

Continuous-Time and Distributionally Robust Mean-Variance Models

Lin Chen

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2020

© 2020

Lin Chen

All Rights Reserved

Abstract

Continuous-Time and Distributionally Robust Mean-Variance Models

Lin Chen

This thesis contains three works in both continuous-time and distributionally robust mean-variance Markowitz models. In the first work, we study naive strategies in the continuous-time mean-variance model. We propose a new type of agent to approximate the dynamic of the naive agent by partitioning the time line into numerous small equal length time intervals. Then, we prove that, the wealth process of the proposed agent converges to that of the naive agent and derive the explicit formula for the limiting wealth process and its corresponding portfolio process. In the end, we compare the naive strategies with two equilibrium strategies in the Black-Scholes market.

The second work contributes to the mean-variance model by considering its distributionally robust counterpart, where the region of distributional uncertainty is around the empirical measure and the discrepancy between probability measures is dictated by the Wasserstein distance. We reduce this problem to an empirical variance minimization problem with an additional regularization term. Moreover, we extend the recently developed inference methodology to our setting in order to select the size of the distributional uncertainty as well as the associated robust target return rate in a data-driven way. Finally, we report extensive backtesting results on the S&P 500 that compares the performance of our model with those of several well-known models, including the Fama–French model and the

Black–Litterman model.

In the last part, we develop a distributionally robust model based on the Sharpe ratio optimization problem. We transform the problem into an equivalent convex optimization problem that can be solved numerically. In this model, we do not need to choose the target return parameter, which has to be decided by subjective judgement in previous distributionally robust mean-variance models. As a result, the distributionally robust Sharpe ratio model is completely data-driven. We also provide guidance on the choice of ambiguity set size by using a much simpler scheme than that employed in the second work. In the end, we compare the performance of this model to that of the second work and some other well-known models on S&P500.

Table of Contents

List of Figures	iii
Acknowledgments	x
1 Introduction	1
1.1 The Classical Single-Period Mean-Variance Markowitz Model	1
1.2 Multi-Period and Continuous-Time MV Models and Time-Inconsistency	2
1.3 Robust MV Models	4
1.4 Main Contributions of This Thesis	9
1.4.1 Naive Strategies in a Continuous-Time MV Model	9
1.4.2 A Distributionally Robust Mean-Variance Model	11
1.4.3 A Distributionally Robust Sharpe Ratio Model	14
1.5 Organization of the Thesis	15
2 Naive Strategies in Continuous-Time MV Model	17
2.1 Problem Formulation	17
2.1.1 Continuous-Time Market	17
2.1.2 Continuous-Time Mean-Variance Model	19
2.2 Strategies of a Naive Agent	21
2.2.1 Time-Inconsistency and the Naive Agent	21
2.2.2 2^n -committed Agent	24
2.3 Naive Strategies and Equilibrium Strategies	38

2.3.1	State-dependent and Myopic Strategies	38
2.3.2	The case $\gamma(x) = \frac{\gamma}{x}$	41
2.3.3	The case $L(t, x) = xe^{k(T-t)}$	44
2.4	Conclusions	46
3	A DRMV Model	48
3.1	Model Formulation	48
3.2	Transformations, Duality, and Regularization	50
3.3	Choice of Model Parameters	61
3.3.1	Choice of δ	62
3.3.2	Choice of $\bar{\alpha}$	71
3.4	Empirical Performance and Comparisons	73
3.4.1	Experiment Design and Data Preparation	74
3.4.2	Comparisons	80
3.4.3	Discussion	86
3.5	Concluding Remarks	96
4	A DRSR Model	97
4.1	Model Formulation	97
4.2	Transformations and Tractability	100
4.3	Choice of Model Parameter δ	110
4.4	Empirical Performance and Comparisons	120
4.4.1	Experiment Design and Data Preparation	120
4.4.2	Comparisons	124
4.5	Concluding Remarks	128
	Bibliography	135

List of Figures

2.1	This graph shows two sample paths of wealth processes. The blue dashed line represents the wealth process corresponding to the optimal portfolio obtained at time 0. The red line is the wealth process corresponding to the optimal portfolio obtained at time t	23
2.2	This figure shows one sample path of wealth process $X_n(s)$. Each different color shows the wealth process corresponding to the optimal portfolio obtained at time $t_k, k = 0, 1, \dots, 2^n - 1$. The process is continuous.	24
3.1	This graph presents the wealth processes of all portfolios (including continuous-time Markowitz) and the S&P 500 from January 2000 to December 2016. All of the portfolios except S&P 500 consist of 100 stocks, and the averages are calculated over 100 numerical experiments. The x -axis indicates the time in months (from 1 to 204) and the y -axis indicates the portfolio wealth. Initial wealth is set at 1.	80
3.2	This graph presents the wealth processes of all portfolios (excluding continuous-time Markowitz) and of the S&P 500 from January 2000 to December 2016. All of the portfolios except the S&P 500 consist of 100 stocks, and the averages are calculated over 100 numerical experiments. The x -axis indicates the time in months (from 1 to 204) and the y -axis indicates the portfolio wealth. Initial wealth is set to be 1.	82

3.3	This graph presents the histograms of the annualized returns of the 100 different experiments on the DRMV (blue) and equal-weighting (orange) portfolios. The x -axis represents the annualized returns and the y -axis represents the number of returns.	83
3.4	This graph presents the histograms of the Sharpe ratio of the 100 different experiments on the DRMV (blue) and equal-weighting (orange) portfolios. The x -axis represents the Sharpe ratio and the y -axis represents the number of Sharpe ratios.	83
3.5	This graph presents the histograms of the kurtosis of the 100 different experiments on the DRMV (blue) and equal-weighting (orange) portfolios. The x -axis represents the kurtosis and the y -axis represents the number of kurtosises.	84
3.6	This graph presents the histograms of the annualized returns of the 100 different experiments on DRMV (blue) and Fama-French (orange) portfolios. The x -axis represents the annualized returns and the y -axis represents the numbers of returns.	85
3.7	This graph presents the histograms of the Sharpe ratio of the 100 different experiments on the DRMV (blue) and Fama-French (orange) portfolios. The x -axis represents the Sharpe ratio and the y -axis represents the number of Sharpe ratios.	85
3.8	This graph presents the histograms of the kurtosises of the 100 different experiments on the DRMV (blue) and Fama-French (orange) portfolios. The x -axis represents the kurtosis and the y -axis represents the number of kurtosises.	86

3.9	This graph presents the histograms of the annualized returns of the 100 different experiments on the DRMV (blue) and Goldfarb–Iyengar (orange) portfolios. There are two experiments in which Goldfarb–Iyengar went into bankruptcy, which are not included in this histogram. The x -axis represents the annualized returns and the y -axis represents the number of returns.	87
3.10	This graph presents the histograms of the Sharpe ratios of the 100 different experiments on the DRMV (blue) and Goldfarb–Iyengar (orange) portfolios. The x -axis represents the Sharpe ratios and the y -axis represents the number of Sharpe ratios.	88
3.11	This graph presents the histograms of the kurtosises of the 100 different experiments on the DRMV (blue) and Goldfarb–Iyengar (orange) portfolios. The x -axis represents the kurtosis and the y -axis represents the number of kurtosises.	88
3.12	This graph presents the histograms of the annualized returns of the 100 different experiments on the DRMV (blue) and single-period Markowitz (orange) portfolios. The x -axis represents the annualized returns and the y -axis represents the number of returns.	89
3.13	This graph presents the histograms of the Sharpe ratios of the 100 different experiments on the DRMV (blue) and single-period Markowitz (orange) portfolios. The x -axis represents the Sharpe ratios and the y -axis represents the number of Sharpe ratios.	90
3.14	This graph presents the histograms of the kurtosises of the 100 different experiments on the DRMV (blue) and single-period Markowitz (orange) portfolios. The x -axis represents the kurtosis and the y -axis represents the number of kurtosises.	90

3.15	This graph presents the histograms of the annualized returns of the 100 different experiments on the DRMV (blue) and Black–Litterman (orange) portfolios. The x -axis represents the annualized returns and the y -axis represents the number of returns.	91
3.16	This graph presents the histograms of the Sharpe ratios of the 100 different experiments on the DRMV (blue) and Black–Litterman (orange) portfolios. The x -axis represents the Sharpe ratios and the y -axis represents the number of Sharpe ratios.	92
3.17	This graph presents the histograms of the kurtosises of the 100 different experiments on the DRMV (blue) and Black–Litterman (orange) portfolios. The x -axis represents the kurtosis and the y -axis represents the number of kurtosises.	92
3.18	This graph presents the portfolio average wealth processes of DRMV ($p = 2$) and DRMV-p1 ($p = 1$) from January 2000 to December 2016. The averages are calculated over 100 numerical experiments. The x -axis indicates the time and the y -axis indicates the portfolio wealth. Initial wealth is set at 1.	93
3.19	This graph presents DRMV’s average wealth processes from January 2000 to December 2016 with different values of ρ . The averages are calculated over 100 numerical experiments. The x -axis indicates the time and the y -axis indicates the portfolio wealth. Initial wealth is set at 1.	93
3.20	This graph presents the histogram of the DRMV monthly turnover rates of 100 experiments. The x -axis represents the turnover rate (%) and the y -axis represents the numbers of turnover rates.	94

3.21	This graph presents portfolios' average wealth processes using shrinkage estimators from January 2000 to December 2016. All the portfolios except S&P 500 consist of 100 stocks and the averages are calculated over 100 numerical experiments. The x -axis indicates the time in months (from 1 to 204) and the y -axis indicates the portfolio wealth. Initial wealth is set to be 1.	95
4.1	This graph presents the wealth processes of all portfolios (allow short-selling) and of the S&P 500 from January 2000 to December 2016. All of the portfolios except the S&P 500 consist of 100 stocks, and the averages are calculated over 100 numerical experiments. The x -axis indicates the time in months (from 1 to 204) and the y -axis indicates the portfolio wealth. Initial wealth is set at 1.	124
4.2	This graph presents the wealth processes of all portfolios (allow short-selling and without classical Sharpe ratio and Fama-French model) and of the S&P 500 from January 2000 to December 2016. All the portfolios except the S&P 500 consist of 100 stocks, and the averages are calculated over 100 numerical experiments. The x -axis indicates the time in months (from 1 to 204) and the y -axis indicates the portfolio wealth. Initial wealth is set at 1.	125
4.3	This graph presents the wealth processes of all portfolios (NOT allowing short-selling) and of the S&P 500 from January 2000 to December 2016. All the portfolios except S&P 500 consist of 100 stocks, and the averages are calculated over 100 numerical experiments. The x -axis indicates the time in months (from 1 to 204) and the y -axis indicates the portfolio wealth. Initial wealth is set at 1.	127

4.4	This graph presents the histograms of the annualized returns of the 100 different experiments on the DRSR (blue) and classical Sharpe ratio (orange) portfolios. The x -axis represents the annualized return and the y -axis represents the number of returns.	127
4.5	This graph presents the histograms of the Sharpe ratio of the 100 different experiments on the DRSR (blue) and classical Sharpe ratio (orange) portfolios. The x -axis represents the Sharpe ratio and the y -axis represents the number of Sharpe ratios.	128
4.6	This graph presents the histograms of the annualized returns of the 100 different experiments on the DRSR (blue) and single-period Markovitz (orange) portfolios. The x -axis represents the annualized return and the y -axis represents the number of returns.	129
4.7	This graph presents the histograms of the Sharpe ratio of the 100 different experiments on the DRSR (blue) and single-period Markovitz (orange) portfolios. The x -axis represents the Sharpe ratio and the y -axis represents the number of Sharpe ratios.	130
4.8	This graph presents the histograms of the annualized returns of the 100 different experiments on the DRSR (blue) and Black-Litterman (orange) portfolios. The x -axis represents the annualized return and the y -axis represents the number of returns.	131
4.9	This graph presents the histograms of the Sharpe ratio of the 100 different experiments on the DRSR (blue) and Black-Litterman (orange) portfolios. The x -axis represents the Sharpe ratio and the y -axis represents the number of Sharpe ratios.	131
4.10	This graph presents the histograms of the annualized returns of the 100 different experiments on the DRSR (blue) and Fama-French (orange) portfolios. The x -axis represents the annualized return and the y -axis represents the number of returns.	132

4.11	This graph presents the histograms of the Sharpe ratio of the 100 different experiments on the DRSR (blue) and Fama-French (orange) portfolios. The x -axis represents the Sharpe ratio and the y -axis represents the number of Sharpe ratios.	132
4.12	This graph presents the histograms of the annualized returns of the 100 different experiments on the DRSR (blue) and DRMV (orange) portfolios. The x -axis represents the annualized return and the y -axis represents the number of returns.	133
4.13	This graph presents the histograms of the Sharpe ratio of the 100 different experiments on the DRSR (blue) and DRMV (orange) portfolios. The x -axis represents the Sharpe ratio and the y -axis represents the number of Sharpe ratios.	133
4.14	This graph presents the histograms of the annualized returns of the 100 different experiments on the DRSR (blue) and equal-weighting (orange) portfolios. The x -axis represents the annualized return and the y -axis represents the number of returns.	134
4.15	This graph presents the histograms of the Sharpe ratio of the 100 different experiments on the DRSR (blue) and equal-weighting (orange) portfolios. The x -axis represents the Sharpe ratio and the y -axis represents the number of Sharpe ratios.	134

Acknowledgments

This thesis concludes my five years of doctoral study and research efforts. In the course of my Ph.D. life, I have become deeply indebted to many excellent scholars and exceptional individuals.

First and foremost, I would like to thank my advisor, Professor Xunyu Zhou, for his careful and enlightening guidance. He is a talented researcher and an insightful mentor. It has been a great honor for me to work with him closely over the course of my Ph.D. career. He came to Columbia University in the second year of my Ph.D. process, and he gave me the confidence and support that put me on the right track. I will always remember lots of precious moments when I went to his office and discussed problems with him using the whiteboard. He has taught me how to be a good researcher and, even more importantly, how to be a good person.

Besides my advisor, I would like to thank the rest of my thesis committee, Professor Agostino Capponi, Professor David Yao, Professor Henry Lam, and Professor Paul Glasserman, for generously offering their precious time to serve on my defense committee and for providing invaluable advice. Also, I want to thank Professor Jose Blanchet, with whom I collaborated along with Professor Zhou on the work of Chapters 3 and 4. I also want to thank Professor Jing Dong, with whom I worked on another research project about exact sampling of the maximum of the Gaussian Processes with infinite memory. I really appreciate the guidance and kind help of these two scholars, as well as my committee members.

I am also grateful to many friends and colleagues whom I have met over the past

five years. I am indebted to several senior colleagues who have given me lots of help during my Ph.D. process: Wei You, Di Xiao, Brian Ward, Cun Mu, Fei He, Chaoxu Zhou. Also, I want to thank the members of my academic family, Professor Xuedong He, Professor Hanqing Jin, Professor Sang Hu, Xiao Xu, and Yilie Huang. I also thank my fellow Ph.D. students, Fengpei Li, Yunjie Sun, Weilong Fu, Yi Ren, Yan Chen, Wenjun Wang, Yunhao Tang, Jingtong Zhao, Raghav Singal, Kumar Goutam, Randy Jia, as well as my friends out of the IEOR department, Yiwen Shen, Weizhuo Sun, Zhongyi Zhang, Xun Wang, Qiliang Lin, Siyu Zhu, Yanru Chen, Tianyang Zhang and Dawei Zhang. Thanks also to all of my friends, this journey has been exciting and enjoyable with your company, and I will always cherish the memories of us attending courses, discussing problems, talking, and laughing together.

I thank my advisor, Columbia University, and the CKGSB doctoral fellowship program for their generous financial support.

Lastly and most importantly, I owe my deepest gratitude to my parents, Yonggang Chen and Lifang Xian, and to my wife, Sixia Su, for their unconditional love and endless support throughout my life. Their faith in me has always encouraged me.

To my parents Yonggang, Lifang and my wife Sixia.

Chapter 1

Introduction

1.1 The Classical Single-Period Mean-Variance Markowitz Model

The Mean-Variance (MV) Markowitz model for portfolio selection formulated in Markowitz (1952) and Markowitz (1959) is one of the best-known classical results in financial economics. It tries to minimize the risk of the terminal portfolio wealth subject to archiving a prescribed return target in a single-period investment. In the MV model, the *return* of a portfolio is quantified by the expectation of the random portfolio return, and the *risk* is measured by the statistical variance of the random portfolio return. The MV model is widely used in the financial industry and has provided a fundamental basis for classical financial theory. One of its important consequences, the mutual fund theorem, contributes to the Capital Asset Pricing Model (CAPM, see Sharpe (1964), Lintner (1965), and Mossin (1966)), which has had a profound impact on the pricing of assets.

Since the emergence of the MV model, a vast number of extensions have been created on this topic. There are two major fields among them. The first consists of attempts to generalize the single-period MV model to its multi-period and continuous-time versions. This extension has generated an interesting problem, called *time-*

inconsistency, which will be elaborated in Section 1.2. Another direction is to *robustify* the single-period MV model. The goal of this extension is to overcome the MV model's sensitivity to parameter estimation. The robust MV literature is discussed in Section 1.3.

This thesis contributes to both of these areas of exploration, particularly, in studying naive agents under the continuous-time MV model and in building distributionally robust MV and Sharpe-ratio models. The following two sections provide a review of the related literature in these two fields.

1.2 Multi-Period and Continuous-Time MV Models and Time-Inconsistency

The problem of multi-period portfolio selection has been extensively studied, e.g., Hakansson (1971), Elton and Gruber (1974a), Elton and Gruber (1974b) and Samuelson (1969). In addition, Chen et al. (1971) examined the difficulties in finding optimal solutions for a multi-period MV model. To be more precise, authors showed that, in order to obtain the optimal solutions for a n -asset and t -period MV problem, we need to solve $(2^n - 2)^{t-1}$ quadratic programming problems, which is computationally expensive. Li and Ng (2000) introduced an embedding technique and successfully derived the analytical optimal portfolio policy and mean-variance efficient frontier for a multi-period MV model. Yin and Zhou (2004) studied discrete-time MV models in a market with regime switching. On the other hand, exploiting the embedding technique and linear-quadratic (LQ) framework control theory, Zhou and Li (2000) solved the continuous-time MV problem with deterministic diffusion coefficients. A more general problem was solved by Lim and Zhou (2002), where the underlying dynamic processes have random drift and diffusion coefficients. Continuous-Time MV model under no bankruptcy constraint was studied in Bielecki et al. (2005a). Dai et al. (2010) studied a continuous-time MV model with transaction costs, and Jin

and Zhou (2015) explored the problem under ambiguity.

In both multi-period and continuous-time MV models, there arises the issue of *time-inconsistency*. Multi-period and continuous-time MV problems are *time-inconsistent* in the sense that the Bellman Optimality Principle does not hold. This is due to the fact that, in the objective function, the variance does not satisfy the so-called *tower rule*. As a consequence, the standard dynamic programming approach can not be directly applied. In the absence of time-consistency, an “optimal” strategy derived at a given moment is generally not optimal when evaluated at the next moment; hence, there is no such things as a “dynamic optimal strategy” good for the entire horizon, as is the case with a time-consistency model. Therefore, in the time-inconsistent situation, instead of trying to find the optimal strategy, researchers concentrate on describing the behaviors of different agents. Economists, starting from Strotz (1956), have described the behaviors of three types of individuals in a time-inconsistent situation. Type 1, the naive agent, always greedily chooses the optimal strategy under his current preference and known information, without realizing that his preference will constantly change in the future. Type 2 is the precommitted agent, who optimizes only once at the initial time and then commits to the resulting strategy during the whole time horizon. Type 3, the sophisticated agent, is aware of the fact that his “future selves” will change his current strategy and chooses the best current action, taking future disobedience as a constraint. The resulting strategy is called an equilibrium, from which no incarnations of the agent at different times have incentive to deviate.

Both precommitted agents and sophisticated agents have been studied in the context of MV models. For the former, Richardson (1989) was probably the first to study precommitted agents in a continuous-time MV model. Zhou and Li (2000) developed an embedding technique to change the originally time-inconsistent MV problem into a stochastic LQ control problem and derived explicit expression for the precommitted optimal portfolio. Further studies and improvements have been put

forward by Lim and Zhou (2002), Bielecki et al. (2005a), and Xia (2005). With respect to sophisticated agents, the game theoretical approach to time-inconsistency for MV models was first studied in Basak and Chabakauri (2010). Bjork and Murgoci (2010) considered a more general class of objective functions. Bjork et al. (2014) studied the equilibrium MV strategy with a state-dependent risk aversion parameter. He and Jiang (2017) investigated the equilibrium strategy (which they designated as the “myopic” strategy) in a model in which they set the target return to be proportional to the agent’s current wealth.

Compared with the rich literature on precommitted and sophisticated agents, there are far fewer works that study the behavior of naive agents, especially in the context of continuous-time MV model. In a discrete-time MV model, to investigate the behavior of naive agent, we only need to find the optimal precommitted strategy in each period. However, in a continuous-time MV model, the number of periods is infinite and the time line is continuous. Therefore, the method applied in a discrete-time MV model is not useful in the continuous-time MV model. Part of this thesis is devoted to solving this problem.

1.3 Robust MV Models

The classical one-period Markowitz mean–variance model involves choosing a portfolio weighting vector $\phi \in \mathbb{R}^d$ (by convention, all of the vectors in this thesis are columns) among d stocks to maximize the risk-adjusted expected return. The precise formulation can be given as¹

$$\min_{\phi \in \mathbb{R}^d} \{ \phi^\top \text{Var}_{\mathbb{P}^*} (R) \phi : \phi^\top \mathbf{1} = 1, \phi^\top \mathbb{E}_{\mathbb{P}^*} (R) = \rho \}, \quad (1.1)$$

where R is the d -dimensional vector of random returns of the stocks; \mathbb{P}^* is the probability measure underlying the distribution of R ; $\mathbb{E}_{\mathbb{P}^*}$ and $\text{Var}_{\mathbb{P}^*}$ are ,respectively,

¹There are several mathematically equivalent formulations of the original mean–variance model.

the expectation and variance under \mathbb{P}^* ; and ρ is the targeted expected return of the portfolio.

It is well-known that this model has a major drawback when applied in practice. On one hand, its solutions are very sensitive to the underlying parameters, namely the mean and the covariance matrix of the stocks (see Michaud (1989) and Chopra and Ziemba (1993)). On the other hand, $\mathbb{E}_{\mathbb{P}^*}[R]$ and $\text{Var}_{\mathbb{P}^*}[R]$ are unknown in practice, so one has to resort to the empirical versions of the mean and the covariance matrix instead, which are usually significantly deviated from the true ones (especially the mean, due to the notorious “mean-blur” problem).

This problem motivates the development of a “robust” formulation of the Markowitz model to reduce the effects of errors brought by estimating $\mathbb{E}_{\mathbb{P}^*}[R]$ and $\text{Var}_{\mathbb{P}^*}[R]$. There are two main techniques to robustify the MV model: the robust estimators method and robust optimization.

The robust estimators method seeks to reduce the estimation error brought by empirical estimators (sample mean and sample covariance). The sample mean and sample covariance matrix are the maximum likelihood estimators (MLE) for $\mathbb{E}_{\mathbb{P}^*}[R]$ and $\text{Var}_{\mathbb{P}^*}[R]$ based on *normally distributed* returns. However, it is well-known that, in financial markets, stock returns do not follow the normal distribution, which strongly influences the efficiency of empirical estimators. Therefore, the purpose of the robust estimator technique is to design estimators that will be stable even when the empirical distribution deviates from the normal distribution. The pioneering work in this regard was done by Huber (1964) and Hampel (1964). Based on their theories, numerous methods have been developed to improve the estimators under the MV model. Perret-Gentil and Victoria-Feser (2005) showed that the use of statistically robust estimators instead of empirical estimators is highly beneficial for the stability properties of MV optimal portfolios under heavy tail distributions. Welsch and Zhou (2007) investigated several robust statistical approaches (e.g., FAST-MCD) and penalization to increase the stability of the MV portfolio. DeMiguel and Nogales (2009) proposed

a one-step framework where the MV portfolio optimization and robust estimation are performed in a single step.

The robust optimization technique attempts to solve the sensitivity problem by introducing uncertainty (ambiguity) sets to incorporate the possible estimation errors. More specifically, it assumes that the true parameters are within a pre-defined uncertainty set. Then, the robust optimization model generates optimal solutions by optimizing the worst-case performance over the uncertainty set. The framework was first introduced in Ben-Tal and Nemirovski (1999) for linear programming and in Ben-Tal and Nemirovski (1998) for general convex programming. Based on their work, Goldfarb and Iyengar (2003) built a robust factor model to overcome the parameter sensitivity in MV. Tütüncü and Koenig (2004) developed a robust MV optimization model whose uncertainty set includes the most likely realizations of the input parameters. Garlappi et al. (2007) formulated a “multi-prior” robust model in which they incorporate both parameter uncertainty and model uncertainty. Lu (2011) considered a robust maximum risk-adjusted return (RMRAR) based on the same factor model as that of Goldfarb and Iyengar (2003). They showed that they could outperform Goldfarb and Iyengar (2003) by combining RMRAR and their joint uncertainty set. Ye et al. (2012) introduced joint uncertainty regions over both the mean vector and the covariance matrix of returns to reduce the conservatism brought by separable uncertainty sets. All the works referred to in the preceding paragraph belong to classical robust optimization, in which the uncertainty sets are typically conic bounded convex sets. Another line of research is distributionally robust optimization (DRO).

The key difference between DRO and classical robust optimization is that, in DRO models, the uncertainty set consists of a set of probability distributions. The history of DRO dates back to the 1950s. Scarf (1958) was the first to study a single-product newsvendor problem in order to maximize the worst-case expected profit, where the worst-case expectation is taken over all the demand probability distributions with only mean and variance known. This approach has been further developed by

Zackova (1966), Dupacova (1987), Lagoa and Barmish (2002), Shapiro and Kleywegt (2002), and Shapiro and Ahmed (2004). More recently, Chen et al. (2007) introduced directional deviations as a way to characterize a family of distributions. Chen and Sim (2009) applied the result of Chen et al. (2007) to build a goal-driven optimization model. Delage and Ye (2010) studied a distributionally robust stochastic program in which they constructed ambiguity regions involving means and covariance matrices of the return vector.

Like the classical approach, DRO and its theoretical extensions have been widely used in the field of portfolio selection. In particular, researchers have considered all the possible distributions of returns as the ambiguity set and formulated distributionally robust models to minimize the worst case portfolio risk. Two methods have typically been used to construct the ambiguity set, the *moment-based* method and the *statistical distance-based* method. The Moment-based technique considers distributions whose moments (such as mean vector and covariance matrix) satisfy certain conditions (such as being bounded). Lobo and Boyd (2000) was among the first to provide a worst-case robust analysis with respect to second-order moment uncertainty within the Markowitz framework. El Ghaoui et al. (2003) built a robust portfolio selection model for worst case Value-at-Risk (VaR) where only the bounds of means and the covariance matrices of the returns are known. Popescu (2007) derived robust solutions to certain stochastic optimization problems, based on mean-covariance information about the distributions underlying the uncertain vector of returns. Natarajan and Sim (2010) considered a robust expected utility model in which no restriction (e.g., normality) was added on return distributions. Wozabal (2012) considered a robust portfolio model with risk constraints based on expected short-fall, resulting in an optimization problem that requires solving multiple convex problems. Zymler et al. (2013) developed a robust joint chance constraints model assuming only the first and second-order moments and support of the uncertain parameters are given.

