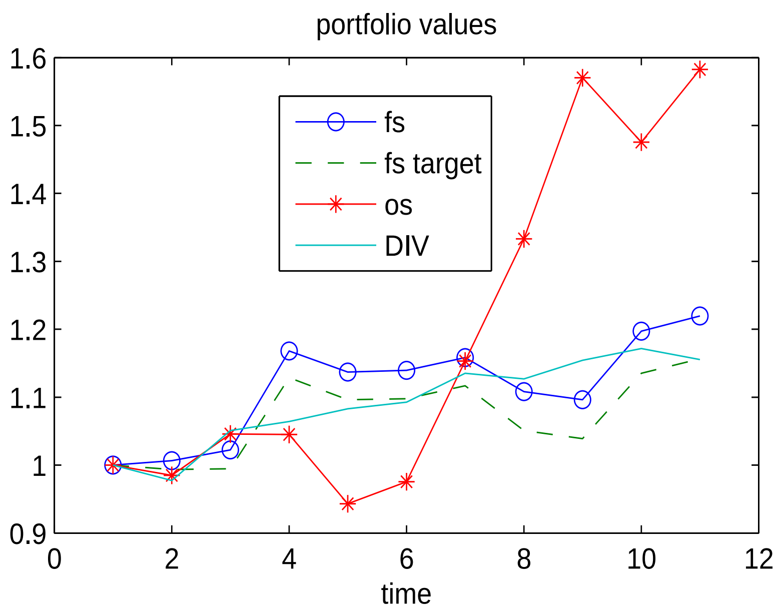


Figure 5.3: DJIA index tracking portfolio value over 10 rounds of investment

$\bar{\rho}$ constructed from targeting the two-fund portfolio return

round	$r_{w_{N,h}}$	$r_{\bar{w}_{N,h}}$	$r_{w_{N+1,\widetilde{os}}}$	r_{DIV}	$SR_{w_{N,h}}$	$SR_{\bar{w}_{N,h}}$	$SR_{w_{N+1,\widetilde{os}}}$	SR_{DIV}
1	-0.02%	-0.07%	0.16%	-0.06%	-0.0207	-0.06	0.5355	-0.0709
2	0.05%	0.15%	0.15%	0.19%	0.0369	0.1286	0.1245	0.3129
3	0.36%	0.32%	-0.07%	0.03%	0.2833	0.2567	-0.0442	0.0506
4	-0.30%	0.00%	-0.26%	0.04%	-0.1014	-0.0016	-0.1811	0.1071
5	-0.03%	-0.07%	0.10%	0.02%	-0.0321	-0.0626	0.09	0.0415
6	0.00%	-0.07%	0.44%	0.10%	0.0022	-0.0541	0.1754	0.108
7	-0.17%	-0.27%	0.54%	-0.02%	-0.1006	-0.158	0.3322	-0.0213
8	-0.03%	-0.02%	0.50%	0.06%	-0.0172	-0.015	0.1725	0.0686
9	0.16%	0.25%	-0.15%	0.04%	0.0369	0.174	-0.0866	0.0513
10	0.04%	0.03%	0.28%	-0.03%	0.0369	0.0308	0.1872	-0.0488
average	0.01%	0.02%	0.17%	0.04%	0.0124	0.0239	0.1305	0.0599