The statistical-distance based method constructs the ambiguity set by consid-

ering distributions that are within a certain distance from a nominal distribution (usually empirical distribution or normal distribution), where the distance is a carefully chosen statistical distance. Popular choices for the statistical distance include the ϕ -divergence (Bayraksan and Love (2015), Wang et al. (2016)), the Prokhorov metric (Erdogan and Iyengar (2006)), the Kullback-Leibler divergence (Jiang and Guan (2016)), and the Wasserstein distance. Glasserman and Xu (2014) built a robust approach to quantify model risk and bound the impact of model error using the KL-divergence, where they also characterized the worst case probability measures. A Wasserstein distance is the optimal value for a specific optimal transport problem. The notion was first formulated by Monge (1781) and its theory developed by Kantorovich (1942). It has been widely used in the study of distributionally robust optimization (DRO) problems. Pflug and Wozabal (2007) presented a Markowitz model with distributional robustness based on the Wasserstein distance. Duality results for Wasserstein DRO formulations in which the probability model appears linearly in the objective function have been studied in Zhao and Guan (2018), Esfahani and Kuhn (2018), and Gao and Kleywegt (2016). A general duality result (with conditions that match the standard assumptions of the general optimal transport theory) is given in Blanchet and Murthy (2019). Esfahani and Kuhn (2018) provide representations for the worst-case expectations in a Wasserstein-based ambiguity set centered at the empirical measure, and then apply their results to portfolio selection using different risk measures, leading to models that differ from the Markowitz model.

Chapters 3 and 4 of this thesis use the Wasserstein distance, owing to several advantages. The first advantage is its tractability. DRO problems with Wasserstein ambiguity sets can often be reformulated as finite convex programs (see Gao and Kleywegt (2016) and Zhao and Guan (2018)), which can in practice usually be quickly solved numerically. Second, the Wasserstein distance does not require two probability distributions to have the same support set. This means that the ambiguity set will include many possible distributions, even when we use the empirical distribution

as the nominal distribution. Third, it has been shown (see Blanchet et al. (2016)) that the Wasserstein ambiguity sets have a close connection with the use of regularization in some classical machine learning models, which motivated us to find the connection between a distributionally robust mean-variance (DRMV) model and the regularized MV model.

1.4 Main Contributions of This Thesis

Despite enormous efforts to develop the Markowitz model in its continuous-time and robust versions, many interesting research problems remain. On one hand, this thesis develops a general methodology to study the naive agent in continuous-time MV portfolio selection, which can also be applied in other time-inconsistent models such as Yaari's dual theory (see Yaari (1987)) and Lope's SP/A theory (see Lopes (1987)). On the other hand, this thesis contributes to the robust MV model by developing a distributionally robust MV (DRMV) model with the Wasserstein distance. We show its tractability and develop an approach to decide the size of the ambiguity set. We also apply DRMV to the real U.S. market and find that it outperforms a number of well-known strategies (e.g., Fama-French and Black-Litterman). In addition, we develop a distributionally robust Sharpe ratio (DRSR) model. The remainder of this section describes the three core problems of this thesis and highlights our contributions.

1.4.1 Naive Strategies in a Continuous-Time MV Model

As mentioned previously, to investigate the behavior of naive agents in the discrete-time multi-period MV model, we need only to find the precommitted strategy in each time period. In a continuous-time MV model, the notion of a "time period" does not apply because the time is continuous. The naive agent in the continuous-time MV model constantly changes his strategy, which makes it difficult to analyze his behavior directly. Motivated by the multi-period case, we propose a new type of

agent who behaves in a manner “in between” that of the precommitted agent and the naive agent. The behavior of the new agent will approximate that of the naive one and, under some conditions, converge to the naive strategy.

In Chapter 2 we propose a new type of agents so called “ 2^n -committed agent”, who discretizes the whole time horizon into 2^n equal length time intervals, and we obtain the precommitted strategies for each interval using the same approach as in Zhou and Li (2000). Then, the 2^n -committed agent implements the precommitted strategy in each time interval, where the initial wealth of every time interval is set to be the terminal wealth of the previous interval. This construction leads to a continuous wealth process that approximates the wealth process of the naive agent as n increases. Then, by letting $n \rightarrow \infty$, we prove that the previously obtained wealth process converges weakly, and we derive explicit expressions for the limiting wealth process and the corresponding portfolio process. This derived portfolio process necessarily chooses optimal strategy at any time in line with the agent’s preference at that point of time generating a naive strategy. Pendersen and Peskir (2017) also considered the naive agent (referred to in their paper as “dynamic optimality”) in a continuous-time MV model with a one-dimensional Black-Scholes market. They conjectured the analytical formula for a naive agent from that of a precommitted agent without showing how they derived it. This method may not work under a more general setting (e.g., one with more than one risky asset) or with some other time-inconsistent problems. In contrast, our approach may be used for a much more general setting.

We make two main contributions in this chapter. First, we derive the analytical formulae for the naive agent in an MV model under a very general setting. Then, we explore the relationship between the naive strategies and two equilibrium strategies, those of Bjork et al. (2014) and He and Jiang (2017), in the Black-Scholes market, where there’s only one risky asset and one risk-free asset. We find that naive strategies are indeed more risk-seeking than equilibrium strategies. To be more precise, naive

strategies tend to allocate more weight to the risky asset than equilibrium strategies. This may be due to the fact that, the sophisticated agents take future disobedience into consideration, while the naive agent only cares about the current state.

More importantly, we develop a general methodology, the “ 2^n -committed agent” approach, to study the behavior of the naive agent. This approach has three main steps. The first step is *discretization*, in which we discretize the time line into many small time periods and obtain the precommitted strategy in each period. By pasting all the precommitted strategies together, we construct the dynamic for the 2^n -committed agent. The second step is *boundedness*. Here, we prove the sequence of wealth processes of the 2^n -committed agent are uniformly bounded by a carefully constructed integrable function. The last step is *convergence*. Here, we derive the limiting process and prove that the sequence of wealth processes weakly converge to the limiting process. This 2^n -committed agent approach is not specially designed for use with a continuous-time MV model. In fact, our methodology can be applied to many other problems, including the general stochastic linear quadratic (LQ) control problem (see Chen and Zhou (2020)).

1.4.2 A Distributionally Robust Mean-Variance Model

In Chapter 3, we are interested in studying a distributionally robust mean-variance (DRMV) model, given by

$$\min_{\phi \in \mathcal{F}_{\delta, \bar{\alpha}}(n)} \max_{\mathbb{P} \in \mathcal{U}_{\delta}(\mathbb{P}_n)} \{ \phi^\top \text{Var}_{\mathbb{P}}(R) \phi \}, \quad (1.2)$$

where \mathbb{P}_n is the empirical probability derived from historical information on the sample size n , $\mathcal{U}_{\delta}(\mathbb{P}_n)$ is the ambiguity set, $\mathcal{F}_{\delta, \bar{\alpha}}(n)$ is the feasible region of portfolios, and $\mathbb{E}_{\mathbb{P}}[R]$ and $\text{Var}_{\mathbb{P}}(R)$ denote, respectively, the mean and the covariance matrix under \mathbb{P} .

Intuitively, formulation (1.2) introduces an artificial adversary \mathbb{P} (whose problem

is that of the inner maximization) as a tool to account for the impact of the model uncertainty around the empirical distribution. There are two key parameters, δ and $\bar{\alpha}$, in this formulation, and they need to be chosen carefully. The parameter δ can be interpreted as the power given to the adversary: The larger the value of δ , the more power is given. If δ is too large relative to the evidence (i.e., the size of n), then the portfolio selection will tend to be unnecessarily conservative. On the other hand, $\bar{\alpha}$ can be regarded as the lowest acceptable target return given the ambiguity set. Naturally, the choice of $\bar{\alpha}$ should be based on the original target ρ given in (1); but one also needs to take into account the size of the distributional uncertainty, δ . Using $\bar{\alpha} = \rho$ will tend to generate portfolios that are too aggressive; it is more sensible to choose $\bar{\alpha} < \rho$ in a way such that $\rho - \bar{\alpha}$ is naturally informed by δ .

Chapter 3 offers three main contributions. First, we show that (1.2) is equivalent to an (explicitly formulated) non-robust minimization problem in terms of the empirical probability measure in which a proper penalty term or “regularization term” is added to the objective function. The explicit regularization term that is derived from (1.2) is given in Theorem 2 below. This connects (and contrasts) to the *directly introduced* use of regularization in variance minimization techniques that is widely employed both in the machine learning literature and in practice to, among other things, address the issue of overfitting. Our result shows that our robust strategies are able to enhance out-of-sample performance with basically the same level of computational tractability as the standard mean-variance selection. Note that the results of Gao and Kleywegt (2016), Zhao and Guan (2018), and Esfahani and Kuhn (2018), all of whom studied the duality and tractability for Wasserstein DRO, can not be applied directly to our setting. The key reason is that, the object function in the DRO formulation used by these authors appears to be linear in probability measure, whereas in (1.2), it is quadratic.

Our second main contribution provides guidance regarding the choice of the size of the ambiguity set, δ , as well as that of the worst mean return target, $\bar{\alpha}$. This

is accomplished by adapting and extending the robust Wasserstein profile inference (RWPI) framework, recently introduced and developed by Blanchet et al. (2016), to our setting in a data-driven way that combines optimization principles and basic statistical theory under suitable mixing conditions on historical data. We also note that the work of Esfahani and Kuhn (2018) provided guidance in choosing the uncertainty size, δ . However, this choice of the uncertainty size deteriorates substantially with an increase in the dimension of the underlying portfolio. Thus, as we will elaborate, we employ an approach similar to that proposed in Blanchet et al. (2016), which must be adapted and extended to our setting.

The last contribution empirically compares the performance of our DRMV strategies with those of several well-known and well-practiced models, including the classical Markowitz model, the Fama–French model and the Black–Litterman model. We also compare our strategies with those of another robust model, the one put forward by Goldfarb and Iyengar (2003), in which robustness is based on vector/matrix distance. All these models (including ours) are static, single-period ones, whereas in practice, a stock market is highly dynamic. In our empirical experiments, we implement these models in a rolling horizon fashion in order to account for the market dynamics. Finally, we include in our comparison dynamically optimal strategies based on the well-calibrated, non-robust model presented in Cui et al. (2012). The experiments are carried out on the S&P 500 for the backtesting period 2000-2017, with the prior ten years as the training period. Our experiments show that DRMV compares favorably against all other models in achieving no worse (far better against most models) average returns and much lower variabilities. This result, we believe, constitutes important insight that we can draw from this research and that merits further investigation.

1.4.3 A Distributionally Robust Sharpe Ratio Model

In Chapter 4 we study a distributionally robust Sharpe-Ratio (DRSR) model, given by

$$\min_{\phi \in \mathcal{F}(n)} \max_{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)} \left\{ \frac{\sqrt{\phi^\top \text{Var}_{\mathbb{P}}(R) \phi}}{\phi^\top \mathbb{E}_{\mathbb{P}}[R]} \right\}, \quad (1.3)$$

where \mathbb{P}_n is the empirical probability derived from historical information of the sample size n , $\mathcal{U}_\delta(\mathbb{P}_n)$ is the ambiguity set, and $\mathcal{F}(n)$ is the feasible region. $\mathbb{E}_{\mathbb{P}}[R]$ and $\text{Var}_{\mathbb{P}}(R)$ denote, respectively, the mean and the covariance matrix under \mathbb{P} .

Note that, instead of maximizing the Sharpe ratio, in (1.3) we choose to optimize the inverse of the Sharpe ratio, which makes it more convenient to derive the tractability of this problem. This is further elaborated in Section 4.2. The inner maximization part represents the worst case Sharpe ratio inverse and the outer minimization part wants to choose the best portfolio ϕ to optimize the worst case performance. Compared with DRMV, the objective function in (1.3) is much more complicated. In DRMV, the objective function is just a quadratic function of \mathbb{P} , while that of DRSR involves the ratio of two functions. However, there is no target return parameter in DRSR. Due to this fact, we are able to make the DRSR a completely data driven model.

Chapter 4 offers three main contributions. First, we show that (1.3) can be transformed into an equivalent tractable convex optimization problem. We first prove that (1.3) is equivalent to an (explicitly formulated) non-robust minimization problem in terms of the empirical probability measure in Theorem 4. Then, we further transform the problem into a convex optimization problem by taking the square of the objective function. In the end, we can achieve a tractable convex optimization problem as stated in corollary 11. One thing to be noted is that the tractability result from DRMV can not be applied directly here. The objective function in the inner maximization problem of (1.3) is more complicated than that of (1.2), which leads to a much more complex derivation procedure.

Our second main contribution provides guidance regarding the choice of the size of the ambiguity set, δ . We follow a similar process that employed in Chapter 3 and derive a more elegant RWP function. The simple form of the RWP function in DRSR is due to the fact that DRSR has no target return equality constraint.

The last contribution empirically compares the performance of our DRSR strategies with those of several well-known and well-practiced models, including the classical Sharpe ratio model, the classical Markowitz model, the DRMV model, the Fama–French model, and the Black–Litterman model. All of these models (including ours) are static, single-period ones whereas in practice, stock markets are highly dynamic. In our empirical experiments, we implement them in the same environment as that used in Chapter 3. Our experiments show that DRSR does generate an improvement over the classical Sharpe ratio model, specifically with respect to realizing more stable average returns. When adding no short-selling constraints, all of the strategies perform better and DRSR outperforms most strategies (except DRMV and the equal weights strategy) by showing lower variabilities. However, we observe that the performance of DRSR is worse than that of DRMV and equally weighted strategies, in both short-selling allowed and no short-selling scenarios. One of the important reasons may be Sharpe ratio criterion will be very unstable when the portfolio return is close to 0. As discussed in Chapter 4, this leads to a very aggressive constraint in the dual form of (4.4).

1.5 Organization of the Thesis

The organization of the rest of this thesis is as follows. In Chapter 2 we study strategies of naive agents in a continuous-time MV model. In Section 2.1 we formulate the continuous-time MV portfolio selection model. In Section 2.2 we introduce the so-called 2^n -committed strategy, in which the agent is committed only in a small interval of length 2^{-n} , study its limiting behavior, and derive the limiting strategy as

a naive strategy. In Section 2.3 we compare the naive agent strategy derived in this chapter with certain equilibrium strategies in the Black-Scholes market. Section 2.4 concludes the chapter.

Chapter 3 is devoted to DRMV. In Section 3.1 we formulate the DRMV model and present necessary preliminaries. Section 3.2 demonstrates the tractability of the DRMV model after a series of transformations, and Section 3.3 studies the choices of distributional uncertainty size and the worst return level. Then, in Section 3.4, we report the empirical performance of our strategies against those of several other models. Concluding remarks are given in Section 3.5.

In the last chapter, we study DRSR. In Section 4.1 we formulate the DRSR model. Section 4.2 demonstrates the tractability of the DRSR model and Section 4.3 studies the choice of distributional uncertainty size. In Section 4.4 we report the empirical performance of our strategies against those of several other models. Concluding remarks are given in Section 4.5.

Chapter 2

Naive Strategies in a Continuous-Time Mean-Variance Model

2.1 Problem Formulation

In this section we formulate the continuous-time market and review the continuous-time MV model with deterministic coefficients.

2.1.1 Continuous-Time Market

Throughout this chapter $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}_{t \geq 0})$ is a fixed filtered complete probability space on which a standard $\{\mathcal{F}_t\}_{t \geq 0}$ -adapted m -dimensional Brownian motion $W(t) \equiv (W^1(t), \dots, W^m(t))^\top$ is defined.

Notation. We make the following additional notation:

M^\top is the transpose of any vector or matrix M ;

$L^2([0, T]; X)$ is the Hilbert space of X -valued integrable functions on $[0, T]$ endowed with the norm $(\int_0^T \|f(t)\|_X^2 dt)^{1/2}$ for a given Hilbert space X . We denote by $L^2_{\mathcal{F}}([0, T]; \mathbb{R}^m)$ the set of all \mathbb{R}^m -valued, $\{\mathcal{F}_t\}_{t \geq 0}$ adapted stochastic processes $f(t)$ such that $\mathbb{E}[\int_0^T \|f(t)\|^2 dt] < +\infty$. Here, $\|\cdot\|$ denotes the L^2 norm in Euclidean space.

We define a continuous-time financial market following Karatzas and Shreve (1998). In the market there are $m + 1$ assets being traded continuously. One of the assets is a bank account whose price process $S_0(t)$ is subject to the following equation:

$$dS_0(t) = r(t)S_0(t)dt, \quad t \in [0, T]; \quad S_0(0) = s_0 > 0, \quad (2.1)$$

where the interest rate function $r(\cdot)$ is deterministic. The other m assets are stocks whose price processes $S_i(t), i = 1, \dots, m$ satisfy the following stochastic differential equation (SDE):

$$dS_i(t) = S_i(t)[b_i(t)dt + \sum_{j=1}^m \sigma_{ij}(t)dW^j(t)], \quad t \in [0, T]; \quad S_i(0) = s_i > 0, \quad (2.2)$$

where $b(\cdot)$ and $\sigma_{ij}(\cdot)$, the appreciation and volatility rates, respectively, are scalar-valued and deterministic.

Set the excess rate of return process as

$$B(t) := (b_1(t) - r(t), \dots, b_m(t) - r(t))^{\top}$$

and define the volatility matrix process as $\sigma(t) := (\sigma_{ij}(t))_{m \times m}$.

Consider an agent, with initial endowment $x_0 > 0$ and an investment horizon $[0, T]$, whose total wealth at time $t \geq 0$ is denoted by $x(t)$. Assume that the trading of shares takes place continuously in a self-financing fashion and that there are no

transaction costs. Then $x(\cdot)$ satisfies (see e.g., Karatzas and Shreve (1998))

$$dx(t) = [r(t)x(t) + B^\top(t)\pi(t)]dt + \pi(t)^\top \sigma(t)dW(t), \quad t \in [0, T]; x(0) = x_0, \quad (2.3)$$

where $\pi_i(t), i = 1, 2, \dots, m$, denotes the total market value of the agent's wealth in the i -th asset at time t . The process $\pi(\cdot) \equiv (\pi_1(\cdot), \dots, \pi_m(\cdot))^\top$ is called a *portfolio* if $\pi(\cdot) \in L^2_{\mathcal{F}}([0, T]; \mathbb{R}^m)$, and it is \mathcal{F}_t -progressively measurable.

Throughout this chapter, we need to impose several technical assumptions.

A1) $r(t), B(t)$ and $\sigma(t)$ are uniformly bounded $\forall t \in [0, T]$.

A2) $\sigma(t)\sigma(t)^\top \geq \delta I, \forall t \in [0, T]$ for some $\delta > 0$. This is the so-called *non-degeneracy* condition.

2.1.2 Continuous-Time Mean-Variance Model

The mean-variance portfolio optimization problem is

$$\min_{\{\pi(\cdot) \text{ is a portfolio}\}} \text{Var}(X(T)) \quad (2.4)$$

$$\text{subject to } \begin{cases} \mathbb{E}[X(T)] = x_0 f(0, T), \\ (x(\cdot), \pi(\cdot)) \text{ satisfy (2.3)}, \end{cases} \quad (2.5)$$

where $f(t, s), 0 \leq t \leq s \leq T$ is a deterministic function satisfying $f(t, t) = 1, \forall t \in [0, T]$. $f(t, s)$ can be regarded as the target growth rate in the time interval $[t, s]$. We add an assumption on $f(t, s)$:

A3) $f \in C^1([0, T] \times [0, T]), f(t, s) > e^{\int_t^s r(v)dv}, \forall 0 \leq t \leq s \leq T$.

This assumption is reasonable since $e^{\int_t^s r(v)dv}$ is the risk-free growth rate in $[t, s]$.

In order to solve the constrained optimization problem (2.4) - (2.5), we first apply the Lagrange multiplier method to transform (2.4) into a form equivalent equivalent to the following:

$$\min_{\{\pi(\cdot) \text{ is a portfolio}\}} \text{Var}(X(T)) - \frac{1}{\mu} \mathbb{E}[X(T)],$$

which turns out to be the same as problem (2.11) in Zhou and Li (2000). It follows from Theorem 4.1 and equations (5.12) and (6.1) in Zhou and Li (2000) that the optimal portfolio with respect to the problem (2.4) - (2.5) is

$$\bar{\pi}(t, X(t)) = [\sigma(t)\sigma(t)^\top]^{-1}B(t)^\top(\bar{\gamma}e^{-\int_t^T r(s)ds} - X(t)), \quad (2.6)$$

and the corresponding wealth process satisfies

$$\begin{cases} dX(t) = \{(r(t) - \rho(t))X(t) + \bar{\gamma}e^{-\int_t^T r(s)ds}\rho(t)\}dt \\ \quad + B(t)(\sigma(t)\sigma(t)^\top)^{-1}\sigma(t)(\bar{\gamma}, e^{-\int_t^T r(s)ds} - X(t))dW(t), \\ X(0) = x_0 \end{cases} \quad (2.7)$$

where

$$\rho(t) = B(t)[\sigma(t)\sigma(t)^\top]^{-1}B(t)^\top, \quad \bar{\gamma} = \frac{\bar{\lambda}}{2\mu}, \quad \bar{\lambda} = e^{\int_0^T \rho(t)dt} + 2\mu x_0 e^{\int_0^T r(t)dt}.$$

In the above solutions, there is still one unknown variable: μ . We need to solve this μ by using the expectation constraint $\mathbb{E}[X(T)] = x_0 f(0, T)$ in (2.5). By taking the expectation on (2.7) and the time to be T , we obtain a linear ODE

$$d\mathbb{E}[X(T)] = (r(t) - \rho(t))\mathbb{E}[X(T)] + \bar{\gamma}e^{-\int_t^T r(s)ds}\rho(t).$$

Solving this linear ODE, we obtain

$$\mathbb{E}[X(T)] = e^{\int_0^T r(t) - \rho(t)dt}x_0 + (1 - e^{-\int_0^T \rho(t)dt})\bar{\gamma}.$$

The expectation constraint yields

$$\mathbb{E}[X(T)] = e^{\int_0^T r(t) - \rho(t)dt}x_0 + (1 - e^{-\int_0^T \rho(t)dt})\bar{\gamma}(\mu) = x_0 f(0, T).$$

Note that deciding the value of μ is equivalent to deciding the value of $\bar{\gamma}$. Thus, it suffices to get

$$\bar{\gamma} = \frac{f(0, T) - e^{\int_0^T r(t) - \rho(t) dt}}{1 - e^{-\int_0^T \rho(t) dt}} x_0.$$

Remark. In order for the above to be true (i.e., in order for theorem 4.1 of Zhou and Li (2000) to be applied here), we require $\mu > 0$, which is equivalent to

$$\bar{\gamma} - x_0 e^{\int_0^T r(t) dt} = \frac{f(0, T) - e^{\int_0^T r(t) dt}}{1 - e^{-\int_0^T \rho(t) dt}} x_0 > 0.$$

This holds because of our Assumption **A3**). If the above inequality is not satisfied, it means that $f(0, T) \leq e^{\int_0^T r(t) dt}$. Recall that $e^{\int_0^T r(t) dt}$ indicates the risk-free growth rate. If $f(0, T) \leq e^{\int_0^T r(t) dt}$, the optimal choice is to put the entire investment into the risk-free asset, which will achieve the target return and yield 0 portfolio variance. Therefore, the above inequality also guarantees that the continuous-time MV model is non-trivial.

2.2 Strategies of a Naive Agent

In this section we discuss the time-inconsistency property of the problem (2.4)-(2.5) and introduce a naive agent facing time-inconsistency.

2.2.1 Time-Inconsistency and the Naive Agent

As discussed in the previous section, at time 0, we solve problem (2.4)-(2.5) to obtain the precommitted optimal portfolio and the corresponding wealth processes, now denoted respectively by $\pi^*(s, X_*(s; 0); 0)$ and $X_*(s; 0)$, $s \in [0, T]$. Here, the notation ; 0 in $\pi^*(s, X_*(s; 0); 0)$ and $X_*(s; 0)$ means they are the precommitted optimal portfolio and wealth processes of the continuous-time MV problem whose starting time is 0.

We retain this portfolio until time $t \in (0, T)$, where t is a fixed value. At time t ,

we observe our current wealth $X_*(t; 0)$. Now, we re-solve the optimization problem (2.4) - (2.5) based on the information until time t . To be more precise, we consider the following problem:

$$\min_{\{\pi(\cdot) \text{ is a portfolio}\}} \text{Var}(X(T)|\mathcal{F}_t) \quad (2.8)$$

$$\text{subject to } \mathbb{E}[X(T)|\mathcal{F}_t] = X_*(t; 0)f(t, T).$$

Note that, conditional on the information until time t , $X_*(t; 0)$ is a deterministic constant. Thus, problem (2.8) is mathematically the same as problem (2.4), and we can use the same approach in Section 2.1.1 to solve problem (2.8) and obtain a new optimal portfolio and wealth processes, denoted by $\pi^*(s, X_*(s; t); t)$ and $X_*(s; t)$, $s \in [t, T]$, respectively. As before, we use notation $; t$ in $\pi^*(s, X_*(s; t); t)$ and $X_*(s; t)$ to denote the portfolio and wealth process are obtained for the problem starting at time t . The expression for $\pi^*(s, X_*(s; t); t)$ is as follows:

$$\pi^*(s, X_*(s; t); t) = [\sigma(s)\sigma(s)^\top]^{-1}B(s)^\top(\gamma^*(t)X_*(t; 0)e^{-\int_s^T r(v)dv} - X_*(s; t)), \quad (2.9)$$

and $X_*(s; t)$ solves the SDE

$$\begin{cases} dX_*(s; t) = \{(r(s) - \rho(s))X_*(s; t) + \gamma^*(t)X_*(t; 0)e^{-\int_s^T r(v)dv} \rho(s)\} ds \\ \quad + B(s)(\sigma(s)\sigma(s)^\top)^{-1}\sigma(s)(\gamma^*(t)X_*(t; 0)e^{-\int_s^T r(v)dv} - X_*(s; t))dW(s), \\ X_*(t; t) = X_*(t; 0), \end{cases} \quad (2.10)$$

where

$$\gamma^*(t) := \frac{f(t, T) - e^{\int_t^T r(v) - \rho(v)dv}}{1 - e^{-\int_t^T \rho(v)dv}}.$$

From the above expressions, it is clear that, in general, for $s \in (t, T]$:

$$X_*(s; 0) \neq X_*(s; t) \quad (2.11)$$

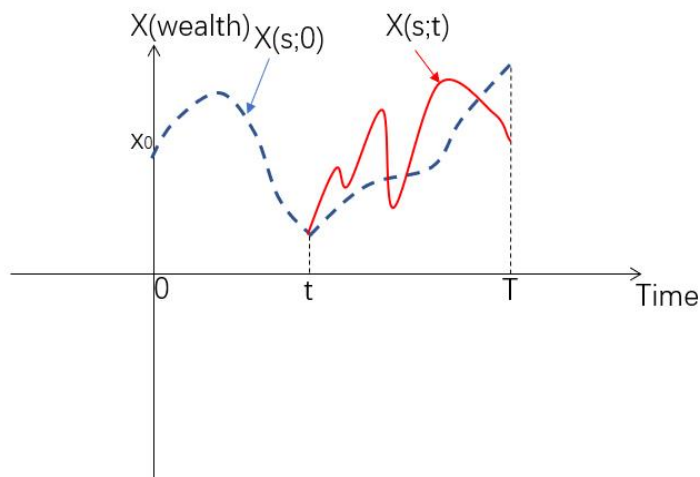


Figure 2.1: This graph shows two sample paths of wealth processes. The blue dashed line represents the wealth process corresponding to the optimal portfolio obtained at time 0. The red line is the wealth process corresponding to the optimal portfolio obtained at time t .

and

$$\pi(s, X_*(s; 0); 0) \neq \pi(s, X_*(s; t); t). \quad (2.12)$$

Figure 2.1 illustrates (2.11).

From (2.11) and (2.12), it follows that the optimal wealth and portfolio processes are always changing over time, which is in sharp contrast to the classical dynamic optimization problem. This property is called *time-inconsistency*.

Recall that a naive agent is one who always chooses the optimal strategy under current information. Therefore, precommitted strategies obtained in (2.6) and (2.7) will not be taken by a naive agent due to time-inconsistency. Our objective in the remainder of this chapter is to derive the explicit expressions of the wealth process and portfolio process for the naive agent, and to compare them with certain other equilibrium strategies derived in the literature. In the following section, we introduce a methodology to approximate the behavior of the naive agent.

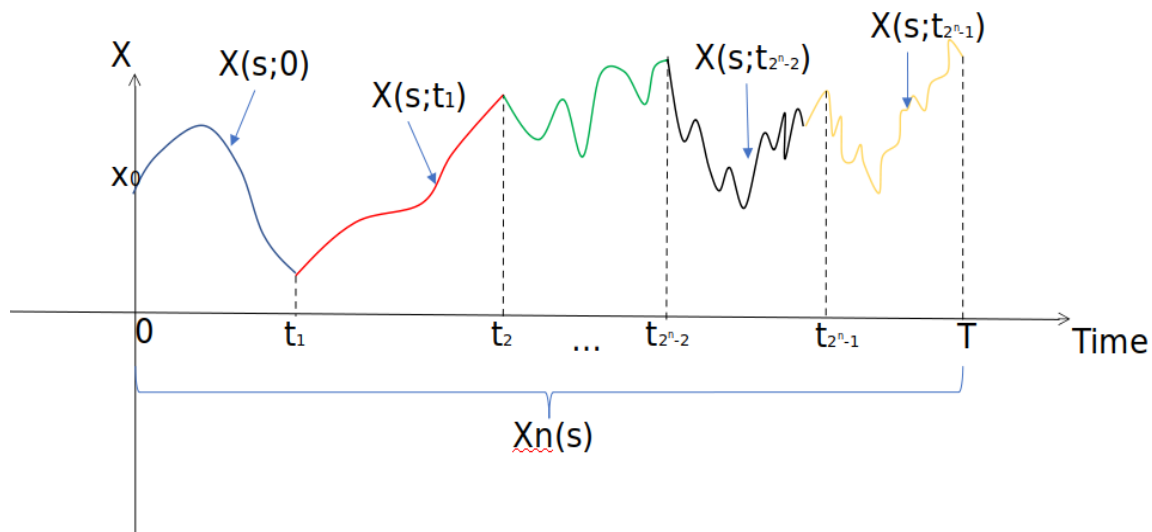


Figure 2.2: This figure shows one sample path of wealth process $X_n(s)$. Each different color shows the wealth process corresponding to the optimal portfolio obtained at time t_k , $k = 0, 1, \dots, 2^n - 1$. The process is continuous.

2.2.2 2^n -committed Agent

Since the naive agent changes her strategy continuously in time, it is difficult to derive the strategy directly. Thus, we first introduce an auxiliary agent, designated the 2^n -committed agent, to approximate the behavior of the naive agent.

The 2^n -committed agent is one who behaves “in between” the precommitted agent and the naive one. To be more precise, the “ 2^n -committed agent” will partition the time horizon $[0, T]$ into 2^n equal-length intervals, and the partitioning points are denoted as $\{t_k\}_{k=0}^{2^n}$, where $t_k = \frac{kT}{2^n}$. This agent first solves problem (2.4)-(2.5) at time 0 to obtain the optimal portfolio $\pi(s, X_*(s;0);0)$. He keeps this portfolio until time t_1 , when he resolves problem (2.8) (where $t = t_1$) to obtain optimal portfolio $\pi(s, X_*(s;t_1);t_1)$. He commits to $\pi(s, X_*(s;t_1);t_1)$ until t_2 . He repeats this pattern until time T . Figure 2.2 illustrates this agent’s wealth process under this strategy.

As a result, we obtain 2^n wealth processes $\{X_*(s;t_k), s \in [t_k, t_{k+1}]\}_{k=0}^{2^n-1}$. By (2.10),

the SDE satisfied by $X_*(s; t_k)$, $s \in [t_k, t_{k+1}]$ ($0 \leq k \leq 2^n - 1$) is as follows:

$$\begin{cases} dX_*(s; t_k) = \{(r(s) - \rho(s))X_*(s; t_k) + \gamma^*(t)X_*(t_k; t_{k-1})e^{-\int_s^T r(v)dv} \rho(s)\} ds \\ \quad + B(s)(\sigma(s)\sigma(s)^\top)^{-1} \sigma(s)(\gamma^*(t)X_*(t_k; t_{k-1})e^{-\int_s^T r(v)dv} - X_*(s; t_k)) dW(s), s \in [t_k, t_{k+1}], \\ X_*(t_k; t_k) = X_*(t_k; t_{k-1}), \end{cases} \quad (2.13)$$

where $X_*(t_0; t_{-1})$ is defined as x_0 . By pasting $X_*(s; t_k)$, $0 \leq k \leq 2^n - 1$, we define

$$X_n(s) := \begin{cases} X_*(s; 0), & 0 \leq s \leq t_1, \\ X_*(s; t_1), & t_1 \leq s \leq t_2, \\ \dots \\ X_*(s; t_{2^n-1}), & t_{2^n-1} \leq s \leq T, \end{cases} \quad (2.14)$$

which is the wealth process of the 2^n -committed agent; see Figure 2.2. Obviously, this process is adapted and continuous in $[0, T]$ and, by following Proposition 1, we know $X_n(\cdot) \in L^2_{\mathcal{F}}([0, T], \mathbb{R})$.

Proposition 1 *If assumptions **A1**, **A2** and **A3** are satisfied, then*

$$\|X_n\|^2 := \mathbb{E}\left[\int_0^T X_n^2(s) ds\right] < \infty, \quad \forall n.$$

Moreover, $\|X_n\|^2$ is uniformly bounded in n .

Proof The main idea of the proof is to find a *deterministic* process $Y^2(\cdot)$ to bound $\mathbb{E}[\int_0^T X_n^2(s) ds]$.

$X_n(\cdot)$ has 2^n parts. It suffices to find $Y^2(\cdot)$ such that

$$\mathbb{E}[X_*^2(s; t_k)] \leq Y^2(s), \quad s \in [t_k, t_{k+1}],$$

where $t_k = \frac{kT}{2^n}$.

The following Lemma 1 gives such $Y^2(\cdot)$ explicitly.

Lemma 1 *Let $Y^2(\cdot)$ satisfying the following ODE*

$$\begin{cases} dY^2(s) = \{R^* + (\gamma^*)^2 e^{-2\int_s^T r(v)dv} \rho(s)\} Y^2(s) ds, \\ Y^2(0) = x_0^2, \end{cases} \quad (2.15)$$

where

$$R^* := \max_{0 \leq s \leq T} \{|2r(s) - \rho(s)|\}, \quad \gamma^* = \max_{0 \leq s \leq T} \{\gamma^*(s)\}.$$

Then, we have, for every $k = 0, 1, \dots, 2^n - 1$,

$$\mathbb{E}[X_*^2(s; t_k)] \leq Y^2(s), \quad s \in [t_k, t_{k+1}].$$

Proof By the uniform boundedness of $B(t)$ and the assumption that $\sigma(t)\sigma(t)^\top \geq \delta I$ for $\delta > 0$, we obtain $\max_{t \in [0, T]} \rho(t) = \max_{t \in [0, T]} B(t)[\sigma(t)\sigma(t)^\top]^{-1} B(t)^\top < \infty$. Together with the uniform boundedness of $r(\cdot)$, it is clear that $R^* < \infty$. Recall that

$$\gamma^*(t) := \frac{f(t, T) - e^{\int_t^T r(v) - \rho(v) dv}}{1 - e^{\int_t^T \rho(v) dv}},$$

which is continuous in $t \in [0, T)$. By L'Hospital's rule and the uniform boundedness of $r(t)$ and $\rho(t)$, we have

$$\gamma^*(t) \rightarrow \frac{\frac{\partial f}{\partial t}(t, T)|_{t=T} + \rho(T) - r(T)}{\rho(T)} < \infty, \quad \text{as } t \rightarrow T.$$

Define $\gamma^*(T) := \frac{\frac{\partial f}{\partial t}(t, T)|_{t=T} + \rho(T) - r(T)}{\rho(T)}$. Then, $\gamma^*(\cdot)$ is continuous on $[0, T]$ with $\gamma^*(T) < \infty$.

We use $X_*(s; t, x_t)$, $s \in [t, T]$, to denote the process satisfying SDE (2.10) with the initial value x_t at t , where x_t is an integrable \mathcal{F}_t -measurable random variable. For $t = t_k$, $k = 0, 1, \dots, 2^n - 1$ and $s \in [t_k, t_{k+1}]$, we consider $\mathbb{E}[X_*^2(s; t, x_t) | \mathcal{F}_t]$. By taking

the conditional expectation on both sides of the SDE for $X_*^2(s; t, x_t)$, we obtain the ODE

$$\begin{cases} d\mathbb{E}[X_*^2(s; t, x_t)|\mathcal{F}_t] = \{(2r(s) - \rho(s))\mathbb{E}[X_*^2(s; t, x_t)|\mathcal{F}_t] + (\gamma^*(t))^2 e^{-2\int_s^T r(v)dv} x_t^2 \rho(s)\} ds, \\ \mathbb{E}[X_*^2(t; t, x_t)|\mathcal{F}_t] = x_t^2. \end{cases} \quad (2.16)$$

Consider a new stochastic process $Z(s; t, x_t)$ satisfying the ODE:

$$\begin{cases} dZ(s; t, x_t) = \{R^* Z(s; t, x_t) + (\gamma^*(t))^2 e^{-2\int_s^T r(v)dv} x_t^2 \rho(s)\} ds, \\ Z(t; t, x_t) = x_t^2. \end{cases} \quad (2.17)$$

Denote $H(s) := (\gamma^*(t))^2 e^{-2\int_s^T r(v)dv} \rho(s) > 0$. By solving (2.16) and (2.17), we can obtain $\forall s \in [t, T]$ and $\omega \in \Omega$,

$$\begin{aligned} Z(s; t, x_t)(\omega) - \mathbb{E}[X_*^2(s; t, x_t)|\mathcal{F}_t](\omega) &= \int_t^s [e^{R^*(s-v)} - e^{\int_v^s (2r(l) - \rho(l))dl}] H(v)(\omega) x_t^2(\omega) dv \\ &\quad + [e^{R^*(s-t)} - e^{\int_t^s (2r(l) - \rho(l))dl}] x_t^2(\omega). \end{aligned} \quad (2.18)$$

By the definition of R^* we obtain $|2r(l) - \rho(l)| \leq R^*$, $l \in [0, T]$. Therefore, we conclude $\forall \omega \in \Omega$

$$\mathbb{E}[X_*^2(s; t, x_t)|\mathcal{F}_t](\omega) \leq Z(s; t, x_t)(\omega). \quad (2.19)$$

Now, we construct the stochastic process $Y^2(s; t, x_t)$ that follows:

$$\begin{cases} dY^2(s; t, x_t) = \{R^* + (\gamma^*)^2 e^{-2\int_s^T r(v)dv} \rho(s)\} Y^2(s; t) ds, \\ Y^2(t; t, x_t) = x_t^2. \end{cases} \quad (2.20)$$

From (2.17) it is easy to see that $\forall \omega \in \Omega$, $Z(s; t, x_t)(\omega)$ is an increasing function

in s ; thus, $Z(s; t, x_t)(\omega) \geq x_t^2(\omega)$. Then, we get

$$\begin{aligned} \frac{dZ(s; t, x_t)}{ds}(\omega) &= R^* Z(s; t, x_t)(\omega) + (\gamma^*(t))^2 e^{-2 \int_s^T r(v) dv} x_t^2(\omega) \rho(s) \\ &\leq \{R^* + (\gamma^*(t))^2 e^{-2 \int_s^T r(v) dv} \rho(s)\} Z(s; t, x_t)(\omega) \\ &\leq \{R^* + (\gamma^*)^2 e^{-2 \int_s^T r(v) dv} \rho(s)\} Z(s; t, x_t)(\omega). \end{aligned} \quad (2.21)$$

It follows from the Gronwall inequality that $\forall s \in [t, T], \forall \omega \in \Omega$,

$$Z(s; t, x_t)(\omega) \leq Y^2(s; t, x_t)(\omega). \quad (2.22)$$

To finish the proof of this lemma we use mathematical induction. When $k = 0$, $s \in [0, t_1]$, it follows from (2.19) and (2.22) that

$$\mathbb{E}[X_*^2(s; 0)] = \mathbb{E}[\mathbb{E}[X^2(s; 0, x_0) | \mathcal{F}_0]] \leq \mathbb{E}[Z(s; 0, x_0)] \leq \mathbb{E}[Y^2(s; 0, x_0)] = Y^2(s). \quad (2.23)$$

Now, assume that when $k = m - 1$, the following holds:

$$\mathbb{E}[X_*^2(s; t_{m-1})] \leq Y^2(s), \quad (2.24)$$

where the initial value of $X_*(s; t_m)$ is $X_*(t_m; t_{m-1})$. By (2.19) and (2.22) we obtain

$$\begin{aligned} \mathbb{E}[X_*^2(s; t_m)] &= \mathbb{E}[\mathbb{E}[X_*^2(s; t_m, X_*(t_m; t_{m-1})) | \mathcal{F}_{t_m}]] \\ &\leq \mathbb{E}[Z(s; t_m, X_*(t_m; t_{m-1}))] \\ &\leq \mathbb{E}[Y^2(s; t_m, X_*(t_m; t_{m-1}))], \end{aligned} \quad (2.25)$$

where the initial value of $\mathbb{E}[Y^2(s; t_m, X_*(t_m; t_{m-1}))]$ is $\mathbb{E}[X_*^2(t_m; t_{m-1})]$. By (2.24) we obtain $\mathbb{E}[X_*^2(t_m; t_{m-1})] \leq Y^2(t_m)$. Note that when $s \in [t_m, t_{m+1}]$, $\mathbb{E}[Y^2(s; t_m, X_*(t_m; t_{m-1}))]$ and $Y^2(s)$ satisfy the same ODE. Combined with (2.25), we conclude that, for

$s \in [t_m, t_{m+1}]$,

$$\mathbb{E}[X_*^2(s; t_m)] \leq \mathbb{E}[Y^2(s; t_m, X_*(t_m; t_{m-1}))] \leq Y^2(s). \quad (2.26)$$

■

Now, we continue proving Proposition 1.

By Lemma 1, we have,

$$\begin{aligned} \|X_n\|^2 &:= \mathbb{E}\left[\int_0^T X_n^2(s) ds\right] \\ &= \mathbb{E}\left[\sum_{k=1}^{2^n} \int_{\frac{k-1}{2^n}T}^{\frac{k}{2^n}T} X_*^2\left(s; \frac{k-1}{2^n}T\right) ds\right] \\ &= \sum_{k=1}^{2^n} \int_{t_{k-1}}^{t_k} \mathbb{E}[X_*^2(s; t_{k-1})] ds \\ &= \sum_{k=1}^{2^n} \int_{t_{k-1}}^{t_k} Y^2(s) ds \\ &= \int_0^T Y^2(s) ds < \infty. \end{aligned} \quad (2.27)$$

In the above, expectation, summation, and integration operators are interchangeable because of Tonelli's theorem. ■

This 2^n -committed agent behaves in a manner in between that of the precommitted agent and the naive one. The behavior is closer to the latter when n becomes larger. Due to Proposition 1, the sequence $\{X_n\}_{n=1}^\infty$ is uniformly bounded in the space $L^2_{\mathcal{F}}([0, T]; \mathbb{R})$, and thus it is weakly compact. By the property of the Hilbert space there exists a weakly convergent subsequence (also denoted as $\{X_n\}_{n=1}^\infty$ without loss of generality) and a process $X \in L^2_{\mathcal{F}}([0, T]; \mathbb{R})$ such that

$$X_n \rightarrow X, \text{ weakly in } L^2_{\mathcal{F}}([0, T]; \mathbb{R}).$$

A natural question is: What is the limiting wealth process X when $n \rightarrow \infty$? We have the following theorem to answer this question.

th

Theorem 1 *If assumptions **A1**), **A2**) and **A3**) are satisfied, the limiting wealth process X satisfies the following SDE:*

$$\begin{cases} dX(t) = [(r(t) - \rho(t)) + e^{-\int_t^T r(v)dv} \rho(t) \gamma^*(t)] X(t) dt \\ \quad + [B(t)(\sigma(t)\sigma(t)^\top)^{-1} \sigma(t) e^{-\int_t^T r(v)dv} \gamma^*(t) - B(t)(\sigma(t)\sigma(t)^\top)^{-1} \sigma(t)] X(t) dW(t), \\ X(0) = x_0, \end{cases} \quad (2.28)$$

where

$$\gamma^*(t) := \frac{f(t, T) - e^{\int_t^T r(v) - \rho(v) dv}}{1 - e^{\int_t^T \rho(v) dv}}.$$

Furthermore, the following feedback portfolio

$$\pi(t, X(t)) = [\sigma(t)\sigma(t)^\top]^{-1} B(t)^\top (\gamma^*(t) e^{-\int_t^T r(s) ds} - 1) X(t)$$

generates X as its wealth process.

Proof To simplify the notation in the rest of the proof we use the following notation

$$\begin{cases} A(s) := (r(s) - \rho(s)), \\ C(s) := e^{-\int_s^T r(v)dv} \rho(s), \\ D(s) := B(s)(\sigma(s)\sigma(s)^\top)^{-1} \sigma(s) e^{-\int_s^T r(v)dv}, \\ F(s) := B(s)(\sigma(s)\sigma(s)^\top)^{-1} \sigma(s), \end{cases} \quad (2.29)$$

with which we rewrite the SDE of wealth process as

$$\begin{cases} dX_*(s; t_k) = \{A(s)X_*(s; t_k) + \gamma^*(t_k)C(s)X_*(t_k; t_{k-1})\} ds \\ \quad + \gamma^*(t_k)D(s)X_*(t_k; t_{k-1}) - F(s)X_*(s; t_k)dW(s), \\ X_*(t_k; t_k) = X_*(t_k; t_{k-1}). \end{cases} \quad (2.30)$$

Due to the boundedness assumptions **A1**) and **A2**), $A(\cdot)$ and $C(\cdot)$ are uniformly bounded. We denote their maximum value over $[0, T]$ as A^* and C^* , respectively. Since $D(t)$ and $F(t)$ are vectors, $\forall t \in [0, T]$, we define $D^* := \max_{t \in [0, T]} \|D(t)\|^2$ and $F^* := \max_{t \in [0, T]} \|F(t)\|^2$.

In order to prove Theorem 1, we need the following lemma.

Lemma 2 *Let $X_*(t_0; t_0) = x_0$ and assume that assumptions **A1**), **A2**) and **A3**) hold. Then, we have*

$$\lim_{n \rightarrow \infty} \max_{k \in \{0, \dots, 2^n - 1\}, s \in [t_k, t_{k+1}]} \mathbb{E}[X_n(s) - X_n(t_k)]^2 = 0.$$

Proof We can bound the term $\mathbb{E}[X_n(s) - X_n(t_k)]^2$ as follows:

$$\begin{aligned} \mathbb{E}[(X_n(s) - X_n(t_k))^2] &= \mathbb{E}[(X_*(s; t_k) - X_*(t_k, t_{k-1}))^2] \\ &= \mathbb{E}\left[\left(\int_{t_k}^s A(t)X_*(t; t_k) + \gamma^*(t_k)C(t)X_*(t_k; t_{k-1})dt + \int_{t_k}^s \gamma^*(t_k)D(t)X_*(t_k; t_{k-1}) - F(t)X_*(t; t_k)dW(t)\right)^2\right] \\ &\leq 2\mathbb{E}\left[\left(\int_{t_k}^s A(t)X_*(t; t_k) + \gamma^*(t_k)C(t)X_*(t_k; t_{k-1})dt\right)^2\right] \\ &\quad + 2\mathbb{E}\left[\left(\int_{t_k}^s \gamma^*(t_k)D(t)X_*(t_k; t_{k-1}) - F(t)X_*(t; t_k)dW(t)\right)^2\right]. \end{aligned} \tag{2.31}$$

Next, we bound the two terms on the right side of last inequality. For the first term we have

$$\begin{aligned} &\mathbb{E}\left[\left(\int_{t_k}^s A(t)X_*(t; t_k) + \gamma^*(t_k)C(t)X_*(t_k; t_{k-1})dt\right)^2\right] \\ &\leq \int_{t_k}^s \mathbb{E}[(A(t)X_*(t; t_k) + \gamma^*(t_k)C(t)X_*(t_k; t_{k-1}))^2]dt \\ &\leq \int_{t_k}^s 2\mathbb{E}[(A(t)X_*(t; t_k))^2] + 2\mathbb{E}[(\gamma^*(t_k)C(t)X_*(t_k; t_{k-1}))^2]dt \\ &\leq \int_{t_k}^s 2A^*\mathbb{E}[X_*(t; t_k)^2] + 2\gamma^*C^*\mathbb{E}[X_*(t_k; t_{k-1})^2]dt \\ &\leq \int_{t_k}^s (2A^* + 2\gamma^*C^*)Y^2(T)dt = (2A^* + 2\gamma^*C^*)(s - t_k)Y^2(T), \end{aligned} \tag{2.32}$$

where the last inequality in (2.32) was by Lemma 1 and the fact that $Y^2(s)$, $s \in [0, T]$,

is increasing in s .

By Ito's isometry, for the second term we have

$$\begin{aligned}
 & \mathbb{E}\left[\left(\int_{t_k}^s \gamma^*(t_k)D(t)X_*(t_k; t_{k-1}) - F(t)X_*(t; t_k)dW(t)\right)^2\right] \\
 &= \int_{t_k}^s \mathbb{E}\left\|\gamma^*(t_k)D(t)X_*(t_k; t_{k-1}) - F(t)X_*(t; t_k)\right\|^2 dt \\
 &\leq \int_{t_k}^s 2\gamma^*D^*\mathbb{E}(X_*(t_k; t_{k-1}))^2 + 2F^*\mathbb{E}(X_*(t; t_k))^2 dt \\
 &\leq (2\gamma^*D^* + 2F^*)(s - t_k)Y^2(T),
 \end{aligned} \tag{2.33}$$

Combining the above, we obtain

$$\mathbb{E}[(X_n(s) - X_n(t_k))^2] \leq 4(s - t_k)(A^* + \gamma^*C^* + \gamma^*D^* + F^*)Y^2(T). \tag{2.34}$$

Note that the above inequality holds for any n . Thus,

$$\max_{k \in \{0, \dots, 2^n - 1\}, s \in [t_k, t_{k+1}]} \mathbb{E}[X_n(s) - X_n(t_k)]^2 \leq \frac{4T}{2^n}(A^* + \gamma^*C^* + \gamma^*D^* + F^*)Y^2(T) \rightarrow 0$$

as $n \rightarrow \infty$. ■

By Mazur's lemma, for each integer $n \geq 1$, there exists positive integer $N(n)$ and a convex combination $V_n := \sum_{k=n}^{N(n)} a_k^n X_k$, where $\sum_{k=n}^{N(n)} a_k^n = 1$ and all a_k^n are non-negative, such that

$$V_n \rightarrow X, \text{ strongly in } L^2_{\mathcal{F}}([0, T]; \mathbb{R}). \tag{2.35}$$

By the definition of V_n , it satisfies the SDE

$$\begin{cases} dV_n(t) = [A(t)V_n(t) + C(t)(\gamma^*X)_{n, N(n)}(t)]dt + [D(t)(\gamma^*X)_{n, N(n)}(t) - F(t)V_n(t)]dW(t), \\ V_n(0) = x_0, \end{cases} \tag{2.36}$$

where

$$(\gamma^* X)_{n,N(n)}(t) := \sum_{k=n}^{N(n)} a_k^n [\gamma^*(m_{t,k}) X_k(m_{t,k})], \quad m_{t,k} := \max\left\{\frac{N}{2^k} T \mid N \in \mathbb{N}, t \geq \frac{N}{2^k} T\right\}.$$

Now consider

$$\begin{cases} dZ(t) = [A(t)X(t) + C(t)\gamma^*(t)X(t)]dt + [D(t)\gamma^*(t)X(t) - F(t)X(t)]dW(t), \\ Z(0) = x_0. \end{cases} \quad (2.37)$$

We now prove that

$$\lim_{n \rightarrow \infty} \int_0^T \mathbb{E}[(Z(t) - V_n(t))^2] dt = 0.$$

To this end, we first analyze $Z(t) - V_n(t)$. We have

$$\begin{aligned} Z(t) - V_n(t) &= \int_0^t A(u)(V_n(u) - X(u)) + C(u)[\gamma^*(u)X(u) - (\gamma^* X)_{n,N(n)}(u)] du \\ &+ \int_0^t D(u)[\gamma^*(u)X(u) - (\gamma^* X)_{n,N(n)}(u)] - F(u)(V_n(u) - X(u)) dW(u) \\ &:= Q_{1,n}(t) + Q_{2,n}(t), \end{aligned} \quad (2.38)$$

where

$$\begin{aligned} Q_{1,n}(t) &:= \int_0^t A(u)(V_n(u) - X(u)) + C(u)[\gamma^*(u)X(u) - (\gamma^* X)_{n,N(n)}(u)] du, \\ Q_{2,n}(t) &:= \int_0^t D(u)[\gamma^*(u)X(u) - (\gamma^* X)_{n,N(n)}(u)] - F(u)(V_n(u) - X(u)) dW(u). \end{aligned}$$

As a result,

$$\begin{aligned} \int_0^T \mathbb{E}[(Z(t) - V_n(t))^2] dt &= \int_0^T \mathbb{E}[(Q_{1,n}(t) + Q_{2,n}(t))^2] dt \\ &\leq 2 \int_0^T \mathbb{E}[Q_{1,n}^2(t) + Q_{2,n}^2(t)] dt \\ &= 2 \int_0^T \mathbb{E}[Q_{1,n}^2(t)] dt + 2 \int_0^T \mathbb{E}[Q_{2,n}^2(t)] dt. \end{aligned} \quad (2.39)$$