Table 5.5: two-fund tracking portfolio 10 times rebalancing performance

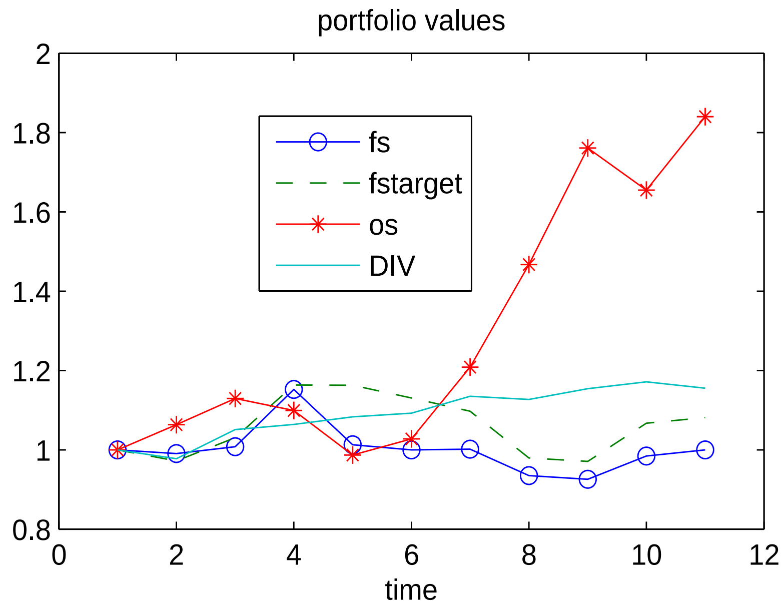


Figure 5.4: DJIA two-fund tracking portfolio value over 10 rounds of investment

Chapter 6

Conclusions and Future work

This chapter summarizes the thesis, and lists the further research directions related to the current work done in this thesis.

6.1 Conclusions

The focus of the thesis has been placed on portfolio optimization. Concretely, the following research problems have been investigated in this thesis.

1. We study a two-phase stochastic linear complementarity approach to seek for a least-0-norm sparse asset allocation of the Markowitz portfolio. In contrast to ℓ_1 penalty regularized portfolio, ℓ_p penalty regularized portfolio, cardinality constrained models and $1/N$ investment strategy, the two-phase approach finds the sparse Markowitz portfolio selection with efficient investment in accordance with the Markowitz minimum variance portfolio structure and preserves the stability of the model. The convergence analysis showed that the SAA method is effective with this two-phase portfolio optimization approach. The application demonstrated the sparsity, and the superior performance of our approach from the perspective of Sharpe ratio, STD, VaR, and CVaR.

2. We study a stochastic penalty method to seek for a least-p-norm sparse asset

allocation of the Markowitz portfolio. The method finds the sparse Markowitz portfolio selection in accordance with the Markowitz minimum variance portfolio structure and preserves the stability of the model. The method we propose is a natural extension of finding sparse solution in the optimal solution set of MV model in Wang and Sun (2017). We have described a NPG algorithm to compute the optimal solutions and implemented it using data sets of 20 A&H stocks, FF12 and FF25 empirically. In contrast to sparse solution of the MV model, ℓ_1 penalty regularized portfolio, cardinality constrained models and $1/N$ investment strategy, the application demonstrated the sparsity, and the superior performance of our approach from the perspective of Sharpe ratio. The convergence analysis showed that SAA method is effective with this least-p-norm with $p \in (0, 1)$ sparse portfolio selection model using penalty method. A randomly generated example illustrates the convergence for p taking several different values.

3. We have used optimization methods for parameter identification in the dynamic portfolio selection. In this case, it is to fit target return with an expected return following a parametric dynamic differential equation with a QP problem. We have described the time-stepping scheme and described the investment procedure. In this framework, we have conducted the re-balancing when parameters in ρ ODE are determined. An empirical example using DJIA data set was given to demonstrate the parametric dynamic portfolio model in useful scenarios such as index tracking strategy and fund tracking.

6.2 Future Work

Related topics for the future research work are listed below.

1. For sparse portfolio studied in Chapter 3 and Chapter 4, we consider the models

when short-selling is allowed. Although the results show our methods bring in sparsity thus transaction cost reduction, it is worthwhile to further explore the portfolio construction when short-selling is not allowed.

2. Although we have studied parameters identification in a dynamic portfolio model with expected return evolution in Chapter 5, due to the complexity, we project the target return into the ODE constraint involving the MV model. We could further consider to project both the target return and target asset allocation into the ODE constraint.

Bibliography

Ammann, M. (2003), "Return guarantees and portfolio allocation of pension funds," *Financial Markets and Portfolio Management*, 17, 277–283.

Andreani, R., Haeser, G., Schuverdt, M. L., and Silva, P. J. (2012), "A relaxed constant positive linear dependence constraint qualification and applications," *Mathematical Programming*, pp. 1–19.