The above inequality shows that the L^2 norm of $Z(t) - V_n(t)$ can be bounded by the sum of L^2 norms of $Q_{1,n}^2(t)$ and $Q_{2,n}^2(t)$. Next, we analyze $\mathbb{E}[Q_{1,n}^2(t)]$ and $\mathbb{E}[Q_{2,n}^2(t)]$, respectively. By the definition of $Q_{1,n}(t)$ and Jensen's inequality, we obtain

$$\begin{aligned}
 \mathbb{E}[Q_{1,n}^2(t)] &= \mathbb{E}\left[\left(\int_0^t A(u)(V_n(u) - X(u)) + C(u)[\gamma^*(u)X(u) - (\gamma^*X)_{n,N(n)}(u)]du\right)^2\right] \\
 &\leq \mathbb{E}\left[\int_0^t \{A(u)(V_n(u) - X(u)) + C(u)[\gamma^*(u)X(u) - (\gamma^*X)_{n,N(n)}(u)]\}^2 du\right] \\
 &\leq 2\mathbb{E}\left[\int_0^t A(u)^2(V_n(u) - X(u))^2 + C(u)^2[\gamma^*(u)X(u) - (\gamma^*X)_{n,N(n)}(u)]^2 du\right] \\
 &\leq 2\{(A^*)^2 \int_0^t \mathbb{E}[(V_n(u) - X(u))^2]du + (C^*)^2 \int_0^t \mathbb{E}[\gamma^*(u)X(u) - (\gamma^*X)_{n,N(n)}(u)]^2 du\}.
 \end{aligned} \tag{2.40}$$

By the strong convergence of V_n , the first term converges to 0 as $n \rightarrow \infty$. For the second term,

$$\begin{aligned}
 &\int_0^t \mathbb{E}[\gamma^*(u)X(u) - (\gamma^*X)_{n,N(n)}(u)]^2 du \\
 &= \int_0^t \mathbb{E}[\gamma^*(u)X(u) - \gamma^*(u)V_n(u) + \gamma^*(u)V_n(u) - (\gamma^*X)_{n,N(n)}(u)]^2 du \\
 &\leq 2 \int_0^t \mathbb{E}[\gamma^*(u)X(u) - \gamma^*(u)V_n(u)]^2 + \mathbb{E}[\gamma^*(u)V_n(u) - (\gamma^*X)_{n,N(n)}(u)]^2 du \\
 &\leq 2(\gamma^*)^2 \int_0^t \mathbb{E}[X(u) - V_n(u)]^2 + 2 \int_0^t \mathbb{E}\left[\sum_{k=n}^{N(n)} \alpha_k^n (\gamma^*(u)X_k(u) - \gamma^*(m_{u,k})X_k(m_{u,k}))\right]^2 du.
 \end{aligned} \tag{2.41}$$

Now,

$$\begin{aligned}
 & \int_0^t \mathbb{E} \left[\sum_{k=n}^{N(n)} \alpha_k^n (\gamma^*(u) X_k(u) - \gamma^*(m_{u,k}) X_k(m_{u,k})) \right]^2 du \\
 &= \int_0^t \mathbb{E} \left[\sum_{k=n}^{N(n)} \alpha_k^n (\gamma^*(u) X_k(u) - \gamma^*(m_{u,k}) X_k(u) + \gamma^*(m_{u,k}) X_k(u) - \gamma^*(m_{u,k}) X_k(m_{u,k})) \right]^2 du \\
 &= \int_0^t \mathbb{E} \left[\sum_{k=n}^{N(n)} \alpha_k^n (\gamma^*(u) - \gamma^*(m_{u,k})) X_k(u) + \alpha_k^n \gamma^*(m_{u,k}) (X_k(u) - X_k(m_{u,k})) \right]^2 du \\
 &\leq 2 \int_0^t \mathbb{E} \left[\sum_{k=n}^{N(n)} \alpha_k^n (\gamma^*(u) - \gamma^*(m_{u,k})) X_k(u) \right]^2 + \mathbb{E} \left[\sum_{k=n}^{N(n)} \alpha_k^n \gamma^*(m_{u,k}) (X_k(u) - X_k(m_{u,k})) \right]^2 du \\
 &\leq 2 \int_0^t \sum_{k=n}^{N(n)} \alpha_k^n \mathbb{E} [(\gamma^*(u) - \gamma^*(m_{u,k})) X_k(u)]^2 + \sum_{k=n}^{N(n)} \alpha_k^n \mathbb{E} [\gamma^*(m_{u,k}) (X_k(u) - X_k(m_{u,k}))]^2 du,
 \end{aligned} \tag{2.42}$$

where the last inequality is by the definition of a convex function. Because $\gamma^*(\cdot)$ is continuous in $[0, T]$, it is uniformly continuous. There exists a sequence of $\{\delta_n\}_{n=1}^\infty$ ($\lim_{n \rightarrow \infty} \delta_n = 0$) such that if $x, y \in [0, T]$, $|x - y| \leq \frac{1}{2^n} T$, then $|\gamma^*(x) - \gamma^*(y)| \leq \delta_n$. Therefore, we are able to bound the right hand side of (2.42) as follows:

$$\begin{aligned}
 & \int_0^t \mathbb{E} \left[\sum_{k=n}^{N(n)} \alpha_k^n (\gamma^*(u) X_k(u) - \gamma^*(m_{u,k}) X_k(m_{u,k})) \right]^2 du \\
 &\leq 2 \int_0^t \sum_{k=n}^{N(n)} \alpha_k^n \mathbb{E} [(\gamma^*(u) - \gamma^*(m_{u,k})) X_k(u)]^2 + \sum_{k=n}^{N(n)} \alpha_k^n \mathbb{E} [\gamma^*(m_{u,k}) (X_k(u) - X_k(m_{u,k}))]^2 du \\
 &\leq 2 \int_0^t \delta_n^2 \max_{n \leq k \leq N(n)} \mathbb{E} [|X_k(u)|^2] + (\gamma^*)^2 \max_{n \leq k \leq N(n)} \mathbb{E} [|X_k(u) - X_k(m_{u,k})|^2] du \\
 &\leq 2 \int_0^t \delta_n^2 Y^2(u) + (\gamma^*)^2 \frac{4T}{2^n} (A^* + \gamma^* C^* + \gamma^* D^* + F^*) Y^2(T) du \\
 &\leq 2 [\delta_n^2 + (\gamma^*)^2 \frac{4T}{2^n} (A^* + \gamma^* C^* + \gamma^* D^* + F^*)] T Y^2(T),
 \end{aligned} \tag{2.43}$$

where the third inequality is by Proposition 1 and Lemma 2. As a result

$$\lim_{n \rightarrow \infty} \int_0^t \mathbb{E} \left[\sum_{k=n}^{N(n)} \alpha_k^n (\gamma^*(u) X_k(u) - \gamma^*(m_{u,k}) X_k(m_{u,k})) \right]^2 du = 0. \tag{2.44}$$

By equations (2.41) and (2.44) we obtain

$$\lim_{n \rightarrow \infty} \int_0^t \mathbb{E}[\gamma^*(u)X(u) - (\gamma^*X)_{n,N(n)}(u)]^2 du = 0. \quad (2.45)$$

Thus, by noting $V_n \rightarrow X$ in $L^2_{\mathcal{F}}([0, T]; \mathbb{R})$, it follows from (2.40) and (2.45) that

$$\lim_{n \rightarrow \infty} \mathbb{E}[Q_{1,n}^2(t)] = 0. \quad (2.46)$$

Note that the bound of $\mathbb{E}[Q_{1,n}^2(t)]$ above does not depend on t , which implies that the convergence of $\mathbb{E}[Q_{1,n}^2(t)]$ is uniform in $t \in [0, T]$. Thus, the limit and the integral operator can be interchanged due to the dominated convergence theorem:

$$\lim_{n \rightarrow \infty} \int_0^T \mathbb{E}[Q_{1,n}^2(t)] dt = \int_0^T \lim_{n \rightarrow \infty} \mathbb{E}[Q_{1,n}^2(t)] dt = 0. \quad (2.47)$$

We now analyze the second term in (2.39), $\int_0^T \mathbb{E}[Q_{2,n}^2(t)] dt$. We have

$$\begin{aligned} \mathbb{E}[Q_{2,n}^2(t)] &= \mathbb{E}\left[\int_0^t D(u)[\gamma^*(u)X(u) - \gamma^*X_{n,N(n)}(u)] - F(u)(V_n(u) - X(u))dW(u)\right]^2 \\ &\leq 2\mathbb{E}\left[\int_0^t D(u)[\gamma^*(u)X(u) - \gamma^*X_{n,N(n)}(u)]dW(u)\right]^2 \\ &\quad + 2\mathbb{E}\left[\int_0^t F(u)(V_n(u) - X(u))dW(u)\right]^2 \\ &= 2\int_0^t D^2(u)\mathbb{E}[\gamma^*(u)X(u) - \gamma^*X_{n,N(n)}(u)]^2 du + 2\int_0^t F^2(u)\mathbb{E}[(V_n(u) - X(u))^2] du \\ &\leq 2(D^*)^2 \int_0^t \mathbb{E}[\gamma^*(u)X(u) - \gamma^*X_{n,N(n)}(u)]^2 du + 2(F^*)^2 \int_0^t \mathbb{E}[(V_n(u) - X(u))^2] du, \end{aligned} \quad (2.48)$$

where the second equality is due to Ito's isometry and Tonelli's theorem. Using a similar argument to that employed in the analysis of $\mathbb{E}[Q_{1,n}^2(t)]$, from the equation (2.45) and the strong convergence of $V_n(\cdot)$, we obtain

$$\lim_{n \rightarrow \infty} \mathbb{E}[Q_{2,n}^2(t)] = 0, \quad (2.49)$$

where the convergence is uniform in $t \in [0, T]$. Thus, by the dominated convergence theorem, we deduce

$$\lim_{n \rightarrow \infty} \int_0^T \mathbb{E}[Q_n^2(t)^2] dt = \int_0^T \lim_{n \rightarrow \infty} \mathbb{E}[Q_n^2(t)^2] dt = 0. \quad (2.50)$$

By plugging equations (2.47) and (2.49) into (2.39) we obtain

$$\lim_{n \rightarrow \infty} \int_0^T \mathbb{E}[(Z(t) - V_n(t))^2] dt \leq 2 \lim_{n \rightarrow \infty} \int_0^T \mathbb{E}[Q_{1,n}^2(t)] dt + 2 \lim_{n \rightarrow \infty} \int_0^T \mathbb{E}[Q_{2,n}^2(t)] dt = 0. \quad (2.51)$$

Thus, $Z(t)(\omega) = X(t)(\omega)$ except on a zero measure set in the space of $[0, T] \times \Omega$. So, X satisfies the same SDE as Z . Specifically, its SDE is

$$\begin{aligned} dX(t) &= [A(t)X(t) + C(t)\gamma^*(t)X(t)]dt + [D(t)\gamma^*(t)X(t) - F(t)X(t)]dW(t) \\ &= [(r(t) - \rho(t))X(t) + e^{-\int_t^T r(v)dv} \rho(t)\gamma^*(t)X(t)]dt \\ &\quad + [B(t)(\sigma(t)\sigma(t)^\top)^{-1}\sigma(t)e^{-\int_t^T r(v)dv} \gamma^*(t)X(t) - B(t)(\sigma(t)\sigma(t)^\top)^{-1}\sigma(t)X(t)]dW(t) \\ &= [(r(t) - \rho(t)) + e^{-\int_t^T r(v)dv} \rho(t)\gamma^*(t)]X(t)dt \\ &\quad + [B(t)(\sigma(t)\sigma(t)^\top)^{-1}\sigma(t)e^{-\int_t^T r(v)dv} \gamma^*(t) - B(t)(\sigma(t)\sigma(t)^\top)^{-1}\sigma(t)]X(t)dW(t). \end{aligned} \quad (2.52)$$

Thus, we obtain

$$\begin{cases} dX(t) = [(r(t) - \rho(t)) + e^{-\int_t^T r(v)dv} \rho(t)\gamma^*(t)]X(t)dt \\ \quad + [B(t)(\sigma(t)\sigma(t)^\top)^{-1}\sigma(t)e^{-\int_t^T r(v)dv} \gamma^*(t) - B(t)(\sigma(t)\sigma(t)^\top)^{-1}\sigma(t)]X(t)dW(t), \\ X(0) = x_0, \end{cases} \quad (2.53)$$

where

$$\gamma^*(t) := \frac{f(t, T) - e^{\int_t^T r(v) - \rho(v)dv}}{1 - e^{\int_t^T \rho(v)dv}}.$$

The corresponding portfolio process is

$$\pi(t, X(t)) = [\sigma(t)\sigma(t)^\top]^{-1}B(t)^\top(\gamma^*(t)e^{-\int_t^T r(s)ds} - 1)X(t).$$

The proof is complete. ■

2.3 Comparison between the Naive Strategy and Two Equilibrium Strategies

In the MV literature, two equilibrium strategies have been developed. These are the state-dependent equilibrium strategy put forward by Bjork et al. (2014) and the myopic strategy developed by He and Jiang (2017). We refer to these as state-dependent strategy and myopic strategy, respectively, through rest of the chapter for simplicity. In this section, we compare the portfolio allocation of the naive agent with those of the two equilibrium strategies, in a Black-Scholes market.

2.3.1 State-dependent and Myopic Strategies

Let us first review the two equilibrium strategies. At a fixed time $t \in [0, T]$, Bjork et al. (2014) consider the following problem:

$$\begin{aligned} & \max_{\pi} \quad \mathbb{E}[X(T)|\mathcal{F}_t] - \frac{\gamma(x_t)}{2} \text{Var}(X(T)|\mathcal{F}_t), \\ & \text{subject to} \quad \begin{cases} dX(s) = [r(s)X(s) + B^\top(s)\pi(s)]ds + \pi(s)^\top \sigma(s)dW(s), \\ X_t = x_t, \end{cases} \end{aligned} \tag{2.54}$$

where x_t is the wealth at time t . In the objective function, there is a risk-aversion term $\gamma(x_t)$ that depends on the current state x_t . For this reason, this model is called the state-dependent model. In this model, Bjork et al. (2014) consider the reward functional

$$J(t, x_t, \pi) = \mathbb{E}[X(T)|\mathcal{F}_t] - \frac{\gamma(x_t)}{2} \text{Var}(X(T)|\mathcal{F}_t).$$

Next, we give the formal definition of the state-dependent equilibrium strategy. Given a portfolio $\hat{\pi}$, construct a portfolio π_h by

$$\pi_h(s) = \begin{cases} \pi, & \text{for } t \leq s < t + h, \\ \hat{\pi}(s), & \text{for } t + h \leq s \leq T, \end{cases} \quad (2.55)$$

where $\pi \in \mathbb{R}^m$, $h > 0$. We say that $\hat{\pi}$ is a state-dependent equilibrium strategy if

$$\liminf_{h \rightarrow 0} \frac{J(t, x_t, \hat{\pi}) - J(t, x_t, \pi_h)}{h} \geq 0, \quad (2.56)$$

for all $\pi \in \mathbb{R}^m$.

On the other hand, He and Jiang (2017) consider the following problem:

$$\begin{aligned} & \min_u \quad \text{Var}(X(T)|\mathcal{F}_t) \\ & s.t \quad \begin{cases} dX(s) = [r(s)X(s) + B^\top(s)\pi(s)]ds + \pi(s)^\top \sigma(s)dW(s), \\ X(t) = x_t, \\ \mathbb{E}[X(T)|\mathcal{F}_t] \geq L(t, x_t), \end{cases} \end{aligned} \quad (2.57)$$

where $L(t, x_t)$ indicates the target for the expected wealth when the start time is t and the initial wealth is x_t . This is a more general constraint than the expected wealth constraint in model (2.8). Specifically, if we pick $L(t, x_t) = x_t f(t, T)$, then problem (2.57) is the same as (2.8). He and Jiang (2017) denote the wealth process with portfolio π as X^π . Assume that there exists $h_1 \in (0, T - t)$ such that $\mathbb{E}[X^{\pi_h}(T)|\mathcal{F}_t] \geq L(t, x_t)$, $\forall h \in (0, h_1)$, where π_h is constructed in the same manner as in (2.55). We say that the portfolio $\hat{\pi}$ is a myopic equilibrium strategy if $E[X^{\hat{\pi}}(T)|\mathcal{F}_t] \geq L(t, x_t)$ and there exists $h_2 \in (0, h_1)$ such that

$$\text{Var}(X^{\pi_\epsilon}(T)|\mathcal{F}_t) - \text{Var}(X^{\hat{\pi}}(T)|\mathcal{F}_t) \geq 0, \quad \forall \epsilon \in (0, h_2). \quad (2.58)$$

The difference between the problems (2.54) and (2.57) is that the former uses a weighting coefficient $\gamma(x_t)/2$ in its objective function while the latter uses $L(t, x)$ in its constraint. The two problems are related in view of the Lagrange multiplier approach. Essentially, if we choose $\gamma(x_t)$ and $L(t, x)$ carefully, the precommitted optimal portfolios for the two problems are the same. This is further explained in the following proposition.

Proposition 2 *If the equation*

$$\frac{1}{\gamma(x_t)} e^{\int_t^T \rho(z) dz} + x_t e^{\int_t^T r(z) dz} = \frac{L(t, x_t) - e^{\int_t^T [r(z) - \rho(z)] dz} x_t}{1 - e^{-\int_t^T \rho(z) dz}} \quad (2.59)$$

holds, the precommitted strategies for (2.54) and (2.57) are the same.

Proof By (2.6), the precommitted strategy of (2.54) is

$$\bar{\pi}_{psd}(s, X(s)) = [\sigma(s)\sigma(s)^\top]^{-1} B(s)^\top (\bar{\gamma}_{sd} e^{-\int_s^T r(v) dv} - X(s)), \quad s \in [t, T], \quad (2.60)$$

where

$$\bar{\gamma}_{sd} = \frac{1}{\gamma(x_t)} e^{\int_t^T \rho(v) dv} + e^{\int_t^T r(v) dv} x_t.$$

By (2.6) and the end of Section 2.1.2, the precommitted strategy of (2.57) is

$$\bar{\pi}_{pm}(s, X(s)) = [\sigma(s)\sigma(s)^\top]^{-1} B(s)^\top (\bar{\gamma}_m e^{-\int_s^T r(v) dv} - X(s)), \quad s \in [t, T], \quad (2.61)$$

where

$$\bar{\gamma}_m = \frac{L(t, x_t) - e^{\int_t^T r(v) - \rho(v) dv} x_t}{1 - e^{-\int_t^T \rho(v) dv}}.$$

It is clear that, if (2.59) is satisfied, we have $\bar{\gamma}_{sd} = \bar{\gamma}_m$, which leads to $\bar{\pi}_{psd} = \bar{\pi}_{pm}$.

■

By Proposition 2, if $\gamma(x_t)$ and $L(t, x_t)$ satisfy (2.59), then we obtain the same precommitted solutions from the two problems. However, the equilibrium strategies

of the two problems are different; see He and Jiang (2017).

Denote the optimal portfolios of problem (2.54) and (2.57) as $\pi_{sd}(t, x)$ and $\pi_m(t, x)$, respectively, and denote the portfolio of the naive agent as $\pi_{na}(t, x)$.

We now compare the above three strategies by considering the problem, for simplicity, in a Black-Scholes market. Specifically, there is a risk-free asset and only one risky asset in the market. Then, we calculate the allocations in the risky asset and in the risk-free asset, respectively, obtained based on the three strategies. Because there is only one risky asset, the portfolios deriving from all the three strategies are just scalars, which can easily be compared.

We carry out the comparisons for two commonly used cases. In Section 2.3.2, we choose $\gamma(x)$ in problem (2.54) to be $\frac{\gamma}{x}$, which is the case examined by Bjork et al. (2014). Section 2.3.3 study the case in which $L(t, x) = xe^{k(T-t)}$. In each subsection, we choose $L(t, x)$ and $\gamma(x_t)$ to satisfy equation (2.59) so as to be consistent in their respective precommitted solutions.

2.3.2 The case $\gamma(x) = \frac{\gamma}{x}$

By choosing $\gamma(x) = \frac{\gamma}{x}$, we can calculate $L(t, x)$ using equation (2.59) to obtain

$$L(t, x_t) = x_t \left[\frac{1}{\gamma} (e^{(T-t)\rho} - 1 + \gamma e^{(T-t)r}) \right]. \quad (2.62)$$

For the case of $\gamma_t = \frac{\gamma}{x}$, by Theorem 4.6 in Bjork et al. (2014), the equilibrium portfolio of problem (2.54) is given by

$$\pi_{sd}(t, X(t)) = c_{sd}(t)X(t), \quad (2.63)$$

where $c_{sd}(t)$ is

$$c_{sd}(t) = \frac{\alpha - r}{\gamma\sigma^2} \left\{ e^{-\int_t^T [r + (\alpha - r)c(s) + \sigma^2 c^2(s)] ds} + \gamma e^{-\int_t^T \sigma^2 c^2(s) ds} - \gamma \right\}. \quad (2.64)$$

In order to obtain the naive strategy, we need to choose $f(t, T)$ in problem (2.8) such that it is consistent with problems (2.54) and (2.57). It follows from (2.62) that the corresponding $f(t, T)$ in (2.8) is

$$f(t, T) = \frac{1}{\gamma}(e^{(T-t)\rho} - 1 + \gamma e^{(T-t)r}).$$

It is easy to check that this $f(t, T)$ satisfies Assumption **A3**). By Theorem 1, the portfolio of the naive agent is

$$\pi(t, X(t)) = [\sigma(t)\sigma(t)^\top]^{-1}B(t)^\top(\gamma^*(t)e^{-\int_t^T r(s)ds} - 1)X(t),$$

where

$$\gamma^*(t) = \frac{e^{\int_t^T \rho(v)dv}}{\gamma} + e^{\int_t^T r(v)dv}.$$

So, the portfolio is given by

$$\pi_{na}(t, X(t)) = \frac{\alpha - r}{\sigma^2} \left(\frac{e^{\frac{(\alpha-r)^2 - r\sigma^2}{\sigma^2}(T-t)}}{\gamma} \right) X(t) = c_{na}(t)X(t), \quad (2.65)$$

where we denote $c_{na}(t) = \frac{\alpha-r}{\sigma^2} \left(\frac{e^{\frac{(\alpha-r)^2 - r\sigma^2}{\sigma^2}(T-t)}}{\gamma} \right)$.

Finally, by (2.62), we should set $L(t, x_t)$ in problem (2.57) to be

$$L(t, x_t) = x_t \left[\frac{1}{\gamma} (e^{(T-t)\rho} - 1 + \gamma e^{(T-t)r}) \right].$$

By Section 4.3 in He and Jiang (2017), the corresponding equilibrium portfolio is

$$\begin{aligned} \pi_m(t, X(t)) &= -\frac{1}{\alpha - r} \left[\frac{(r - \rho)e^{(T-t)\rho} - r}{e^{(T-t)\rho} - 1 + \gamma e^{(T-t)r}} \right] X(t) \\ &= c_m(t)X(t), \end{aligned} \quad (2.66)$$

where $c_m(t) = -\frac{1}{\alpha-r} \left[\frac{(r-\rho)e^{(T-t)\rho} - r}{e^{(T-t)\rho} - 1 + \gamma e^{(T-t)r}} \right]$. Next, we show that the naive strategy allocates more weight to the risky asset than the two equilibrium strategies by the following

proposition.

Proposition 3 *In the Black-Scholes market, we have*

$$c_{sd}(t) < c_{na}(t), c_m(t) < c_{na}(t)$$

for any $\gamma > 0$.

Proof We first prove $c_{sd}(t) < c_{na}(t)$. We consider three cases: $\rho = \frac{(\alpha-r)^2}{\sigma^2} > r$, $\rho < r$, and $\rho = r$.

If $\rho = \frac{(\alpha-r)^2}{\sigma^2} > r$, then $c_{na}(t)$ is decreasing in t and $c(t)$ is increasing in t . Thus,

$$c_{na}(t) \geq c_{na}(T) = \frac{\alpha - r}{\sigma^2} = c_{sd}(T) \geq c_{sd}(t), \forall t \in [0, T].$$

If $\rho = \frac{(\alpha-r)^2}{\sigma^2} < r$, then $c(t) \geq 0, \forall t \in [0, T]$. Then

$$\begin{aligned} c_{sd}(t) &= \frac{\alpha - r}{\gamma\sigma^2} \left\{ e^{-\int_t^T [r+(\alpha-r)c(s)+\sigma^2c^2(s)]ds} + \gamma e^{-\int_t^T \sigma^2c^2(s)ds} - \gamma \right\} \\ &\leq \frac{\alpha - r}{\gamma\sigma^2} \left\{ e^{-\int_t^T [r+(\alpha-r)c(s)+\sigma^2c^2(s)]ds} + \gamma - \gamma \right\} \\ &= \frac{\alpha - r}{\gamma\sigma^2} \left\{ e^{-\int_t^T [r+(\alpha-r)c(s)+\sigma^2c^2(s)]ds} \right\} \\ &\leq \frac{\alpha - r}{\gamma\sigma^2} \left\{ e^{-r(T-t)} \right\} \\ &\leq \frac{\alpha - r}{\gamma\sigma^2} \left\{ e^{-r(T-t)+\rho(T-t)} \right\} = c_{na}(t), \forall t \in [0, T]. \end{aligned} \tag{2.67}$$

Next, we prove $c_m(t) < c_{na}(t)$. The risky asset weight for the myopic strategy is

$$c_m(t) = -\frac{1}{\alpha - r} \left[\frac{(r - \rho)e^{(T-t)\rho} - r}{e^{(T-t)\rho} - 1 + \gamma e^{(T-t)r}} \right]. \tag{2.68}$$

Because $\rho \geq 0$, we have

$$\begin{aligned}
 c_{na}(t) - c_m(t) &= \frac{1}{\beta} \left[\rho \left(\frac{e^{(\rho-r)(T-t)}}{\gamma} \right) + \frac{(r-\rho)e^{(T-t)\rho} - r}{e^{(T-t)\rho} - 1 + \gamma e^{(T-t)r}} \right] \\
 &= \frac{1}{\beta} \left[\frac{\frac{\rho e^{(2\rho-r)(T-t)}}{\gamma} - \frac{\rho e^{(\rho-r)(T-t)}}{\gamma} + \rho e^{\rho(T-t)} + (r-\rho)e^{(T-t)\rho} - r}{e^{(T-t)\rho} - 1 + \gamma e^{(T-t)r}} \right] \quad (2.69) \\
 &= \frac{1}{\beta} \left[\frac{\frac{\rho e^{(2\rho-r)(T-t)}}{\gamma} - \frac{\rho e^{(\rho-r)(T-t)}}{\gamma} + r e^{(T-t)\rho} - r}{e^{(T-t)\rho} - 1 + \gamma e^{(T-t)r}} \right] \geq 0.
 \end{aligned}$$

The proof is complete. ■

2.3.3 The case $L(t, x) = x e^{k(T-t)}$

In this case, the parameter k should be greater than the risk-free rate r , otherwise, problem (2.57) is trivial. Specifically, if $k \leq r$, the optimal portfolio is putting all of one's money in the bank account. We can calculate $\gamma(x)$ using equation (2.59) to obtain

$$\gamma(x_t) = \frac{1}{x_t} \frac{e^{(T-t)\rho} - 1}{e^{(T-t)k} - e^{(T-t)r}}. \quad (2.70)$$

Let $\phi(t) := \frac{e^{(T-t)\rho} - 1}{e^{(T-t)k} - e^{(T-t)r}}$. By Bjork et al. (2014), the equilibrium portfolio is given by

$$\pi'_{sd}(t, X(t)) = c'_{sd}(t) X(t), \quad (2.71)$$

where $c'_{sd}(t)$ is given as follows:

$$c'_{sd}(t) = \frac{\beta}{\phi(t)\sigma^2} \left\{ e^{-\int_t^T [r + (\alpha-r)c(s) + \sigma^2 c^2(s)] ds} + \phi(t) e^{-\int_t^T \sigma^2 c^2(s) ds} - \phi(t) \right\}. \quad (2.72)$$

By the definition of $L(t, x)$, it is easy to obtain the corresponding $f(t, T)$ in (2.5) as

$$f(t, T) = e^{k(T-t)},$$

which satisfies Assumption **A3**). Thus, by Theorem 1, the naive portfolio is

$$\pi(t, X(t)) = [\sigma(t)\sigma(t)^\top]^{-1}B(t)^\top(\gamma^*(t)e^{-\int_t^T r(s)ds} - 1)X(t),$$

where

$$\gamma^*(t) = \frac{e^{k(T-t)} - e^{(r-\rho)(T-t)}}{1 - e^{-\rho(T-t)}}.$$

Therefore, the portfolio for the naive agent is given by

$$\pi'_{na}(t, X(t)) = \frac{\rho}{\beta} \left(\frac{e^{(k-r)(T-t)} - 1}{1 - e^{-\rho(T-t)}} \right) X(t) = c'_{na}(t)X(t), \quad (2.73)$$

where $c'_{na}(t) := \frac{\rho}{\beta} \left(\frac{e^{(k-r)(T-t)} - 1}{1 - e^{-\rho(T-t)}} \right)$. Finally, by Section 4.3 of He and Jiang (2017), the corresponding equilibrium portfolio is

$$\pi'_m(t, X(t)) = \frac{1}{\beta}(k - r)X(t) = c'_m(t)X(t), \quad (2.74)$$

where $c'_m(t) := \frac{1}{\beta}(k - r)$. Next, we show that the naive strategy allocates more weight to the risky asset than the two equilibrium strategies by the following proposition.

Proposition 4 *In the Black-Scholes market, we have*

$$c'_{sd}(t) < c'_{na}(t), \quad c'_m(t) < c'_{na}(t)$$

for any $\gamma > 0$.

Proof We divide the proof into two parts. In the first part, we prove that $c'_{sd}(t) < c'_{na}(t)$. The portfolio coefficient for the naive strategy is given by

$$c'_{na}(t) = \frac{\rho}{\beta} \left(\frac{e^{(k-r)(T-t)} - 1}{1 - e^{-\rho(T-t)}} \right). \quad (2.75)$$

The coefficient for the state-dependent strategy is

$$c'_{sd}(t) = \frac{\beta}{\phi(t)\sigma^2} \left\{ e^{-\int_t^T [r+(\alpha-r)c'_{sd}(s)+\sigma^2(c'_{sd})^2(s)]ds} + \phi(t)e^{-\int_t^T \sigma^2(c'_{sd})^2(s)ds} - \phi(t) \right\}, \quad (2.76)$$

where $\phi(t) := \frac{e^{(T-t)\rho}-1}{e^{(T-t)k}-e^{(T-t)r}}$. Note that $c'(t) \geq 0$ always holds. Due to the fact that $k > r$, we have

$$\begin{aligned} c'_{sd}(t) &= \frac{\beta}{\phi(t)\sigma^2} \left\{ e^{-\int_t^T [r+(\alpha-r)c(s)+\sigma^2c^2(s)]ds} + \phi(t)e^{-\int_t^T \sigma^2c^2(s)ds} - \phi(t) \right\} \\ &= \frac{\beta}{\phi(t)\sigma^2} \left\{ e^{-\int_t^T [r+(\alpha-r)c(s)+\sigma^2c^2(s)]ds} \right\} \\ &\leq \frac{\beta}{\phi(t)\sigma^2} \left\{ e^{-r(T-t)} \right\} \\ &= \frac{(e^{(T-t)(k-r)} - 1)\beta}{(e^{(T-t)\rho} - 1)\sigma^2} \\ &= \frac{(e^{(T-t)(k-r)} - 1)\rho}{(e^{(T-t)\rho} - 1)\beta} \\ &\leq \frac{\rho}{\beta} \frac{e^{(T-t)(k-r)} - 1}{1 - e^{-(T-t)\rho}} = c_{na}(t), \quad \forall t \in [0, T]. \end{aligned} \quad (2.77)$$

In the second part, we prove that $c_m(t) < c_{na}(t)$. Because $\rho \geq 0$, we have

$$\begin{aligned} c_{na}(t) - c_m(t) &= \frac{1}{\beta} \left[\frac{\rho e^{(k-r)(T-t)} - \rho - (k-r) + (k-r)e^{-\rho(T-t)}}{1 - e^{-\rho(T-t)}} \right] \\ &= \frac{1}{\beta} \left[\frac{\rho e^{(k-r)(T-t)} + (k-r)e^{-\rho(T-t)} - \rho - (k-r)}{1 - e^{-\rho(T-t)}} \right] \geq 0. \end{aligned} \quad (2.78)$$

The proof is complete. ■

2.4 Conclusions

We consider a naive agent in a continuous-time mean-variance model. By partitioning the time line into 2^n parts of equal length, we are able to construct the portfolio and wealth processes for the 2^n -committed agent using the precommitted strategy of each time interval. Next, we use the behavior of the 2^n -committed agent to approximate that of the naive agent. Under some boundedness assumptions, we

are able to prove that the wealth process of the 2^n -committed agent will converge to a limiting process as $n \rightarrow \infty$, where this limiting process is the wealth process for the naive agent. We also derive the wealth and portfolio processes explicitly. After that, we compare the derived naive strategy to the two equilibrium strategies developed by Bjork et al. (2014) and by He and Jiang (2017). We derive the allocations of all three agents in the Black-Scholes market, where there is only one risky asset and one risk-free asset. We prove that the naive agent will allocate more weight to the risky-asset than that of the two equilibrium strategies. These comparisons show that the agent in the naive strategy tends to be more “greedy” than the equilibrium strategies.

Chapter 3

A Distributionally Robust Mean-Variance Model

3.1 Model Formulation

In this section, we formulate a distributionally robust Markowitz (DRM) model while reviewing some useful concepts.

Let $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ be the space of all Borel probability measures supported on $\mathbb{R}^d \times \mathbb{R}^d$. A given element $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ can be assumed to be the joint distribution of a random vector (U, V) , where $U \in \mathbb{R}^d$ and $V \in \mathbb{R}^d$. We use π_U and π_V to denote the marginal distributions of U and V under π . In particular, $\pi_U(A) = \pi(A \times \mathbb{R}^d)$ and $\pi_V(A) = \pi(\mathbb{R}^d \times A)$ for every Borel set $A \subset \mathbb{R}^d$.

We start with a “cost” function $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, \infty]$, which we assume to be lower semicontinuous and such that $c(u, u) = 0$ for any $u \in \mathbb{R}^d$. For a given such cost function c , we introduce $D_c(\cdot, \cdot)$ representing some “discrepancy” between two probability measures, as follows:

$$D_c(\mathbb{P}, \mathbb{Q}) := \inf\{\mathbb{E}_\pi[c(U, V)] : \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), \pi_U = \mathbb{P}, \pi_V = \mathbb{Q}\}, \quad (3.1)$$

where \mathbb{P} and \mathbb{Q} are two probability measures supported on \mathbb{R}^d . This can be interpreted

as the optimal (minimal) transportation cost (also known as the optimal transport discrepancy or the Wasserstein discrepancy) of moving the mass from \mathbb{P} into the mass of \mathbb{Q} under a cost $c(x, y)$ per unit of mass transported from x to y .

If for a given $p > 0$, $c^{1/p}(\cdot, \cdot)$ is a metric, then so is $D_c^{1/p}(\cdot, \cdot)$; see Villani (2003). Such a metric $D_c^{1/p}(\cdot, \cdot)$ is known as a Wasserstein distance of the order p . Most of the times in this chapter, we choose the following cost function

$$c(u, v) = \|u - v\|_q^2, \quad (3.2)$$

where $q \geq 1$ is fixed (which leads to a Wasserstein distance of the order 2).¹

Recall that R is the d -dimensional vector of the random returns of d stocks. Let \mathbb{P}_n be the empirical probability measure on \mathbb{R}^d with a sample size n , i.e.,

$$\mathbb{P}_n(dx) = \frac{1}{n} \sum_{i=1}^n \delta_{R_i}(dx),$$

where R_i ($i = 1, 2, \dots, n$) are realizations of R and $\delta_{R_i}(\cdot)$ is the indicator function.

Define the ambiguity set as

$$\mathcal{U}_\delta(\mathbb{P}_n) = \{\mathbb{P} : D_c(\mathbb{P}, \mathbb{P}_n) \leq \delta\}$$

and the feasible region of portfolios as

$$\mathcal{F}_{\delta, \bar{\alpha}}(n) = \{\phi \in \mathbb{R}^d : \phi^\top \mathbf{1} = 1, \min_{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)} [\mathbb{E}_\mathbb{P}(\phi^\top R)] \geq \bar{\alpha}\}.$$

The first constraint in the feasible region $\mathcal{F}_{\delta, \bar{\alpha}}(n)$ is the budget constraint, and the second constraint represents the *worst case target* for our portfolios when the underlying probability measure is selected from the ambiguity set $\mathcal{U}_\delta(\mathbb{P}_n)$. The DRMV

¹Different cost functions can be used, resulting in different regularization penalties, we discuss this at the end of Section 3.2 and in Section 3.5.

model is formulated as follows:

$$\min_{\phi \in \mathcal{F}_{\delta, \bar{\alpha}}(n)} \max_{\mathbb{P} \in \mathcal{U}_{\delta}(\mathbb{P}_n)} \{ \phi^\top \text{Var}_{\mathbb{P}}(R) \phi \}. \quad (3.3)$$

The inner maximization part of (3.3) represents the worst case portfolio variance. The objective of the DRMV model is to choose a portfolio $\phi \in \mathcal{F}_{\delta, \bar{\alpha}}(n)$ that achieves the optimal worst case value in (3.3).

3.2 Transformations, Duality, and Regularization

Problem (3.3) appears, in principle, very complex. First of all, the inner maximization problem is over a set of probability measures, which renders it an infinite dimensional optimization problem. Second, it is not clear whether the outer minimization problem, while finite dimensional, is convex. Therefore (3.3) at its outset seems computationally insurmountable. In this section, we reformulate (3.3), through a series of transformations and a duality argument, as an equivalent problem that is computationally tractable.

The first step is to show that the feasible region over ϕ in the outer minimization part can be explicitly evaluated. This is given in the following proposition:

Proposition 5 For $c(u, v) = \|u - v\|_q^2$, $q \geq 1$, we have

$$\min_{\mathbb{P} \in \mathcal{U}_{\delta}(\mathbb{P}_n)} \mathbb{E}_{\mathbb{P}}(\phi^\top R) = \mathbb{E}_{\mathbb{P}_n}(\phi^\top R) - \sqrt{\delta} \|\phi\|_p, \quad (3.4)$$

where p satisfies $1/p + 1/q = 1$.

Proof We consider the following problem:

$$\min_{\mathbb{P} \in \mathcal{D}_c(\mathbb{P}, \mathbb{P}_n) \leq \delta} \phi^\top \mathbb{E}_{\mathbb{P}}[R] \quad (3.5)$$

or, equivalently,

$$- \max_{\mathbb{P} \in D_c(\mathbb{P}, \mathbb{P}_n) \leq \delta} \mathbb{E}_{\mathbb{P}}[(-\phi)^\top R]. \quad (3.6)$$

By checking Slater's condition and using Proposition 4 of Blanchet et al. (2016), we obtain the dual problem:

$$\max_{\mathbb{P} \in D_c(\mathbb{P}, \mathbb{P}_n) \leq \delta} \mathbb{E}_{\mathbb{P}}[(-\phi)^\top R] = \inf_{\lambda \geq 0} \left[\lambda \delta + \frac{1}{n} \sum_{i=1}^n \Phi_\lambda(R_i) \right] \quad (3.7)$$

where

$$\begin{aligned} \Phi_\lambda(R_i) &= \sup_u \{h(u) - \lambda c(u, R_i)\} \\ &= \sup_u \{(-\phi^\top)u - \lambda \|u - R_i\|_q^2\} \\ &= \sup_\Delta \{(-\phi^\top)(\Delta + R_i) - \lambda \|\Delta\|_q^2\} \\ &= \sup_\Delta \{(-\phi^\top)\Delta - \lambda \|\Delta\|_q^2\} - \phi^\top R_i \\ &= \sup_\Delta \{ \|\phi\|_p \|\Delta\|_q - \lambda \|\Delta\|_q^2 \} - \phi^\top R_i \\ &= \frac{\|\phi\|_p^2}{4\lambda} - \phi^\top R_i. \end{aligned}$$

Thus, (3.7) becomes

$$\begin{aligned} \max_{\mathbb{P} \in D_c(\mathbb{P}, \mathbb{P}_n) \leq \delta} \mathbb{E}_{\mathbb{P}}[(-\phi)^\top R] &= \inf_{\lambda \geq 0} \left\{ \lambda \delta + \frac{1}{n} \sum_{i=1}^n \left[\frac{\|\phi\|_p^2}{4\lambda} - \phi^\top R_i \right] \right\} \\ &= \inf_{\lambda \geq 0} \left\{ \lambda \delta + \frac{\|\phi\|_p^2}{4\lambda} - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] \right\} \\ &= \sqrt{\delta} \|\phi\|_p - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] \end{aligned}$$

or

$$\min_{\mathbb{P} \in D_c(\mathbb{P}, \mathbb{P}_n) \leq \delta} \phi^\top \mathbb{E}_{\mathbb{P}}[R] = \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] - \sqrt{\delta} \|\phi\|_p. \quad (3.8)$$

■

Therefore, the feasible region can be rewritten as

$$\mathcal{F}_{\delta, \bar{\alpha}}(n) = \{\phi \in \mathbb{R}^d : \phi^\top \mathbf{1} = 1, \mathbb{E}_{\mathbb{P}_n}(\phi^\top R) \geq \bar{\alpha} + \sqrt{\delta} \|\phi\|_p\},$$

which can now be seen as clearly convex. Note that, using the same approach in the proof of Proposition 5, we can prove

$$\max_{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)} \mathbb{E}_{\mathbb{P}}(\phi^\top R) = \mathbb{E}_{\mathbb{P}_n}(\phi^\top R) + \sqrt{\delta} \|\phi\|_p. \quad (3.9)$$

Although it is difficult to characterize the worst case probability distribution, equations (3.4) and (3.9) have given us an intuitive sense about what effects it will bring in our model. To be more specific, under the worst case probability distribution, the portfolio return will deviate from the its empirical estimation with the size of $\sqrt{\delta} \|\phi\|_p$.

Next, by fixing $\mathbb{E}_{\mathbb{P}}(\phi^\top R) = \alpha \geq \bar{\alpha}$ in the inner maximization part of problem (3.3), we obtain the following equivalent formulation

$$\min_{\phi \in \mathcal{F}_{\delta, \bar{\alpha}}} \left\{ \max_{\alpha \geq \bar{\alpha}} \left[\max_{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n), \mathbb{E}_{\mathbb{P}}(\phi^\top R) = \alpha} \{\phi^\top \mathbb{E}_{\mathbb{P}}(RR^\top) \phi\} - \alpha^2 \right] \right\}. \quad (3.10)$$

Introducing $\mathbb{E}_{\mathbb{P}}(\phi^\top R) = \alpha$ is useful because the inner-most maximization problem in the above is now linear in \mathbb{P} . So, let us concentrate on the problem

$$\max_{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n), \mathbb{E}_{\mathbb{P}}(\phi^\top R) = \alpha} \phi^\top \mathbb{E}_{\mathbb{P}}(RR^\top) \phi. \quad (3.11)$$

The following proposition solves this problem in terms of a general cost function c .

Proposition 6 *For any cost function c that is lower semicontinuous and non-negative, the optimal value function of problem (3.11) is given by*

$$\inf_{\lambda_1 \geq 0, \lambda_2} \left[\frac{1}{n} \sum_{i=1}^n \Phi(R_i) + \lambda_1 \delta + \lambda_2 \alpha \right], \quad (3.12)$$

where

$$\Phi(R_i) := \sup_u [(\phi^\top u)^2 - \lambda_1 c(u, R_i) - \lambda_2 \phi^\top u].$$

Proof The proof is based on a duality argument. Introduce a slack random variable $S \equiv v$, where v is a deterministic number. Then, we can recast problem (3.11) as

$$\max\{\mathbb{E}_{\mathbb{P}}[(U^\top \phi)^2] : \mathbb{E}_{\pi}[c(U, R) + S] = \delta, \pi_R = \mathbb{P}_n, \pi(S = v) = 1, \quad (3.13)$$

$$\mathbb{E}_{\pi}[U^\top \phi] = \alpha, \pi \in \mathcal{P}(\mathcal{R}^m \times \mathcal{R}^m \times \mathcal{R}_+)\}. \quad (3.14)$$

Define

$$\Omega := \{(u, r, s) : c(u, r) < \infty, s \geq 0, r \in \{R_1, \dots, R_n\}\}$$

and let

$$f(u, r, s) = \begin{bmatrix} 1_{r=R_1}(u, r, s) \\ \dots \\ 1_{r=R_n}(u, r, s) \\ \phi^\top u \\ 1_{s=v}(u, r, s) \\ c(u, r) + s \end{bmatrix} \quad \text{and} \quad q = \begin{bmatrix} \frac{1}{n} \\ \dots \\ \frac{1}{n} \\ \alpha \\ 1 \\ \delta \end{bmatrix}. \quad (3.15)$$

Thus, (3.13) can be written as

$$\max\{\mathbb{E}_{\pi}[(U^\top \phi)^2] : \mathbb{E}_{\pi}[f(U, R, S)] = q, \pi \in \mathcal{P}_{\Omega}\}. \quad (3.16)$$

Let $f_0 = \mathbf{1}_{\Omega}$, $\tilde{f} = (f_0, f)$, $\tilde{q} = (1, q)$, $\mathcal{Q}_{\tilde{f}} := \{\int \tilde{f}(x) d\mu(x) : \mu \in \mathcal{M}_{\Omega}^+\}$ where \mathcal{M}_{Ω}^+ denotes the set of non-negative measures on Ω . If $\phi \neq 0$, then it is easy to see that \tilde{q} lies in the interior of $\mathcal{Q}_{\tilde{f}}$. By Proposition 6 in Blanchet et al. (2016), the optimal

value of problem (3.16) equals that of its dual problem, i.e.,

$$\begin{aligned} \max\{\mathbb{E}_\pi[(U^\top \phi)^2] : \mathbb{E}_\pi[f(U, R, S)] = q, \pi \in \mathcal{P}_\Omega\} \\ = \inf_{a=(a_0, \dots, a_{n+3}) \in A} \left\{ a_0 + \frac{1}{n} \sum_{i=1}^n a_i + \alpha a_{n+1} + a_{n+2} + \delta a_{n+3} \right\}, \end{aligned} \quad (3.17)$$

where

$$\begin{aligned} A := \{a = (a_0, \dots, a_{n+3}) \in \mathbb{R}^{n+4} : a_0 + \frac{1}{n} \sum_{i=1}^n a_i \mathbf{1}_{r=R_i}(u, r, s) + a_{n+1} \phi^\top u \\ + a_{n+2} \mathbf{1}_{s=v}(u, r, s) + a_{n+3} [c(u, r) + s] \geq (\phi^\top u)^2, \forall (u, r, s) \in \Omega\}. \end{aligned}$$

From the definition of A , replacing $r = R_i$, we obtain that the inequality

$$a_0 + a_i + a_{n+2} \geq \sup_{(u,s) \in \Omega} \{(\phi^\top u)^2 - a_{n+3} [c(u, R_i) + s] - a_{n+1} \phi^\top u\} \quad (3.18)$$

holds for each $i \in \{1, \dots, n\}$. It follows directly that

$$\sup_{(u,s) \in \Omega} \{(\phi^\top u)^2 - a_{n+3} [c(u, R_i) + s] - a_{n+1} \phi^\top u\} \quad (3.19)$$

$$= \begin{cases} +\infty, & \text{if } a_{n+3} < 0 \\ \sup_u \{(\phi^\top u)^2 - a_{n+3} c(u, R_i) - a_{n+1} \phi^\top u\}, & \text{if } a_{n+3} \geq 0. \end{cases} \quad (3.20)$$

Thus, the dual problem can be expressed as

$$\inf \left\{ a_0 + \frac{1}{n} \sum_{i=1}^n a_i + \alpha a_{n+1} + a_{n+2} + \delta a_{n+3} : a_{n+3} \geq 0, a_0 + a_i + a_{n+2} \geq \sup_u \{(\phi^\top u)^2 - a_{n+3} c(u, R_i) - a_{n+1} \phi^\top u\} \right\}, \quad (3.21)$$

which can be transformed into

$$\inf_{a_{n+3} \geq 0} \left\{ \frac{1}{n} \sum_{i=1}^n \Phi(R_i) + \alpha a_{n+1} + \delta a_{n+3} \right\}, \quad (3.22)$$

with

$$\Phi(R_i) := \sup_u \{(\phi^\top u)^2 - a_{n+3}c(u, R_i) - a_{n+1}\phi^\top u\}.$$

Using λ_1 to replace a_{n+3} and λ_2 to replace a_{n+1} , the dual problem becomes

$$\inf_{\lambda_1 \geq 0} \left\{ \frac{1}{n} \sum_{i=1}^n \Phi(R_i) + \lambda_2 \alpha + \lambda_1 \delta \right\}, \quad (3.23)$$

where

$$\Phi(R_i) := \sup_u \{(\phi^\top u)^2 - \lambda_1 c(u, R_i) - \lambda_2 \phi^\top u\}.$$

■

Thanks to this proposition, we are able to reduce the inner (infinite dimensional) optimization problem in (3.3) to a two-dimensional optimization problem in terms of λ_1 and λ_2 , which can be further simplified if the cost function c has additional structure. We make this statement precise in the case of a quadratic l_q cost.

Proposition 7 *Let $c(u, v) = \|u - v\|_q^2$ with $q \geq 1$ and $1/p + 1/q = 1$. If $(\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2 - \delta \|\phi\|_p^2 \leq 0$, then the value of (3.11) is equal to*

$$\begin{aligned} h(\alpha, \phi) := & \mathbb{E}_{\mathbb{P}_n} [(\phi^\top R)^2] + 2(\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] + \delta \|\phi\|_p^2 \\ & + 2\sqrt{\delta \|\phi\|_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2} \sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R) \phi}. \end{aligned}$$

Proof Writing $\Delta := u - R_i$, we have

$$\begin{aligned}
\Phi(R_i) &= \sup_u \{(\phi^\top u)^2 - \lambda_1 c(u, R_i) - \lambda_2 \phi^\top u\} \\
&= \sup_u \{(\phi^\top u)^2 - \lambda_1 \|u - R_i\|_q^2 - \lambda_2 \phi^\top u\} \\
&= \sup_\Delta \{(\phi^\top (\Delta + R_i))^2 - \lambda_1 \|\Delta\|_q^2 - \lambda_2 \phi^\top (R_i + \Delta)\} \\
&= \sup_\Delta \{(\phi^\top R_i)^2 + (\phi^\top \Delta)^2 + 2(\phi^\top R_i)(\phi^\top \Delta) - \lambda_1 \|\Delta\|_q^2 - \lambda_2 \phi^\top (R_i + \Delta)\} \\
&= (\phi^\top R_i)^2 - \lambda_2 \phi^\top R_i + \sup_\Delta \{(\phi^\top \Delta)^2 + 2(\phi^\top R_i)(\phi^\top \Delta) - \lambda_1 \|\Delta\|_q^2 - \lambda_2 \phi^\top \Delta\} \\
&= (\phi^\top R_i)^2 - \lambda_2 \phi^\top R_i + \sup_\Delta \{(\|\phi\|_p^2 - \lambda_1) \|\Delta\|_q^2 + |2(R_i^\top \phi) - \lambda_2| (\|\phi\|_p \|\Delta\|_q)\}.
\end{aligned}$$

We consider four cases: 1) $\|\phi\|_p^2 > \lambda_1$, $\Phi(R_i) = +\infty$; 2) $\|\phi\|_p^2 = \lambda_1$, $2R_i^\top \phi \neq \lambda_2$, $\Phi(R_i) = +\infty$; 3) $\|\phi\|_p^2 = \lambda_1$, $2R_i^\top \phi = \lambda_2$, $\Phi(R_i) = 0$; 4) $\|\phi\|_p^2 < \lambda_1$, $\Phi(R_i) = (\phi^\top R_i)^2 - \lambda_2 \phi^\top R_i + \frac{(2R_i^\top \phi - \lambda_2)^2 \|\phi\|_p^2}{4(\lambda_1 - \|\phi\|_p^2)}$.