Bertsimas, D. and Shioda, R. (2009), "Algorithm for cardinality-constrained quadratic optimization," *Computational Optimization and Applications*, 43, 1–22.

Bian, W. and Chen, X. (2014), "Neural network for nonsmooth, nonconvex constrained minimization via smooth approximation," *IEEE Transactions on Neural Networks and Learning Systems*, 25, 545–556.

Bonami, P. and Lejeune, M. A. (2009), "An exact solution approach for portfolio optimization problems under stochastic and integer constraints," *Operations Research*, 57, 650–670.

Bonnans, J. F. and Shapiro, A. (2000), *Perturbation Analysis of Optimization Problems*, Springer, New York.

Borwein, J. M. and Zhu, Q. J. (1999), "A survey of subdifferential calculus with applications," *Nonlinear Analysis: Theory, Methods & Applications*, 38, 687–773.

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., and Loris, I. (2009), "Sparse and stable Markowitz portfolios," *Proceedings of the National Academy of Sciences*, 106, 12267–12272.

Candes, E. J. and Tao, T. (2005), "Decoding by linear programming," *IEEE Transactions on Information Theory*, 51, 4203–4215.

Cesarone, F., Scozzari, A., and Tardella, F. (2013), "A new method for mean-variance portfolio optimization with cardinality constraints," *Annals of Operations Research*, 205, 213–234.

Chen, C., Li, X., Tolman, C., Wang, S., and Ye, Y. (2013), "Sparse portfolio selection via quasi-norm regularization," *arXiv preprint arXiv:1312.6350.*

Chen, X. (2012), "Smoothing methods for nonsmooth, nonconvex minimization," *Mathematical Programming*, pp. 1–29.

Chen, X. and Wang, Z. (2013), "Convergence of regularized time-stepping methods for differential variational inequalities," *SIAM Journal on Optimization*, 23, 1647–1671.

Chen, X. and Wang, Z. (2014), "Differential variational inequality approach to dynamic games with shared constraints," *Mathematical Programming*, 146, 379–408.

Chen, X. and Xiang, S. (2016), "Sparse solutions of linear complementarity problems," *Mathematical Programming*, 159, 539–556.

Chen, X., Lu, Z., and Pong, T. K. (2016), "Penalty methods for a class of non-Lipschitz optimization problems," *SIAM Journal on Optimization*, 26, 1465–1492.

Chen, X., Guo, L., Lu, Z., and Ye, J. J. (2017), "An augmented Lagrangian method for non-Lipschitz nonconvex programming," *SIAM Journal on Numerical Analysis*, 55, 168–193.

Cottle, R. W., Pang, J.-S., and Stone, R. E. (1992), *The Linear Complementarity Problem*, Academic Press, Boston, MA.

Cowell, F. (2013), *Risk-Based Investment Management in Practice*, Palgrave Macmillan UK.

Cui, X., Gao, J., Li, X., and Li, D. (2014), "Optimal multi-period mean–variance policy under no-shorting constraint," *European Journal of Operational Research*, 234, 459–468.

DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009a), "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms," *Management Science*, 55, 798–812.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009b), "Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy?" *Review of Financial Studies*, 22, 1915–1953.

Deng, G.-F., Lin, W.-T., and Lo, C.-C. (2012), "Markowitz-based portfolio selection with cardinality constraints using improved particle swarm optimization," *Expert Systems with Applications*, 39, 4558–4566.

Diamond, D. W. and Verrecchia, R. E. (1987), "Constraints on short-selling and asset price adjustment to private information," *Journal of Financial Economics*, 18, 277–311.

Evstigneev, I., Hens, T., and Schenk-Hoppé, K. R. (2015), *Mathematical Financial Economics: A Basic Introduction*, Springer.

Gao, J. and Li, D. (2013), "Optimal cardinality constrained portfolio selection," *Operations Research*, 61, 745–761.

Ge, D., Jiang, X., and Ye, Y. (2011), "A note on the complexity of L p minimization," *Mathematical Programming*, 129, 285–299.

Holthausen, D. M. (1981), "A risk-return model with risk and return measured as deviations from a target return," *The American Economic Review*, 71, 182–188.