For the first three cases, the value of $\frac{1}{n} \sum_{i=1}^n \Phi(R_i)$ is $+\infty$. Hence, only the fourth case is non-trivial. In this case, problem (3.12) is transformed into

$$\begin{aligned}
&\inf_{\lambda_1 \geq 0, \lambda_2} \left[\frac{1}{n} \sum_{i=1}^n \Phi(R_i) + \lambda_2 \alpha + \lambda_1 \delta \right] \\
&= \inf_{\lambda_1 \geq \|\phi\|_p^2, \lambda_2} \left\{ \frac{1}{n} \sum_{i=1}^n \left[(\phi^\top R_i)^2 - \lambda_2 \phi^\top R_i + \frac{(2R_i^\top \phi - \lambda_2)^2 \|\phi\|_p^2}{4(\lambda_1 - \|\phi\|_p^2)} \right] + \lambda_2 \alpha + \lambda_1 \delta \right\}.
\end{aligned} \tag{3.24}$$

Define

$$H = \frac{1}{n} \sum_{i=1}^n \left[(\phi^\top R_i)^2 - \lambda_2 \phi^\top R_i + \frac{(2R_i^\top \phi - \lambda_2)^2 \|\phi\|_p^2}{4(\lambda_1 - \|\phi\|_p^2)} \right] + \lambda_2 \alpha + \lambda_1 \delta.$$

Taking a partial derivative with respect to λ_2 and setting it to be 0, we get

$$\frac{\partial H}{\partial \lambda_2} = \alpha - \frac{1}{n} \sum_{i=1}^n \left[\phi^\top R_i + \frac{(2\phi^\top R_i - \lambda_2) \|\phi\|_p^2}{2(\lambda_1 - \|\phi\|_p^2)} \right] = 0,$$

which implies (note that $\phi^\top 1 = 1$ guarantees that $\|\phi\|_p^2 > 0$)

$$\lambda_2 = 2\alpha - 2C \frac{\lambda_1}{\|\phi\|_p^2}, \quad (3.25)$$

where $C := \alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R]$. Moreover, λ_2 is optimal because

$$\frac{\partial^2 H}{\partial \lambda_2^2} = \frac{\|\phi\|_p^2}{2(\lambda_1 - \|\phi\|_p^2)} > 0. \quad (3.26)$$

We plug (3.25) into (3.24) and obtain

$$\begin{aligned} & \inf_{\lambda_1 \geq 0, \lambda_2} \left[\frac{1}{n} \sum_{i=1}^n \Phi(R_i) + \lambda_2 \alpha + \lambda_1 \delta \right] \\ &= \frac{1}{n} \sum_{i=1}^n (\phi^\top R_i)^2 + \inf_{\lambda_1 \geq \|\phi\|_p^2, \lambda_2} \left\{ \frac{1}{n} \sum_{i=1}^n \left[-\lambda_2 \phi^\top R_i + \frac{(2R_i^\top \phi - \lambda_2)^2 \|\phi\|_p^2}{4(\lambda_1 - \|\phi\|_p^2)} \right] + \lambda_2 \alpha + \lambda_1 \delta \right\} \\ &= \frac{1}{n} \sum_{i=1}^n (\phi^\top R_i)^2 + \inf_{\lambda_1 \geq \|\phi\|_p^2, \lambda_2} \left\{ \frac{1}{n} \sum_{i=1}^n \left[\frac{(2R_i^\top \phi - \lambda_2)^2 \|\phi\|_p^2}{4(\lambda_1 - \|\phi\|_p^2)} \right] + \lambda_2 C + \lambda_1 \delta \right\} \\ &= \frac{1}{n} \sum_{i=1}^n (\phi^\top R_i)^2 + \inf_{\lambda_1 \geq \|\phi\|_p^2} \left\{ \frac{1}{n} \sum_{i=1}^n \left[\frac{(2R_i^\top \phi - 2\alpha + 2C \frac{\lambda_1}{\|\phi\|_p^2})^2 \|\phi\|_p^2}{4(\lambda_1 - \|\phi\|_p^2)} \right] + (2\alpha - 2C \frac{\lambda_1}{\|\phi\|_p^2})C + \lambda_1 \delta \right\}. \end{aligned}$$

Writing $\lambda_1 = \kappa + \|\phi\|_p^2$, we have

$$\begin{aligned} & \inf_{\lambda_1 \geq 0, \lambda_2} \left[\frac{1}{n} \sum_{i=1}^n \Phi(R_i) + \lambda_2 \alpha + \lambda_1 \delta \right] \\ &= \frac{1}{n} \sum_{i=1}^n (\phi^\top R_i)^2 + \inf_{\kappa \geq 0} \left\{ \frac{1}{n} \sum_{i=1}^n \left[\frac{(R_i^\top \phi - \alpha + C \frac{\kappa + \|\phi\|_p^2}{\|\phi\|_p^2})^2 N}{\kappa} \right] + (2\alpha - 2C \frac{\kappa + \|\phi\|_p^2}{\|\phi\|_p^2})C + (\kappa + \|\phi\|_p^2)\delta \right\} \\ &= \frac{1}{n} \sum_{i=1}^n (\phi^\top R_i)^2 + \inf_{\kappa \geq 0} \left\{ \frac{C_1^2}{\|\phi\|_p^2} k + 2\|\phi\|_p^2 (\phi^\top \bar{W} - \alpha + C) + \frac{1}{n} \sum_{i=1}^n \frac{(R_i^\top \phi - \alpha + C)^2 \|\phi\|_p^2}{\kappa} \right. \\ & \quad \left. + 2\alpha C - 2C^2 + \kappa(\delta - \frac{2C^2}{\|\phi\|_p^2}) + \|\phi\|_p^2 \delta \right\} \\ &= \frac{1}{n} \sum_{i=1}^n (\phi^\top R_i)^2 + 2\alpha C - 2C^2 + \|\phi\|_p^2 \delta + \inf_{\kappa \geq 0} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{(R_i^\top \phi - \mathbb{E}_{\mathbb{P}_n}[R]\phi)^2 \|\phi\|_p^2}{\kappa} + \kappa(\delta - \frac{C^2}{\|\phi\|_p^2}) \right\}. \end{aligned}$$

If $\delta - C^2/\|\phi\|_p^2 < 0$, then the optimal value of the above problem is $-\infty$, which means

that the primal problem (3.11) is not feasible. If $\delta - C^2/||\phi||_p^2 \geq 0$, then

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n (\phi^\top R_i)^2 + 2\alpha C - 2C^2 + ||\phi||_p^2 \delta + \inf_{\kappa \geq 0} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{(R_i^\top \phi - \mathbb{E}_{\mathbb{P}_n}[R]\phi)^2 ||\phi||_p^2}{\kappa} + \kappa \left(\delta - \frac{C^2}{||\phi||_p^2} \right) \right\} \\
&= \frac{1}{n} \sum_{i=1}^n (\phi^\top R_i)^2 + 2(\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] + \delta ||\phi||_p^2 \\
&+ 2\sqrt{\delta ||\phi||_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2} \sqrt{\frac{1}{n} \phi^\top \sum_{i=1}^n (R_i - \mathbb{E}_{\mathbb{P}_n}[R])(R_i - \mathbb{E}_{\mathbb{P}_n}[R])^\top \phi} \\
&= \frac{1}{n} \sum_{i=1}^n (\phi^\top R_i)^2 + 2(\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] + \delta ||\phi||_p^2 \\
&+ 2\sqrt{\delta ||\phi||_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2} \sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}[R]\phi}.
\end{aligned}$$

Thus, problem (3.11) can be written as

$$\begin{aligned}
& \min_{\phi} \frac{1}{n} \sum_{i=1}^n (\phi^\top R_i)^2 + 2(\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] + \delta ||\phi||_p^2 \\
&+ 2\sqrt{\delta ||\phi||_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2} \sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}[R]\phi},
\end{aligned}$$

subject to $1^\top \phi = 1$ and $(\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2 - \delta ||\phi||_p^2 \leq 0$. ■

The condition $(\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2 - \delta ||\phi||_p^2 \leq 0$ ensures that (3.11) is feasible, failing which the optimal value $h(\alpha, \phi) = -\infty$. Proposition 7 ultimately leads to the following main result of the chapter, one that transforms (3.3) into a non-robust portfolio selection problem in terms of the empirical measure \mathbb{P}_n , with an additional “regularization” term.

Theorem 2 *The primal formulation given in (3.3) is equivalent to the following dual problem*

$$\min_{\phi \in \mathcal{F}_{\delta, \alpha}(n)} \left(\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi} + \sqrt{\delta} ||\phi||_p \right)^2, \quad (3.27)$$

in the sense that the two problems have the same optimal solutions and optimal value.

Proof Note that

$$\begin{aligned}
& h(\alpha, \phi) - \alpha^2 \\
&= \mathbb{E}_{\mathbb{P}_n} [(\phi^\top R)^2] + 2(\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] - \alpha^2 + \delta \|\phi\|_p^2 \\
&+ 2\sqrt{\delta \|\phi\|_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2} \sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R) \phi} \\
&= \mathbb{E}_{\mathbb{P}_n} [(\phi^\top R)^2] + 2\alpha\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] - (\phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2 - \alpha^2 - (\phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2 + \delta \|\phi\|_p^2 \\
&+ 2\sqrt{\delta \|\phi\|_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2} \sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R) \phi} \\
&= \phi^\top \text{Var}_{\mathbb{P}_n}(R) \phi + \{\delta \|\phi\|_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2\} \\
&+ 2\sqrt{\delta \|\phi\|_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2} \sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R) \phi} \\
&= \left(\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R) \phi} + \sqrt{\delta \|\phi\|_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2} \right)^2.
\end{aligned}$$

Therefore, it follows from Proposition 7 that

$$\max_{\alpha \geq \bar{\alpha}, (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2 - \delta \|\phi\|_p^2 \leq 0} [h(\alpha, \phi) - \alpha^2] = \left(\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R) \phi} + \sqrt{\delta} \|\phi\|_p \right)^2,$$

with the optimal $\alpha_{opt} = \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] \geq \bar{\alpha}$. This concludes the proof. \blacksquare

It is not difficult to verify that the mapping $\phi \rightarrow \sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R) \phi} + \sqrt{\delta} \|\phi\|_p$ is convex, and we have shown that the feasible region $\mathcal{F}_{\delta, \bar{\alpha}}(n)$ is convex. So (3.2) and therefore (3.3) are both convex optimization problems. As such, they are tractable optimization problems.

Problem (3.2) has an additional term, $\sqrt{\delta} \|\phi\|_p$, in its objective function. In the asset management industry, fund managers using a mean–variance portfolio selection model often add a “penalty” or “regularization” term – in the form of $k \|\phi\|$, where $\|\cdot\|$ is an appropriately chosen norm - in order to enhance the sparsity of the vector as a way to include fewer stocks in the portfolio and to address the issue of overfitting.²

²In practice, it is not desirable to include, in the case of S&P 500 stocks, for example, all of the 500 stocks in one’s portfolio, even though one of the key implications of the mean–variance model is diversification. From a practical perspective, including too many stocks is costly and prone to mismanagement. Therefore, adding a proper regularization term not only reduces overfitting, it also helps to achieve a balance between diversification and manageability.

Here, we provide *interpretability* of this regularization technique (which is based on experience or heuristics) by means of a well-established robustification idea, backed by precise rationality principles; see for example Delage et al. (2019). Moreover, the parameter δ , which reflects the level of regularization, will also be endogenously informed by data, as we show in the next section.

As indicated earlier, it should be noted that Theorem 2 does not follow directly from any of the strong duality results mentioned in Section 1.4.2. This is because the portfolio variance in the objective function (3.3) is *not* a linear function of the probability measure. A related work, Gao et al. (2017), proves only an asymptotic equivalence to regularization. On the other hand, we have obtained the exact equivalence between (3.3) and the regularized optimization problem given in Theorem 2.

To conclude this section, we note that while the cost function is chosen as $c(u, v) = \|u - v\|_q^2$ in the study as presented here, our result (Theorem 2) actually holds for any cost function of the form $c(u, v) = \|u - v\|^2$ where $\|\cdot\|$ is any given norm with a suitable dual. To be more precise, define the dual norm as $\|x\|_* = \sup_{\|z\|=1} |x^\top z|$. Then, the primal distributionally robust model under this alternative cost function is equivalent to the following dual problem:

$$\min_{\phi \in \mathcal{F}_{\delta, \bar{\alpha}}(n)} \left(\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R) \phi} + \sqrt{\delta} \|\phi\|_* \right)^2,$$

where the feasible region is modified as

$$\mathcal{F}_{\delta, \bar{\alpha}}(n) = \{\phi \in \mathbb{R}^d : \phi^\top \mathbf{1} = 1, \mathbb{E}_{\mathbb{P}_n}(\phi^\top R) \geq \bar{\alpha} + \sqrt{\delta} \|\phi\|_*\}.$$

For example, consider a norm as $\|x\| = (x^\top \Sigma x)^{1/2}$ where Σ is a strictly positive definite matrix. Then $\|x\|_* = (x^\top \Sigma^{-1} x)^{1/2}$. Interested readers may refer to Blanchet and Kang (2017) for discussion of some other interesting norms.

3.3 Choice of Model Parameters

There are two key parameters, δ and $\bar{\alpha}$, in the formulation (3.3), the choice of which is not only curious in theory, but also crucial in practical implementation and for the success of our algorithm. The idea is that the choice of these parameters should be informed by the data (i.e., in a data-driven way) based on some statistical principles, rather than being arbitrarily exogenous. Specifically, we define the distributional uncertainty region just large enough that the correct optimal portfolio (the one that we would apply if the underlying distribution was known) becomes a plausible choice with a sufficiently high confidence level. Once this is determined, we then determine the feasible set of portfolios just large enough that the correct optimal portfolio is feasible with adequately high confidence.

We need to impose several technical/statistical assumptions.

A1) The underlying return time series $(R_k : k \geq 0)$ is a stationary, ergodic process satisfying $\mathbb{E}_{\mathbb{P}^*} (\|R_k\|_2^4) < \infty$ for each $k \geq 0$. Moreover, for each measurable $g(\cdot)$ such that $|g(x)| \leq c(1 + \|x\|_2^2)$ for some $c > 0$, the limit

$$\Upsilon_g := \lim_{n \rightarrow \infty} \text{Var}_{\mathbb{P}^*} \left(n^{-1/2} \sum_{k=1}^n g(R_k) \right)$$

exists and the central limit theorem holds:

$$n^{1/2} [\mathbb{E}_{\mathbb{P}_n} (g(R)) - \mathbb{E}_{\mathbb{P}^*} (g(R))] \Rightarrow N(0, \Upsilon_g),$$

where (and henceforth) “ \Rightarrow ” denotes weak convergence.

A2) For any matrix $\Lambda \in \mathbb{R}^{d \times d}$ and any vector $\zeta \in \mathbb{R}^d$ such that either $\Lambda \neq 0$ or $\zeta \neq 0$,

$$\mathbb{P}^* (\|\Lambda R + \zeta\|_2 > 0) > 0.$$

A3) The classical model (1.1) has a unique solution ϕ^* . Moreover, $\text{Var}_{\mathbb{P}^*} [\mathbb{E}_{\mathbb{P}^*} (R^\top) R] > 0$.

Assumption **A1)** is standard for most time series models (after removing seasonality). Assumption **A2)** holds, assuming, for example, that R has a density. Assumption **A3)** is a technical assumption that can be relaxed, but then the evaluation of the optimal choice of δ would become more cumbersome, as we explain below.

3.3.1 Choice of δ

The choice of the uncertainty size δ is crucial. If δ is too large, then there is too much model ambiguity and the available data becomes less relevant. In this case, the resulting optimal portfolio will tend to consist merely of equal allocations. If δ is too small, then the effect of robustification will be negligible. Therefore, the choice of δ should *not* be exogenously specified; rather, it should be endogenously informed by the data.

Theorem 2 actually suggests an appropriate order of $\delta = \delta_n$ (here, n is the size of the available return time series data) in terms of n . Because the differences between the optimal standard deviation obtained by solving (1.1) and that obtained by solving the empirical version of (1.1) are of the order $O(n^{-1/2})$, it follows from Theorem 2 that any choice of δ_n in the order of $o(n^{-1})$ would be too small. Hence, an “optimal” order of δ_n should be of the order $O(n^{-1})$.

In order to choose an appropriate δ_n , we follow here the idea behind the robust Wasserstein profile inference (RWPI) approach introduced in Blanchet et al. (2016). Intuitively, δ should be chosen such that the set $\mathcal{U}_\delta(\mathbb{P}_n) = \{\mathbb{P} : D_c(\mathbb{P}, \mathbb{P}_n) \leq \delta\}$ contains all of the probability measures that constitute plausible variations of the data represented by \mathbb{P}_n . Denote by $\mathcal{Q}(\mathbb{P})$ the classical Markowitz portfolio selection

problem with target return ρ assuming that \mathbb{P} is the underlying model:

$$\begin{aligned} \min_{1^\top \phi = 1} \quad & \phi^\top \mathbb{E}_{\mathbb{P}}[RR^\top] \phi \\ \text{subject to} \quad & \phi^\top \mathbb{E}_{\mathbb{P}}[R] = \rho, \end{aligned} \tag{3.28}$$

and denote by $\phi_{\mathbb{P}}$ a solution to $\mathcal{Q}(\mathbb{P})$ and by $\Phi_{\mathbb{P}}$ the set of all such solutions. According to Assumption **A3**), we have $\Phi_{\mathbb{P}^*} = \{\phi^*\}$ for some portfolio ϕ^* . Therefore, there exist (unique) Lagrange multipliers λ_1^* and λ_2^* such that

$$\begin{aligned} 2\mathbb{E}_{\mathbb{P}^*}(RR^\top)\phi^* - \lambda_1^*\mathbb{E}_{\mathbb{P}^*}[R] - \lambda_2^*1 &= 0, \\ (\phi^*)^\top \mathbb{E}_{\mathbb{P}^*}[R] - \rho &= 0. \end{aligned} \tag{3.29}$$

Now, when δ is suitably chosen so that $\mathcal{U}_\delta(\mathbb{P}_n)$ constitutes the models that are plausible variations of \mathbb{P}_n , any $\phi_{\mathbb{P}}$ with $\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)$ is a plausible estimate of ϕ^* . This intuition motivates the definition of the following set:

$$\Lambda_\delta(\mathbb{P}_n) = \cup_{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)} \Phi_{\mathbb{P}},$$

which corresponds to all the plausible estimates of ϕ^* . As a result, $\Lambda_\delta(\mathbb{P}_n)$ is a natural confidence region for ϕ^* and, therefore, δ should be chosen as the smallest number δ_n^* such that ϕ^* belongs to this region with a given confidence level. Namely,

$$\delta_n^* = \min\{\delta > 0 : \mathbb{P}^*(\phi^* \in \Lambda_\delta(\mathbb{P}_n)) \geq 1 - \delta_0\},$$

where $1 - \delta_0$ is a user-defined confidence level (typically 95%).

However, by mere definition, it is difficult to compute δ_n^* . We now provide a simpler representation for δ_n^* via an auxiliary function called the robust Wasserstein profile (RWP) function. To this end, first observe that any $\phi \in \Lambda_\delta(\mathbb{P}_n)$ if and only if

there exist $\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)$ and $\lambda_1, \lambda_2 \in (-\infty, \infty)$ such that

$$\begin{aligned} 2\mathbb{E}_{\mathbb{P}}(RR^\top)\phi - \lambda_1\mathbb{E}_{\mathbb{P}}[R] - \lambda_2\mathbf{1} &= 0, \\ \phi^\top\mathbb{E}_{\mathbb{P}}(R) - \rho &= 0. \end{aligned}$$

From these two equations, multiplying the first equation by ϕ , substituting the expression in the second equation, and noting that $\phi \cdot \mathbf{1} = 1$, we obtain

$$\lambda_2 = 2(\phi)^\top\mathbb{E}_{\mathbb{P}}(RR^\top)\phi - \lambda_1\rho.$$

We now define the following RWP function:

$$\bar{\mathcal{R}}_n(\phi, \lambda_1, \Sigma, \mu) := \inf \left\{ D_c(\mathbb{P}, \mathbb{P}_n) : \begin{cases} 2\Sigma\phi - \lambda_1\mu = \left(2(\phi)^\top\Sigma\phi - \lambda_1\mu \cdot \phi\right)\mathbf{1} \\ \mu = \mathbb{E}_{\mathbb{P}}[R], \Sigma = \mathbb{E}_{\mathbb{P}}(RR^\top) \end{cases} \right\},$$

for $(\phi, \lambda_1, \Sigma, \mu) \in \mathbb{R}^d \times \mathbb{R} \times \mathcal{S}_+^{d \times d} \times \mathbb{R}^d$ where $\mathcal{S}_+^{d \times d}$ is the set of all the symmetric positive semidefinite matrices, and we convent that $\inf \emptyset := +\infty$. Moreover, define

$$\bar{\mathcal{R}}_n^*(\phi^*) := \inf_{\Sigma \in \mathcal{S}_+^{d \times d}, \mu \in \mathbb{R}^d, \lambda_1 \in \mathbb{R}} \bar{\mathcal{R}}_n(\phi^*, \lambda_1, \Sigma, \mu).$$

It follows directly from the definitions that

$$\phi^* \in \Lambda_\delta(\mathbb{P}_n) \implies \bar{\mathcal{R}}_n^*(\phi^*) \leq \delta, \tag{3.30}$$

while for any given $\epsilon > 0$

$$\bar{\mathcal{R}}_n^*(\phi^*) \leq \delta + \epsilon \implies \phi^* \in \Lambda_\delta(\mathbb{P}_n). \tag{3.31}$$

Let us define

$$\tilde{\delta}_n^* = \inf\{\delta > 0 : \mathbb{P}^*(\bar{\mathcal{R}}_n^*(\phi^*) \leq \delta) \geq 1 - \delta_0\}.$$

It follows from (3.30) and (3.31) that $\tilde{\delta}_n^* \leq \delta_n^* \leq \tilde{\delta}_n^* + \epsilon$. As $\epsilon > 0$ is arbitrary, we obtain

$$\delta_n^* = \tilde{\delta}_n^* = \inf\{\delta : \mathbb{P}^*(\bar{\mathcal{R}}_n^*(\phi^*) \leq \delta) \geq 1 - \delta_0\}.$$

In other words, δ_n^* is the quantile corresponding to the $1 - \delta_0$ percentile of the distribution of $\bar{\mathcal{R}}_n^*(\phi^*)$.³

Still, even under A3), the statistic $\bar{\mathcal{R}}_n^*(\phi^*)$ is somewhat cumbersome to work with, as it is derived from solving a minimization problem in terms of the mean and variance. So, instead, we define an alternative statistic involving only the empirical mean and variance while producing an upper bound of $\delta = \delta_n$, which still preserves the target rate of convergence to zero as $n \rightarrow \infty$ (which, as we have argued, should be of the order $O(n^{-1})$).

Denote $\Sigma_n = \mathbb{E}_{\mathbb{P}_n}(RR^\top)$ and let λ_1^* be the Lagrange multiplier in (3.29). Set

$$\mu_n = \rho 1 + 2(\Sigma_n \phi^* - \phi^{*T} \Sigma_n \phi^* 1) / \lambda_1^*. \quad (3.32)$$

Define

$$\mathcal{R}_n(\Sigma_n, \mu_n) := \bar{\mathcal{R}}_n(\phi^*, \lambda_1^*, \Sigma_n, \mu_n).$$

It is clear that

$$\mathcal{R}_n(\Sigma_n, \mu_n) \geq \bar{\mathcal{R}}_n^*(\phi^*).$$

Therefore,

$$\mathcal{R}_n(\Sigma_n, \mu_n) \leq \delta \implies \bar{\mathcal{R}}_n^*(\phi^*) \leq \delta$$

and, consequently,

$$\bar{\delta}_n^* = \inf\{\delta \geq 0 : \mathbb{P}^*(\mathcal{R}_n(\Sigma_n, \mu_n) \leq \delta) \geq 1 - \delta_0\} \geq \delta_n^*. \quad (3.33)$$

³Herein the analysis is under Assumption A3). If $\Phi_{\mathbb{P}^*}$ contained more than just one element, then there would be several possible options to formulate an optimization problem for choosing δ . For example, we might choose δ as the smallest uncertainty size such that $\Phi_{\mathbb{P}^*} \subset \Lambda_\delta(\mathbb{P}_n)$ with probability $1 - \delta_0$, in which case we would need to study $\sup_{\phi^* \in \Phi_{\mathbb{P}^*}} \bar{\mathcal{R}}_n^*(\phi^*)$.

Moreover, because of the choice of Σ_n and μ_n , we have

$$\mathcal{R}_n(\Sigma_n, \mu_n) = \inf \{ \mathcal{D}_c(\mathbb{P}, \mathbb{P}_n) : \mathbb{E}_{\mathbb{P}}[RR^\top] = \Sigma_n, \mathbb{E}_{\mathbb{P}}[R] = \mu_n \}.$$

The next result shows $\bar{\delta}_n^* = O(n^{-1})$ as $n \rightarrow \infty$.

Theorem 3 *Assume that **A1**) and **A2**) hold and write $\mu_* = \mathbb{E}_{\mathbb{P}^*}(R)$ and $\Sigma_* = \mathbb{E}_{\mathbb{P}^*}(RR^\top)$. Define $g(x) = x + 2(xx^\top \cdot \phi^* - \phi^{*T}xx^\top\phi^*1)/\lambda_1^*$. Then,*

$$n\mathcal{R}_n(\Sigma_n, \mu_n) \Rightarrow L_0 := \sup_{\bar{\lambda} \in \mathbb{R}^d} \left(\bar{\lambda}^\top Z - \inf_{\bar{\Lambda} \in \mathbb{R}^{d \times d}} \mathbb{E}_{\mathbb{P}^*} [\|\bar{\Lambda}R + \bar{\lambda}\|_p^2] \right)$$

where $Z \sim N(0, \Upsilon_g)$. Moreover, if $p = 2$, then

$$L_0 = \frac{\|Z\|_2^2}{4(1 - \|\mu_*\|_2^4 / \mu_*^\top \Sigma_* \mu_*)}.$$

Proof Define

$$h_0(R, \Sigma) = RR^\top - \Sigma \quad \text{and} \quad h_1(R, \mu) = R - \mu.$$

Then, by Proposition 1 of Blanchet et al. (2016), we have that for any given μ and Σ

$$\mathcal{R}_n(\Sigma, \mu) = \sup_{\Lambda \in \mathbb{R}^{d \times d}, \lambda \in \mathbb{R}^d} \left\{ -\mathbb{E}_{\mathbb{P}_n} \left[\sup_{u \in \mathbb{R}^d} \{ \text{Tr}(\Lambda h_0(u, \Sigma)) + \lambda^\top h_1(u, \mu) - \|u - R\|_q^2 \} \right] \right\}.$$

Observe that

$$\begin{aligned} & \sup_{u \in \mathbb{R}^d} \{ \text{Tr}(\Lambda h_0(u, \Sigma)) + \lambda^\top h_1(u, \mu) - \|u - R\|_q^2 \} \\ &= \sup_{\Delta \in \mathbb{R}^d} \{ \text{Tr}(\Lambda h_0(\Delta + R, \Sigma)) + \lambda^\top h_1(\Delta + R, \mu) - \|\Delta\|_q^2 \} \\ &= \sup_{\Delta \in \mathbb{R}^d} \{ \text{Tr}(\Lambda [h_0(\Delta + R, \Sigma) - h_0(R, \Sigma)]) + \lambda^\top \Delta - \|\Delta\|_q^2 \} \\ &+ \text{Tr}(\Lambda h_0(R, \Sigma)) + \lambda^\top h_1(R, \mu). \end{aligned}$$

Moreover, let us write

$$\text{Tr}(\Lambda [h_0(\Delta + R, \Sigma) - h_0(R, \Sigma)]) = \int_0^1 \frac{d}{dt} \text{Tr}(\Lambda h_0(R + t\Delta)) dt.$$

However,

$$\begin{aligned} \frac{d}{dt} \text{Tr}(\Lambda h_0(R + t\Delta)) &= 2\text{Tr}(\Lambda (R + t\Delta) \Delta^\top) \\ &= 2\text{Tr}(\Lambda R \Delta^\top) + 2t\Delta^\top \Lambda \Delta. \end{aligned}$$

Furthermore,

$$\mathbb{E}_{\mathbb{P}_n} [\text{Tr}(\Lambda h_0(R, \Sigma))] |_{\Sigma = \Sigma_n} = 0. \quad (3.34)$$

So, we deduce

$$\begin{aligned} \mathcal{R}_n(\Sigma_n, \mu) &= \sup_{\lambda \in \mathbb{R}^d} \{-\mathbb{E}_{\mathbb{P}_n}[\lambda^\top (R - \mu)] + \\ &\quad \sup_{\Lambda \in \mathbb{R}^{d \times d}} (-\mathbb{E}_{\mathbb{P}_n}[\sup_{\Delta} \{2\text{Tr}(\Lambda R \Delta^\top) + \Delta^\top \Lambda \Delta + \lambda^\top \Delta - \|\Delta\|_q^2\}])\}. \end{aligned}$$

Introduce the scaling $\Delta = \bar{\Delta}/n^{1/2}$ and $\bar{\lambda} = \lambda n^{1/2}$ and $\bar{\Lambda} = \Lambda n^{1/2}$. Then, we obtain

$$\begin{aligned} n\mathcal{R}_n(\Sigma_n, \mu_n) &= \sup_{\bar{\lambda} \in \mathbb{R}^d} \{-n^{-1/2} \sum_{i=1}^n \bar{\lambda}^\top (R_i - \mu_n) + \\ &\quad \sup_{\bar{\Lambda} \in \mathbb{R}^{d \times d}} (-\mathbb{E}_{\mathbb{P}_n}[\sup_{\bar{\Delta}} \{2\text{Tr}(\bar{\Lambda} R \bar{\Delta}^\top) + \bar{\Delta}^\top \bar{\Lambda} \bar{\Delta}/n^{1/2} + \bar{\lambda}^\top \bar{\Delta} - \|\bar{\Delta}\|_q^2\}])\}. \end{aligned}$$

In the proof of Proposition 3 in Blanchet et al. (2016), under Assumption **A2**), a technique is introduced to show that $\bar{\Delta}$ and $\bar{\lambda}$ can be restricted to compact sets with high probability, and therefore the term $\bar{\Delta}^\top \bar{\Lambda} \bar{\Delta}/n^{1/2}$ is asymptotically negligible. On

the other hand,

$$\begin{aligned} & \sup_{\Delta} \{2\text{Tr}(\bar{\Delta}^\top \bar{\Lambda} R) + \bar{\Delta}^\top \bar{\lambda} - \|\bar{\Delta}\|_q^2\} \\ &= \sup_{\Delta} \{2 \|\bar{\Lambda} R + \bar{\lambda}\|_p \|\bar{\Delta}\|_q - \|\bar{\Delta}\|_q^2\} = \|\bar{\Lambda} R + \bar{\lambda}\|_p^2. \end{aligned}$$

Therefore, if

$$n^{-1/2} \sum_{i=1}^n (R_i - \mu_n) \Rightarrow -Z$$

for some Z (to be characterized momentarily), then we conclude that

$$\mathcal{R}_n(\Sigma_n, \mu_n) \Rightarrow L_0 = \sup_{\bar{\lambda} \in \mathbb{R}^d} \{\bar{\lambda}^\top Z - \inf_{\bar{\Lambda} \in \mathbb{R}^{d \times d}} \mathbb{E}_{\mathbb{P}^*} [\|\bar{\Lambda} R + \bar{\lambda}\|_p^2]\}.$$

If $p = 2$, then we have

$$\mathbb{E}_{\mathbb{P}^*} [\|\bar{\Lambda} R + \bar{\lambda}\|_2^2] = \sum_i \mathbb{E}_{\mathbb{P}^*} (\bar{\Lambda}_i \cdot R + \bar{\lambda}_i)^2.$$

So, taking the derivative with respect to the i -th row, $\bar{\Lambda}_i$, of the matrix $\bar{\Lambda}$, $\bar{\Lambda}_i$, we obtain

$$\nabla_{\bar{\Lambda}_i} \mathbb{E}_{\mathbb{P}^*} [\|\bar{\Lambda} R + \bar{\lambda}\|_2^2] = 2\mathbb{E}_{\mathbb{P}^*} ((R^\top \bar{\Lambda}_i + \bar{\lambda}_i) R) = 2\mathbb{E}_{\mathbb{P}^*} (R^\top \bar{\Lambda}_i R) + 2\bar{\lambda}_i \mathbb{E}_{\mathbb{P}^*} (R) = 0. \quad (3.35)$$

Writing

$$\mu_* = \mathbb{E}_{\mathbb{P}^*} (R) \text{ and } \Sigma_* = \mathbb{E}_{\mathbb{P}^*} (RR^\top),$$

and then multiplying (3.35) by $\bar{\Lambda}_i^\top$, we obtain

$$\bar{\Lambda}_i^\top \Sigma_* \bar{\Lambda}_i = -\bar{\lambda}_i \bar{\Lambda}_i^\top \mu_*.$$

To solve this equation, take

$$\bar{\Lambda}_i = a_i \mu_*,$$

leading to

$$a_i \mu_*^\top \Sigma_* \mu_* = -\bar{\lambda}_i \|\mu_*\|_2^2,$$

or

$$a_i = -\bar{\lambda}_i \|\mu_*\|_2^2 / \mu_*^\top \Sigma_* \mu_*.$$

Therefore,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^*} (\bar{\Lambda}_i R + \bar{\lambda}_i)^2 &= \bar{\Lambda}_i^\top \Sigma_* \bar{\Lambda}_i + 2\bar{\lambda}_i \bar{\Lambda}_i^\top \mu_* + \bar{\lambda}_i^2 \\ &= \bar{\lambda}_i^2 + \bar{\lambda}_i \bar{\Lambda}_i^\top \mu_* = \bar{\lambda}_i^2 (1 - \|\mu_*\|_2^4 / \mu_*^\top \Sigma_* \mu_*). \end{aligned}$$

Observe that $\|\mu_*\|_2^4 / \mu_*^\top \Sigma_* \mu_* < 1$ if and only if

$$Tr (\mathbb{E}_{\mathbb{P}^*} (RR^\top) \mathbb{E}_{\mathbb{P}^*} (R) \mathbb{E}_{\mathbb{P}^*} (R^\top)) > Tr (\mathbb{E}_{\mathbb{P}^*} (R) \mathbb{E}_{\mathbb{P}^*} (R^\top) \mathbb{E}_{\mathbb{P}^*} (R) \mathbb{E}_{\mathbb{P}^*} (R^\top)),$$

which in turn holds if and only if

$$\begin{aligned} Tr (\mathbb{E}_{\mathbb{P}^*} (R^\top) [\mathbb{E}_{\mathbb{P}^*} (RR^\top) - \mathbb{E}_{\mathbb{P}^*} (R) \mathbb{E}_{\mathbb{P}^*} (R^\top)] \mathbb{E}_{\mathbb{P}^*} (R)) \\ = \text{Var}_{\mathbb{P}^*} (\mathbb{E}_{\mathbb{P}^*} (R^\top) R) > 0. \end{aligned}$$

It follows from **A3**) that $\text{Var}_{\mathbb{P}^*} (\mathbb{E}_{\mathbb{P}^*} (R^\top) R) > 0$. Hence,

$$\begin{aligned} L_0 &= \sup_{\bar{\lambda} \in \mathbb{R}^d} \{ \bar{\lambda}^\top Z - \inf_{\bar{\Lambda} \in \mathbb{R}^{d \times d}} \mathbb{E}_{\mathbb{P}^*} [\|\bar{\Lambda} R + \bar{\lambda}\|_2^2] \} \\ &= \sup_{\bar{\lambda} \in \mathbb{R}^d} \{ \bar{\lambda}^\top Z - \|\bar{\lambda}\|_2^2 (1 - \|\mu_*\|_2^4 / \mu_*^\top \Sigma_* \mu_*) \} \\ &= \frac{\|Z\|_2^2}{4 (1 - \|\mu_*\|_2^4 / \mu_*^\top \Sigma_* \mu_*)}. \end{aligned}$$

It remains to identify Z . Observe that

$$\begin{aligned}\mu_n &= \rho 1 + 2 (\Sigma_n \phi^* - \phi^{*T} \Sigma_n \phi^* 1) / \lambda_1^* \\ &= \rho 1 + 2 (\Sigma_* \phi^* - \phi^{*T} \Sigma_* \phi^* 1) / \lambda_1^* \\ &\quad + 2 (H_n \phi^* - \phi^{*T} H_n \phi^* 1) / \lambda_1^* \\ &= \mu_* + 2 (H_n \phi^* - \phi^{*T} H_n \phi^* 1) / \lambda_1^*,\end{aligned}$$

where $H_n := \Sigma_n - \Sigma_*$. By **A1**) we have

$$\begin{aligned}n^{-1/2} \sum_{i=1}^n (R_i - \mu_*) &\Rightarrow Z_0 \sim N(0, \Upsilon_{g_1}), \\ n^{1/2} H_n &\Rightarrow Y \sim N(0, \Upsilon_{g_2}).\end{aligned}$$

Thus,

$$\begin{aligned}n^{-1/2} \sum_{i=1}^n \bar{\lambda}^\top (R_i - \mu_*) + 2n^{1/2} \bar{\lambda}^\top (H_n \phi^* - \phi^{*T} H_n \phi^* 1) / \lambda_1^* \\ \Rightarrow \bar{\lambda}^\top Z = \bar{\lambda}^\top (Z_0 + Z_1),\end{aligned}$$

where

$$Z_1 := 2 (Y \phi^* - \phi^{*T} Y \phi^* 1) / \lambda_1^*.$$

■

Note that L_0 has an explicit expression when $p = 2$. When $p \neq 2$, using the inequalities that $\|x\|_p^2 \geq \|x\|_2^2$ if $p < 2$ and $d^{(\frac{1}{2}-\frac{1}{p})} \|x\|_p^2 \geq \|x\|_2^2$ if $p > 2$, we can find a stochastic upper bound of L_0 that can be explicitly expressed. In that case, we can obtain $\bar{\delta}_n^*$ in exactly the same way; namely, first compute the $1 - \delta_0$ quantile of L_0 and then let $\bar{\delta}_n^*$ be such a quantile multiplied by $1/n$. The distribution of L_0 can be calibrated using a natural plug-in estimator, leading to an asymptotically equivalent estimator of $\bar{\delta}_n^*$. The validity of this type of (plug-in) approach is explained in the

following section in the context of choosing $\bar{\alpha}$, but the principle applies directly to the setting of L_0 as well. Simply put, whenever an asymptotic limiting distribution depends continuously on various parameters and consistent estimators are available for those parameters, then consistent plug-in estimators can be safely used and still preserve exactly the same asymptotic distributions. The details of this approach are investigated in Proposition 2 of Blanchet et al. (2019), and the performance of such plug-in estimators is tested empirically in Section 3.4 of this chapter.

3.3.2 Choice of $\bar{\alpha}$

Once δ has been chosen, the next step is to choose $\bar{\alpha}$. The idea is to select $\bar{\alpha}$ just large enough to make sure that we do not rule out the inclusion $\phi^* \in \mathcal{F}_{\delta, \bar{\alpha}}(n)$, with a given confidence level chosen by the user, where ϕ^* is the optimal solution to (1.1). It is equivalent to choosing v_0 where

$$\bar{\alpha} = \rho - \sqrt{\delta} \|\phi^*\|_p v_0.$$

Therefore, it follows from Proposition 5 that $\phi^* \in \mathcal{F}_{\delta, \bar{\alpha}}(n)$ if and only if

$$(\phi^*)^\top \mathbb{E}_{\mathbb{P}_n}(R) - \sqrt{\delta} \|\phi^*\|_p \geq \rho - \sqrt{\delta} \|\phi^*\|_p v_0.$$

However, $\rho = (\phi^*)^\top \mathbb{E}_{\mathbb{P}^*}(R)$; so the previous inequality holds if and only if

$$(\phi^*)^\top [\mathbb{E}_{\mathbb{P}_n}(R) - \mathbb{E}_{\mathbb{P}^*}(R)] \geq \|\phi^*\|_p \sqrt{\delta} (1 - v_0). \quad (3.36)$$

Hence, we can choose $\sqrt{\delta} (1 - v_0) < 0$ sufficiently negative so that the previous inequality holds with a specified confidence level. We hope to choose a v_0 such that ϕ^* will satisfy (3.36) with confidence level $1 - \epsilon$. This can be achieved asymptotically by a central limit theorem, as the following result indicates.

Proposition 8 *Suppose that **A1**) and **A3**) hold and let $\{\phi_n^*\}_{n=1}^\infty$ be any consistent*

sequence of estimators of ϕ^* in the sense that $\phi_n^* \rightarrow \phi^*$ in probability as $n \rightarrow \infty$.

Then,

$$n^{1/2} \left\{ \frac{(\phi_n^*)^\top [\mathbb{E}_{\mathbb{P}_n}(R) - \mathbb{E}_{\mathbb{P}^*}(R)]}{\|\phi_n^*\|_p} \right\} \Rightarrow N(0, \Upsilon_{\phi^*})$$

as $n \rightarrow \infty$, where

$$\Upsilon_{\phi^*} := \lim_{n \rightarrow \infty} \text{Var}_{\mathbb{P}^*} \left(n^{-1/2} \sum_{k=1}^n (\phi^*)^\top R_k / \|\phi^*\|_p \right).$$

Proof Note that

$$\begin{aligned} n^{1/2} \left\{ \frac{(\phi_n^*)^\top [\mathbb{E}_{\mathbb{P}_n}(R) - \mathbb{E}_{\mathbb{P}^*}(R)]}{\|\phi_n^*\|_p} \right\} &= n^{1/2} \left\{ \frac{(\phi_n^* - \phi^*)^\top [\mathbb{E}_{\mathbb{P}_n}(R) - \mathbb{E}_{\mathbb{P}^*}(R)]}{\|\phi_n^*\|_p} \right\} \\ &\quad + n^{1/2} \left\{ \frac{(\phi^*)^\top [\mathbb{E}_{\mathbb{P}_n}(R) - \mathbb{E}_{\mathbb{P}^*}(R)]}{\|\phi^*\|_p} \cdot \frac{\|\phi^*\|_p}{\|\phi_n^*\|_p} \right\}. \end{aligned}$$

By the standard central limit theorem and the fact that $\phi_n^* \rightarrow \phi^*$ in probability, we conclude that

$$n^{1/2} \left\{ \frac{(\phi_n^*)^\top [\mathbb{E}_{\mathbb{P}_n}(R) - \mathbb{E}_{\mathbb{P}^*}(R)]}{\|\phi_n^*\|_p} - \frac{(\phi^*)^\top [\mathbb{E}_{\mathbb{P}_n}(R) - \mathbb{E}_{\mathbb{P}^*}(R)]}{\|\phi^*\|_p} \right\} \Rightarrow 0$$

as $n \rightarrow \infty$. However, again by the central limit theorem, we have

$$n^{1/2} \left\{ \frac{(\phi^*)^\top [\mathbb{E}_{\mathbb{P}_n}(R) - \mathbb{E}_{\mathbb{P}^*}(R)]}{\|\phi^*\|_p} \right\} \Rightarrow N(0, \Upsilon_{\phi^*}),$$

which yields the desired result. ■

Using the previous result we can estimate v_0 asymptotically. Let ϕ_n denote the optimal solution of problem $\mathcal{Q}(\mathbb{P}_n)$. We know that ϕ_n converges to ϕ^* in probability. So, we choose a v_0 such that the following inequality will hold with confidence level $1 - \epsilon$,

$$\frac{1}{\|\phi_n\|_p} (\phi_n)^\top [\mathbb{E}_{\mathbb{P}_n}(R) - \mathbb{E}_{\mathbb{P}^*}(R)] \geq \sqrt{\delta}(1 - v_0). \quad (3.37)$$

According to Proposition 8, the left-hand side of (3.37) is approximately normally

distributed, and thus we can choose its $1 - \epsilon$ quantile and consequently decide the value of $v_0 > 1$.

We now present a simple “menu” for estimating δ and $\bar{\alpha}$.

1. Choose the target return rate ρ .
2. Collect return data $\{R_i\}_{i=1}^n$.
3. Use the sample mean $\mu_n = \mathbb{E}_{\mathbb{P}_n}(R)$ and the sample second-moment matrix $\Sigma_n = \mathbb{E}_{\mathbb{P}_n}(RR^\top)$ to approximate μ_* and Σ_* , respectively, appearing in Theorem 3.
4. Use the solution ϕ_n , which is the solution to problem $\mathcal{Q}(\mathbb{P}_n)$ (see (3.28)), to approximate ϕ^* in Theorem 3.
5. Apply Theorem 3 and (3.33) to determine $\delta = \bar{\delta}^*$ with the 95% confidence level.
6. Choose v_0 based on the 95% quantile according to (3.37) and Proposition 8, and consequently obtain $\bar{\alpha}$.

3.4 Empirical Performance and Comparisons

In this section we report the results of our backtesting experiments on S&P 500 constituents in order to compare the performance of our DRMV portfolios with those of the portfolios based on the following models: classical (non-robust) single-period Markowitz, continuous-time Markowitz, Fama–French, Black–Litterman, robust Goldfarb–Iyengar, and an equally weighted portfolio. The first four models are well-known and have been widely used in practice, and the fifth is an alternative robust model not based on distributional uncertainty. The equally weighted strategy is actually an extreme outcome of the DRMV model when the uncertainty size $\delta = \infty$.

3.4.1 Experiment Design and Data Preparation

We backtested for the period January 2000–December 2016 with the training (estimation) period being January 1991–December 1999 (i.e., the previous 10 years).⁴ All the stock monthly price data had been obtained from the database of Wharton Business School. At the beginning of 2000, we *randomly* chose 100 stocks from the constituents of the S&P 500 that had at least ten years’ historical price data available.⁵ The basic period is set to be one year in all the single-period models involved, with target annual mean return rate fixed at $\rho = 10\%$. Then, we used the training period to estimate the out-of-sample parameters, namely the mean and the variance, in order to construct the optimal strategies of the various models tested.

DRMV model

Let us first describe in detail the construction of the DRMV strategy for the selected 100 stocks. We generated this 17-year long strategy in an (overlapping) rolling horizon fashion, with each horizon being a month. Specifically, on the first trading day of January 2000, we solved our DRMV model to obtain a portfolio, denoted as ϕ_R . In doing so, we set $p = q = 2$ and $\rho = 10\%$, and obtained the parameters δ and $\bar{\alpha}$ using the menu presented at the end of Section 3.3. We then substituted δ and $\bar{\alpha}$ in the optimization problem described in Theorem 2 to obtain ϕ_R .

We kept ϕ_R until only the first trading day of February 2000. At that point we

⁴We chose the period 2000–2016 for our backtestings for a reason: the market was overall very volatile during this period, experiencing two major crashes: the dot com bubble burst and the subprime financial crisis, followed by a long bull run. We were particularly interested to see how “robust” our DRMV strategies would have been when sailing through such a bumpy journey.

⁵In theory, we should have included *all* the constituents of S&P 500 in our portfolios. However, that would be computationally inefficient and practically (almost) infeasible for most of the models under testing (e.g., the original Markowitz model). Therefore, it is desirable to choose a small subset of stocks based on which to apply various models. This “stock selection” is ultimately an important part of the overall portfolio management. In this chapter, however, we aim to test the performance of “stock allocation” (namely, to allocate wealth among the stocks that have been *already* selected in order to achieve the best risk-adjusted return) of these models. That is why we randomly selected a small subset of stocks in order to focus on the part of the *stock allocation*. On the other hand, the requirement that the selected stocks have at least ten years’ price data is due to the length of the training period.

re-estimated the parameters δ and $\bar{\alpha}$ using the *immediate* previous ten-year (namely, February 1991 – January 2000) price data and re-solved the DRMV model, and generated a new portfolio ϕ_R for February 2000, the second month in our backtesting period. We repeated the same steps for all of the subsequent months.

If at the beginning of a month, some stocks in our portfolio had been removed from the S&P 500 during the previous month, we also removed them from our portfolio, replace them by the same number of stocks that were randomly picked from the S&P 500 (yet having at least ten years' historical data), and then re-balanced based on our DRMV model. We still denoted by ϕ_R the overall portfolio for the 17-year period and kept track of the wealth process that had been updated at the end of each month.

In what follows we describe the implementations of the other models, mentioned at the beginning of this section, under comparison. Except for the continuous-time Markowitz model, all of the rest are single-period models, so we applied the same monthly rolling horizon approach to build the respective strategies. Moreover, for these single-period models, whenever stocks were dropped from the S&P 500, we replaced them with exactly the same set of stocks as in the DRMV model, so as to maintain consistency across models. The case of the continuous-time model is slightly more complicated, and we explain below how we dealt with the issues it raised.

Single-Period Markovitz Model

For the single-period Markovitz model, we consider the following problem:

$$\min_{\phi: \mathbf{1}^\top \phi = 1, \mathbb{E}_{\mathbb{P}_n}[\phi^\top R] = \rho} \text{Var}_{\mathbb{P}_n}(\phi^\top R), \quad (3.38)$$

where ρ is the targeted expected return of the portfolio. We used the sample mean and sample covariance matrix of the previous ten-year return data to estimate $\mathbb{E}_{\mathbb{P}_n}[R]$ and $\text{Var}_{\mathbb{P}_n}[R]$ in each month. Then we generated the optimal portfolio for the single-period Markowitz model, ϕ_M , by setting $\rho = 10\%$ and solving problem (3.38), on

exactly the same rolling horizon basis as for the DRMV model.

Continuous-Time Markovitz Model

The continuous-time Markowitz mean-variance model is based on Cui et al. (2012), in which portfolios were constructed on risky stocks only. This setting is consistent with ours.⁶ It is assumed that the stock price process follows correlated time-inhomogeneous Black-Scholes dynamics. Let $\{X(t) : t \in [0, T]\}$ be the wealth process (also called an admissible wealth process) under any given admissible portfolio. The mean-variance problem is

$$\text{Minimize } \text{Var}(X(T)) \tag{3.39}$$

subject to

$$\{X(t) : t \in [0, T]\} \text{ is admissible, } X(0) = x_0, \mathbb{E}[X(T)] = z,$$

where z is a given parameter representing the expected payoff at the end of the investment horizon, T . An optimal strategy is given explicitly in Theorem 1.1 of Cui et al. (2012), which provides the portfolio at each given time $t \in [0, T]$ as a function of the wealth, a couple of auxiliary feedback processes, and gives estimates (at time t) of the (time-inhomogeneous) diffusion and drift coefficients.

In theory, a continuous-time model requires *continuous* rebalancing of all the times. Naturally, this is not possible (indeed, not necessary) in practice, nor was it attempted in our empirical implementation. In our experiments, we set $T = 1$ (year) and $z = 1.1x_0$ corresponding to an annual expected return $\rho = 10\%$, and we rebalance only monthly (instead of continuously). The one-year period is consistent with the other models under comparison. Therefore, on the first trading day in January of 2000, we estimated all the necessary parameters/coefficients based on the previous ten-year data and then applied the explicit formula for the optimal portfolio,

⁶There is an extensive literature on continuous-time Markowitz models; however, to our best knowledge all of the other existing models include a risk-free asset.

denoted as ϕ_C , given by Theorem 1.1 in Cui et al. (2012). On the first trading day of February 2000, we applied the same formula but with updated estimates of the parameters/coefficients based on the immediate previous ten-year data. In this way, we have constructed a strategy for the whole year of 2000. For all of the subsequent years, we repeat the same procedure to generate a 17-year long strategy ϕ_C and the corresponding wealth process.

It is important to note that this model does not have an explicit no-bankruptcy constraint (i.e., it does not rule out the possibility that the wealth process may go negative during $[0, T]$). Indeed, as we will see in the discussions below, this model led to bankruptcy in *all* of our numerical experiments for portfolios of 100 stocks.⁷ Once a bankruptcy occurred, we considered it “game over” and retained the zero wealth until December 2016.⁸

Fama–French model and Black–Litterman model

Both the Fama–French model and the Black–Litterman model were developed to address the mean-blur problem, namely, the fact that compared with variance, it is much more difficult to estimate within a workable accuracy the expected returns of stocks based purely on sample means. These models estimate the stock returns by their respective methods while keeping the sample covariance matrix and feed them into the classical Markowitz model to obtain the corresponding strategies.

In implementing the Fama–French model, we first downloaded the monthly data of the three factors (i.e., Rm-Rf, SMB, and HML)⁹ from Kenneth French’s personal web-

⁷We have also tested for portfolios with 20 stocks and observed bankruptcy in more than half of our experiments. On the other hand, although the other six single-period models have no explicit no-bankruptcy constraint either, a total of only two instances of bankruptcy occurred in our experiments.

⁸Bielecki et al. (2005b) solved a continuous-time mean–variance model with the no-bankruptcy constraint. However, there is a risk-free asset in that model. To have a fair comparison with the other models, in which there is no risk-free account available, it is proper to choose the model of Cui et al. (2012) in our experiments. We are not aware of a work on a continuous-time Markowitz model without a risk-free asset and with a bankruptcy prohibition.

⁹We assume that the factors have been processed according to the available papers (Fama and French (1992) and Fama and French (1993)).

site.¹⁰ Then, on the first trading day of each month during January 2000–December 2016, for each stock, we used its immediate prior ten-year history returns to fit the three-factor Fama–French model. Next, we plugged in the factors data available on that day to obtain an estimate of the stock’s return for the month. We used these estimates for all of the randomly chosen 100 stocks as the mean vector and solved the single-period classical Markowitz model with $\rho = 10\%$. The generated portfolio was denoted as ϕ_F . This process was then repeated in the subsequent months on a rolling horizon basis.

For the Black–Litterman model, on the first trading day of each month during January 2000–December 2016, we calculated the implied returns of all the S&P 500 constituent stocks having at least ten years’ historical data, using the following formula:

$$R_{implied} = \lambda \Sigma \phi_{market},$$

where $\lambda = 3.07$, Σ was the sample covariance matrix of the previous ten years’ returns of these stocks and ϕ_{market} was the corresponding market portfolio (i.e., ϕ_{market} is a vector whose components add up to 1 and are proportional to the capitalizations of the S&P 500 constituents having at least ten years’ historical data) based on the closing prices of the previous trading day; see Idzorek (2002).¹¹ Then, we picked from $R_{implied}$ the implied returns of the 100 stocks that had been randomly chosen. We inputted these returns and the sample covariance matrix into the classical Markowitz model with $\rho = 10\%$ to obtain the portfolio ϕ_B . This process was repeated for the subsequent months on a rolling horizon fashion.

Goldfarb–Iyengar Robust Model

Goldfarb and Iyengar (2003) consider the following robust Markowitz problem

¹⁰http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

¹¹Here we include only those having at least ten years’ historical data to be consistent with the other models.

with a factor model for the return rate:

$$\begin{aligned} & \text{Minimize } \phi \quad \max_{\mathbf{V} \in S_v, \mathbf{D} \in S_d} \text{Var}[\mathbf{r}^\top \phi] \\ & \text{subject to} \quad \min_{\mu \in S_m} \mathbb{E}[\mathbf{r}^\top \phi] \geq \rho \\ & \quad \mathbf{1}^\top \phi = 1, \end{aligned}$$

where

$$\mathbf{r} = \mu + \mathbf{V}^\top \mathbf{f} + \epsilon, \quad \epsilon \sim N(0, \mathbf{D}),$$

$$S_v = \{V : V = V_0 + W, \|W_i\|_g \leq \rho_i, i = 1, \dots, n\}$$

with W_i being the i th column of W and $\|w\|_g = \sqrt{w^\top G w}$ for some positive definite matrix G ,

$$S_d = \{D : D = \text{diag}(d), d_i \in [d_i^{\min}, d_i^{\max}], i = 1, \dots, n\},$$

$$S_v = \{V : V = V_0 + W, \|W_i\|_g \leq \lambda_i, i = 1, \dots, n\},$$

and

$$S_m = \{\mu : \mu = \mu_0 + \xi, |\xi_i| \leq \gamma_i, i = 1, \dots, n\}.$$

So the uncertainty set of this model is based on vector/matrix distance, as opposed to our uncertainty set, which is defined through the Wasserstein distance between probability measures.