Horneff, W., Maurer, R., and Rogalla, R. (2010), "Dynamic portfolio choice with deferred annuities," *Journal of Banking & Finance*, 34, 2652–2664.

Li, D. and Ng, W.-L. (2000), "Optimal dynamic portfolio selection: multiperiod mean-variance formulation," *Mathematical Finance*, 10, 387–406.

Li, X., Zhou, X. Y., and Lim, A. E. (2002), "Dynamic mean-variance portfolio selection with no-shorting constraints," *SIAM Journal on Control and Optimization*, 40, 1540–1555.

Liu, Q., Dang, C., and Huang, T. (2013), "A one-layer recurrent neural network for real-time portfolio optimization with probability criterion," *IEEE Transactions on Cybernetics*, 43, 14–23.

Luenberger, D. G. (1997), *Investment Science*, Oxford University Press.

Luo, X.-D. and Luo, Z.-Q. (1994), "Extension of Hoffman's error bound to polynomial systems," *SIAM Journal on Optimization*, 4, 383–392.

Markowitz, H. (1952), "Portfolio selection," *The Journal of Finance*, 7, 77–91.

Markowitz, H. M., Todd, G. P., and Sharpe, W. F. (2000), *Mean-Variance Analysis in Portfolio Choice and Capital Markets*, John Wiley & Sons.

McNeil, A. J., Frey, R., and Embrechts, P. (2015), *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press.

Merton, R. C. (1980), "On estimating the expected return on the market: An exploratory investigation," *Journal of Financial Economics*, 8, 323–361.

Michaud, R. O. and Michaud, R. O. (2008), *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation Includes CD*, Oxford University Press.

Natarajan, B. K. (1995), "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, 24, 227–234.

Pang, J.-S. and Stewart, D. E. (2008), "Differential variational inequalities," *Mathematical Programming*, 113, 345–424.

Qi, H. and Sun, D. (2006), "A quadratically convergent Newton method for computing the nearest correlation matrix," *SIAM Journal on Matrix Analysis and Applications*, 28, 360–385.

Rockafellar, R. T. and Uryasev, S. (2000), "Optimization of conditional value-at-risk," *Journal of Risk*, 2, 21–42.

Sharpe, W. F. (1994), "The sharpe ratio," *The Journal of Portfolio Management*, 21, 49–58.

Shleifer, A. and Vishny, R. W. (1997), "The limits of arbitrage," *The Journal of Finance*, 52, 35–55.

Sun, W., Xu, C., and Zhu, D. (2004), *Optimization Method*, Higher Education Press.

Tian, Y., Fang, S., Deng, Z., and Jin, Q. (2016), "Cardinality constrained portfolio selection problem: a completely positive programming approach," *Journal of Industrial and Management Optimization*, 12, 1041–1056.

Wang, Q. and Sun, H. (2017), "Sparse portfolio selection via linear complementarity approach," *Journal of Industrial and Management Optimization*.

Xu, F., Wang, G., and Gao, Y. (2014), "Nonconvex $\ell_{1/2}$ regularization for sparse portfolio selection," *Pacific Journal of Optimization*, 10, 163–176.

Xu, F., Lu, Z., and Xu, Z. (2016), "An efficient optimization approach for a cardinality-constrained index tracking problem," *Optimization Methods and Software*, 31, 258–271.

Xu, H. and Zhang, D. (2012), "Monte Carlo methods for mean-risk optimization and portfolio selection," *Computational Management Science*, 9, 3–29.

Xue, L., Ma, S., and Zou, H. (2012), "Positive-definite ℓ_1-penalized estimation of large covariance matrices," *Journal of the American Statistical Association*, 107, 1480–1491.

Ye, J. (2000), "Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints," *SIAM Journal on Optimization*, 10, 943–962.

Zheng, X., Sun, X., and Li, D. (2014), "Improving the performance of MIQP solvers for quadratic programs with cardinality and minimum threshold constraints: A semidefinite program approach," *INFORMS Journal on Computing*, 26, 690–703.

Zhou, X. Y. and Li, D. (2000), "Continuous-time mean-variance portfolio selection: A stochastic LQ framework," *Applied Mathematics & Optimization*, 42, 19–33.