In implementing this model, we followed the instructions in Section 7.2 of Goldfarb and Iyengar (2003). Specifically, we calculated the ten years' sample returns \mathbf{r} of the chosen 100 stocks. Then we used the top five principal components of \mathbf{r} , together with the return data of DJA, NDX, SPC, RUT, and TYX, to be the factor vector \mathbf{f} . By choosing the confidence threshold ω to be 95%, we estimated μ_0 , \mathbf{V}_0 , σ_i^2 , γ_i , G , and λ_i . With the target annual return $\rho = 10\%$ and plugging in all of the parameters from the above steps, we used SeDuMi to solve the SOCP formulation (problem (32))

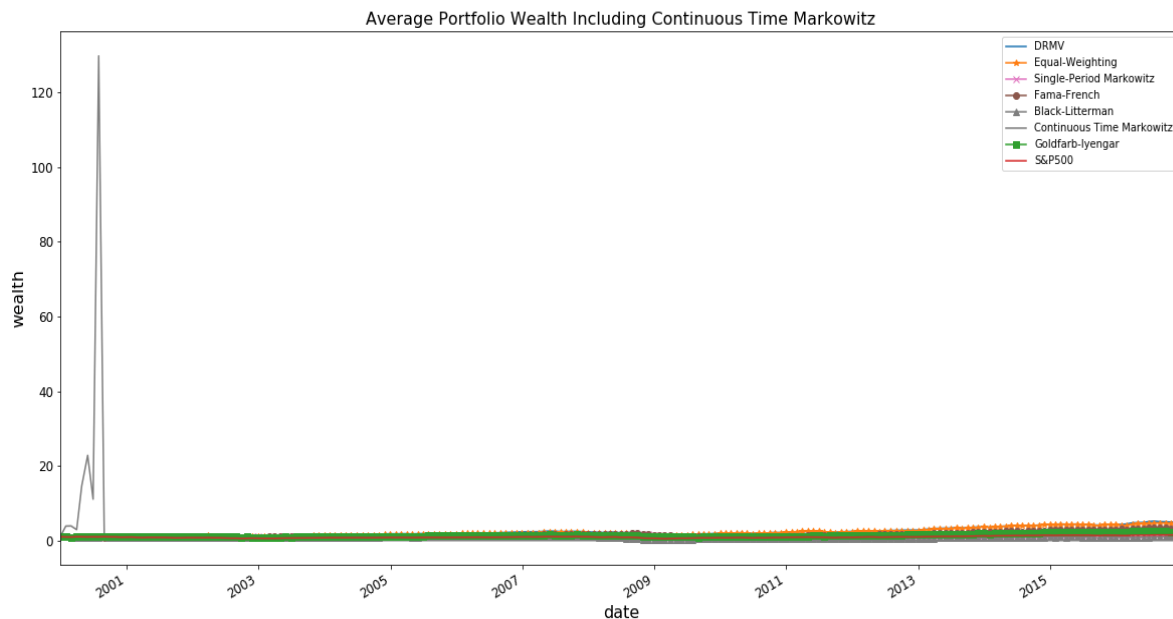


Figure 3.1: This graph presents the wealth processes of all portfolios (including continuous-time Markowitz) and the S&P 500 from January 2000 to December 2016. All of the portfolios except S&P 500 consist of 100 stocks, and the averages are calculated over 100 numerical experiments. The x -axis indicates the time in months (from 1 to 204) and the y -axis indicates the portfolio wealth. Initial wealth is set at 1.

in Goldfarb and Iyengar (2003)) to obtain the portfolio ϕ_G for each month on a rolling horizon basis, starting from January 2000.

3.4.2 Comparisons

Assume that the initial wealth at the start of the backtesting period (i.e., January 2000) is 1. For each randomly selected set of 100 stocks, we generate the wealth process for the period 2000-2016 under each of the six models as described in the previous subsection, as well as that under the equal weighting. Then, we repeat the experiments on 100 such sets of 100 stocks and obtain the *average* realized wealth process for each model. These processes, along with that of the S&P500 (normalized to start from 1 at the start of the testing period), are plotted in Figure 3.1.

Graphically, Figure 3.1 is “corrupted” because of the extreme behavior of the continuous-time Markowitz model. Its average performance went “through the roof” initially and then quickly dived to zero (all of the 100 experiments ended in bankruptcy).

So, the continuous-time Markowitz can be considered an extremely volatile model. This may be explained as follows. The dynamic strategies incorporate considerable feedback effects, which are computed assuming the underlying model is correct. As such, model misspecifications are compounded precisely due to feedback effects. The inclusion of feedback in the optimal dynamic policy, in outputs that are close to typical realizations of the underlying assumed model, result in highly profitable portfolios. On the other hand, even moderate discrepancies from the underlying model dynamics can lead to relatively poor performance. As a consequence, the dynamic model exhibits significantly higher variability than its static-rolling-horizon robust counterpart.

In order to be able to visualize the comparisons among other portfolios, it is necessary to remove the continuous-time Markowitz from Figure 3.1, resulting in Figure 3.2. It is evident that all of the six models except the Black–Litterman (almost) uniformly and substantially outperform S&P500 for the 17-year backtesting period. In terms of the final realized wealth, of the six models, DRMV and equal-weighting dominate over the other three models by a substantial margin. Specifically, Fama–French and Single-period Markowitz lag behind, while Goldfarb–Iyengar stands in between.

The average performance of DRMV and the equal-weighting model are close, although the former outperforms the latter most of the time. This is no surprise, as the latter can be regarded as an extreme case of the distributionally robust model when the uncertainty size $\delta = \infty$, whereas the former has a nearly “optimal” choice of δ informed by the data. We can study more closely the variability of the performance and the overall return–risk efficiency of the two models by examining their histograms of annualized returns (i.e., the distributions of the annualized returns of the 100 experiments) and those of the Sharpe ratios. These data are plotted in Figures 3.3 and 3.4, respectively. Equal-weighting is more concentrated than DRMV, although the difference is not that significant. This indicates that both strategies have stable

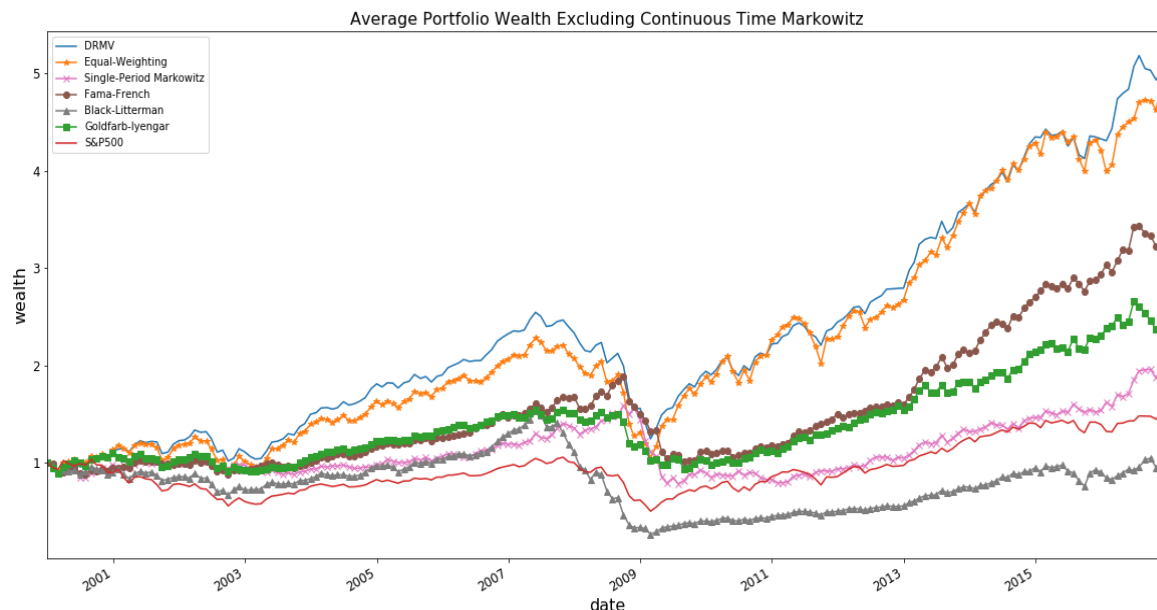


Figure 3.2: This graph presents the wealth processes of all portfolios (excluding continuous-time Markowitz) and of the S&P 500 from January 2000 to December 2016. All of the portfolios except the S&P 500 consist of 100 stocks, and the averages are calculated over 100 numerical experiments. The x -axis indicates the time in months (from 1 to 204) and the y -axis indicates the portfolio wealth. Initial wealth is set to be 1.

performance, although equal-weighting appears to be slightly more stable. On the other hand, DRMV outperforms equal-weighting significantly in terms of Sharpe ratio. We can also compare the histograms of kurtosis for the two models; see Figure 3.5. There is no statistically significant differences between the two: most return distributions under both strategies are platykurtic (i.e., the kurtosis values are less than 3), implying that there are fewer extreme outliers than the standard normal. Overall, we can conclude that both DRMV and equal-weighting are robust and stable, the latter is slightly more so, but the former is markedly superior to the latter in terms of return–risk efficiency.

Similarly, we compare the two histograms for DRMV and Fama–French; see Figures 3.6– 3.8. DRMV has a much more concentrated return histogram indicating a significantly more robust performance, a much more right-shifted Sharpe ratio histogram, and a much more left-shifted kurtosis histogram, implying significantly fewer extreme returns. We can therefore conclude that DRMV compares favorably with

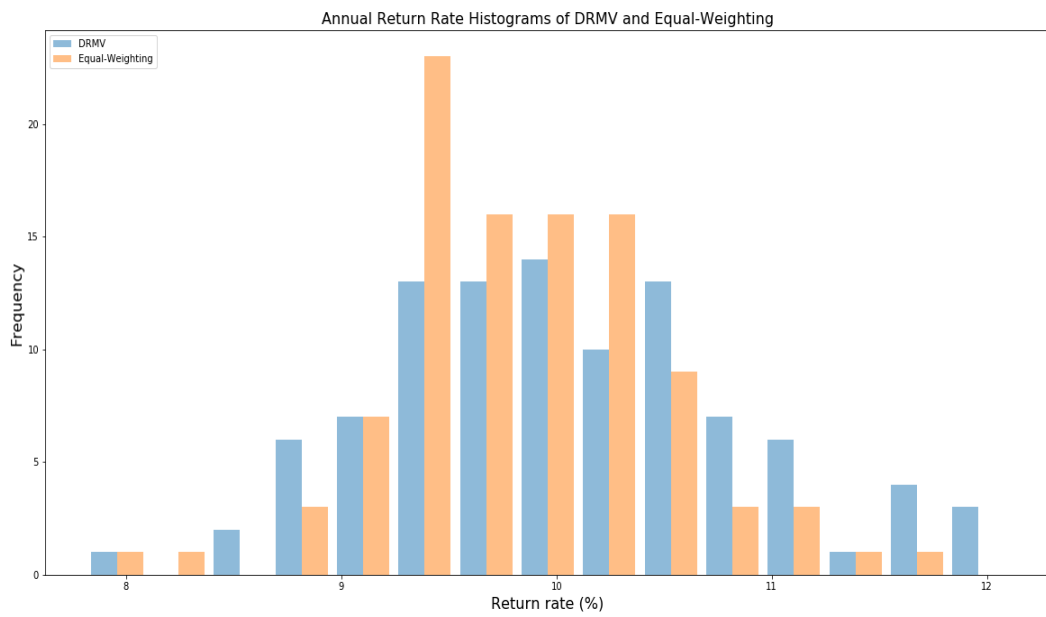


Figure 3.3: This graph presents the histograms of the annualized returns of the 100 different experiments on the DRMV (blue) and equal-weighting (orange) portfolios. The x -axis represents the annualized returns and the y -axis represents the number of returns.

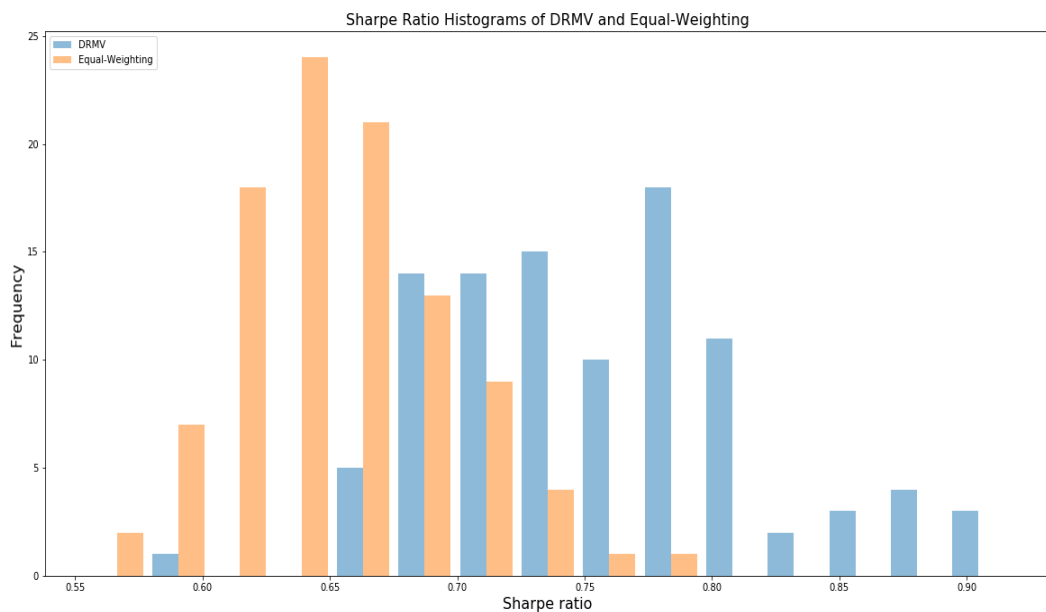


Figure 3.4: This graph presents the histograms of the Sharpe ratio of the 100 different experiments on the DRMV (blue) and equal-weighting (orange) portfolios. The x -axis represents the Sharpe ratio and the y -axis represents the number of Sharpe ratios.

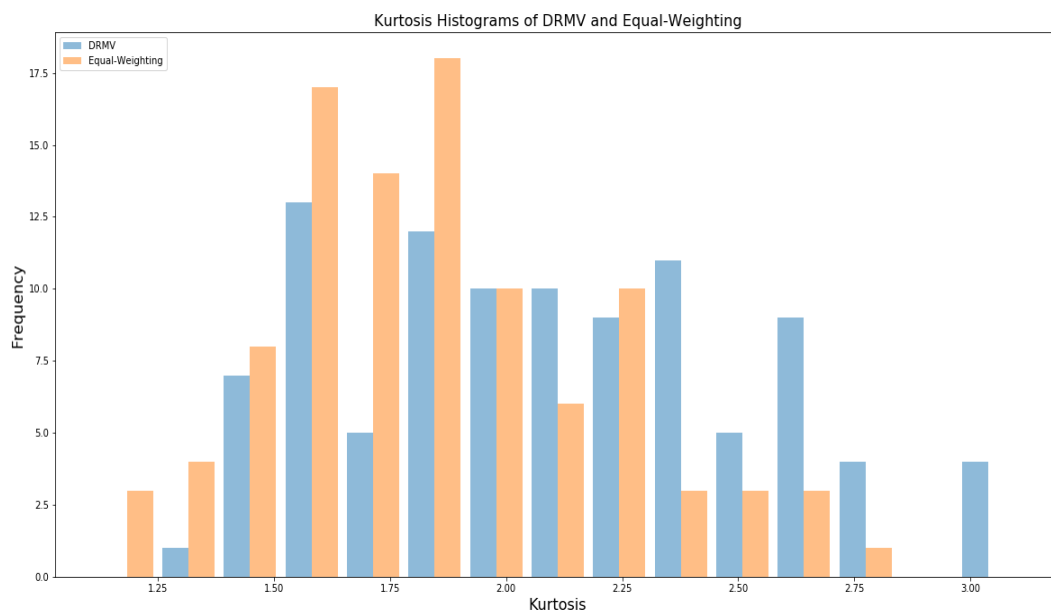


Figure 3.5: This graph presents the histograms of the kurtosises of the 100 different experiments on the DRMV (blue) and equal-weighting (orange) portfolios. The x -axis represents the kurtosis and the y -axis represents the number of kurtosises.

Fama–French in all of the key metrics reported. However, it is interesting to note that DRMV utilizes *only* the price data, whereas Fama–French requires additional fundamental information on the companies concerned.

As for the robust portfolio model developed by Goldfarb and Iyengar (2003), we note that it has a reasonably concentrated return histogram (Figure 3.9), indicating robustness. However, DRMV’s returns are not only more concentrated, but also distributed more to the right than Goldfarb–Iyengar’s. Together with the Sharpe ratio histogram (Figure 3.10), the kurtosis histogram (Figure 3.11), and the average wealth comparison (Figure 3.2), this clearly demonstrates that our uncertainty set formulation based on Wasserstein distance is a significant improvement over the one using the matrix/vector distance.

Finally, we provide comparisons between the histograms for DRMV and single-period Markowitz, as well as between DRMV and Black–Litterman; see Figures 3.12–3.17. Clearly, DRMV has far superior performance in all metrics.

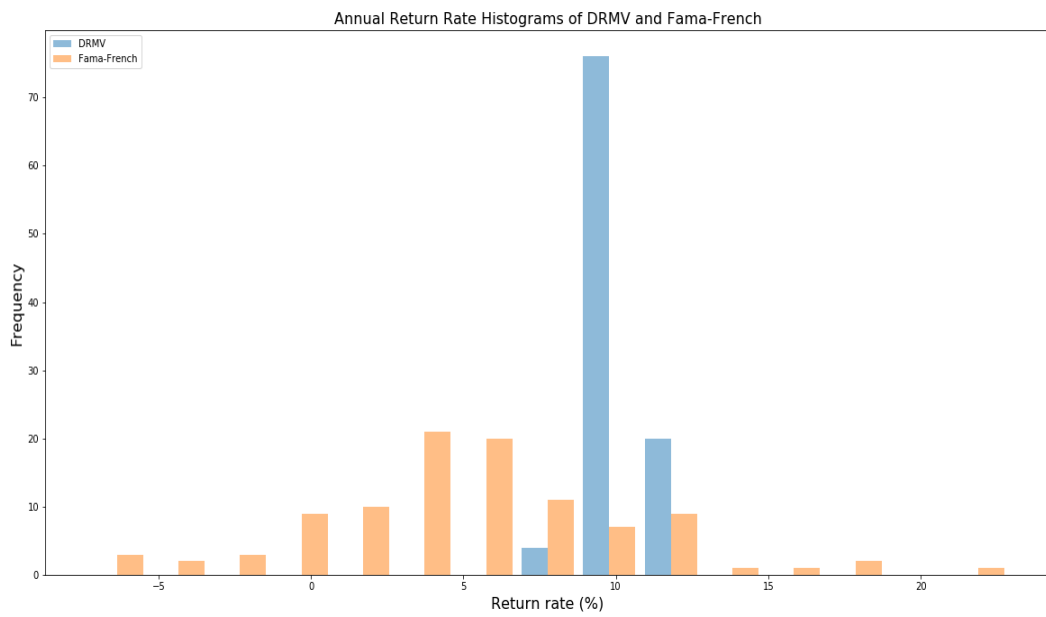


Figure 3.6: This graph presents the histograms of the annualized returns of the 100 different experiments on DRMV (blue) and Fama-French (orange) portfolios. The x -axis represents the annualized returns and the y -axis represents the numbers of returns.

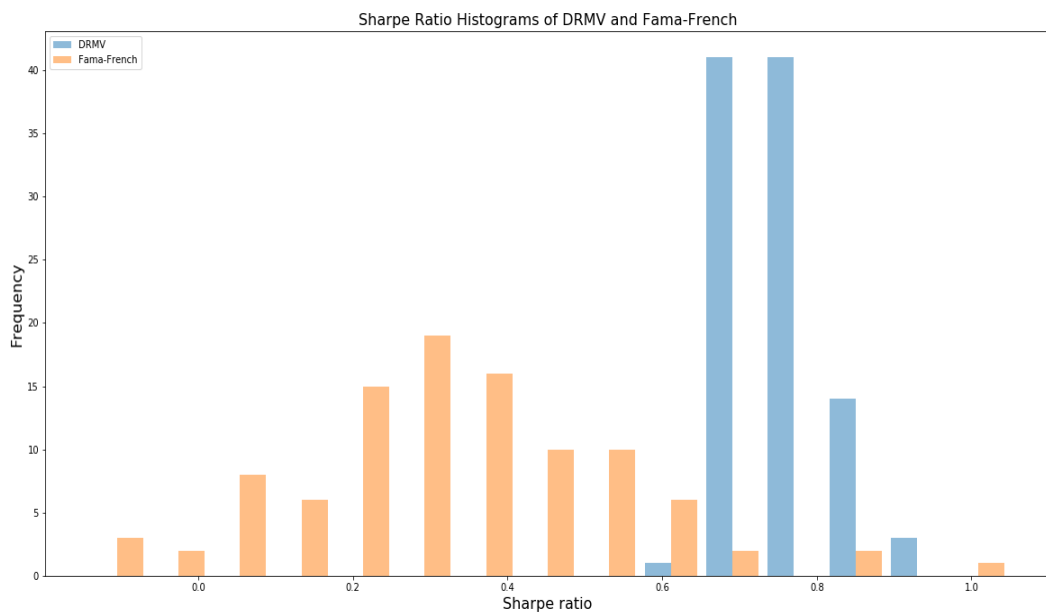


Figure 3.7: This graph presents the histograms of the Sharpe ratio of the 100 different experiments on the DRMV (blue) and Fama-French (orange) portfolios. The x -axis represents the Sharpe ratio and the y -axis represents the number of Sharpe ratios.

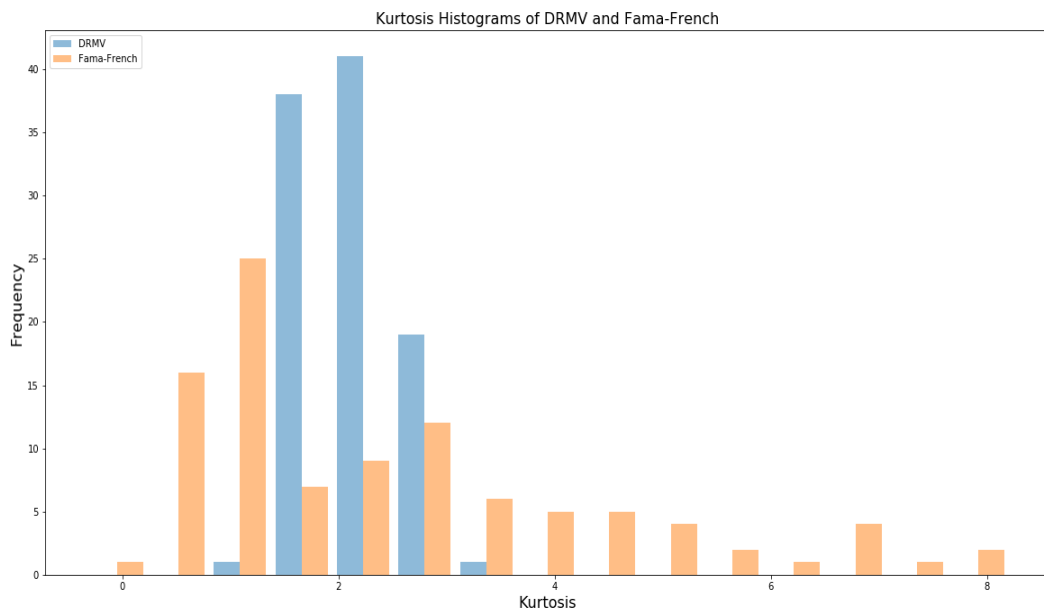


Figure 3.8: This graph presents the histograms of the kurtosises of the 100 different experiments on the DRMV (blue) and Fama-French (orange) portfolios. The x -axis represents the kurtosis and the y -axis represents the number of kurtosises.

3.4.3 Discussion

In this subsection we offer discussions of a variety of issues related to our empirical experiments.

Other values of the Wasserstein order

In all of the previously reported experiments, we set the order of the Wasserstein distance to be $p = 2$. We have also tried different values of p , notably when $p = 1$ (and hence $q = \infty$), and found that the impact on the performance of the corresponding portfolios is minimal. As an example, Figure 3.18 shows a comparison of the average wealth processes under $p = 1$ and $p = 2$ for portfolios of $d = 100$ stocks.

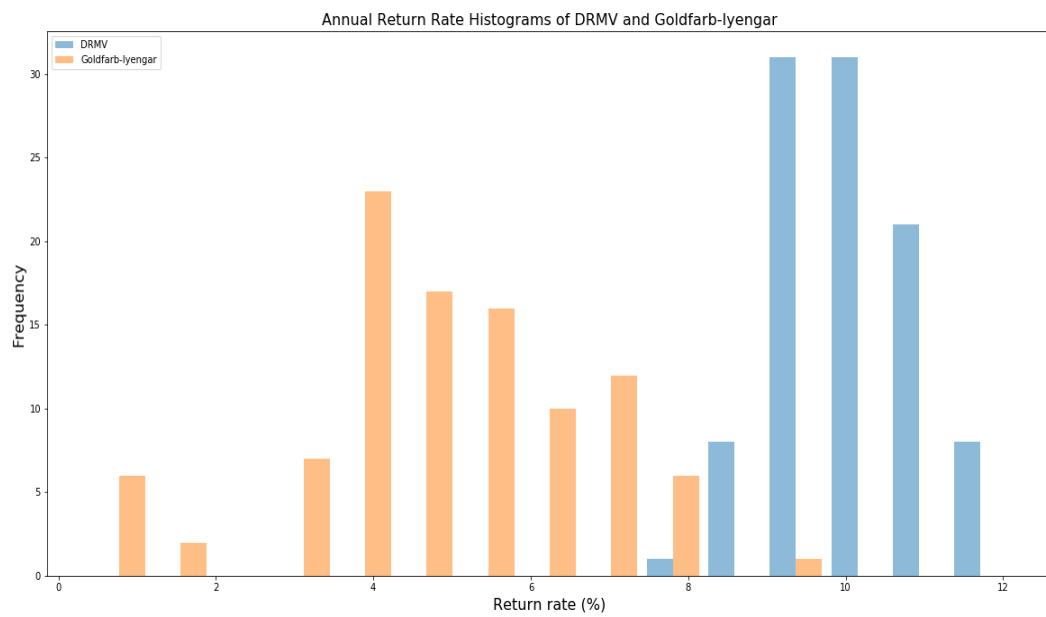


Figure 3.9: This graph presents the histograms of the annualized returns of the 100 different experiments on the DRMV (blue) and Goldfarb–Iyengar (orange) portfolios. There are two experiments in which Goldfarb–Iyengar went into bankruptcy, which are not included in this histogram. The x -axis represents the annualized returns and the y -axis represents the number of returns.

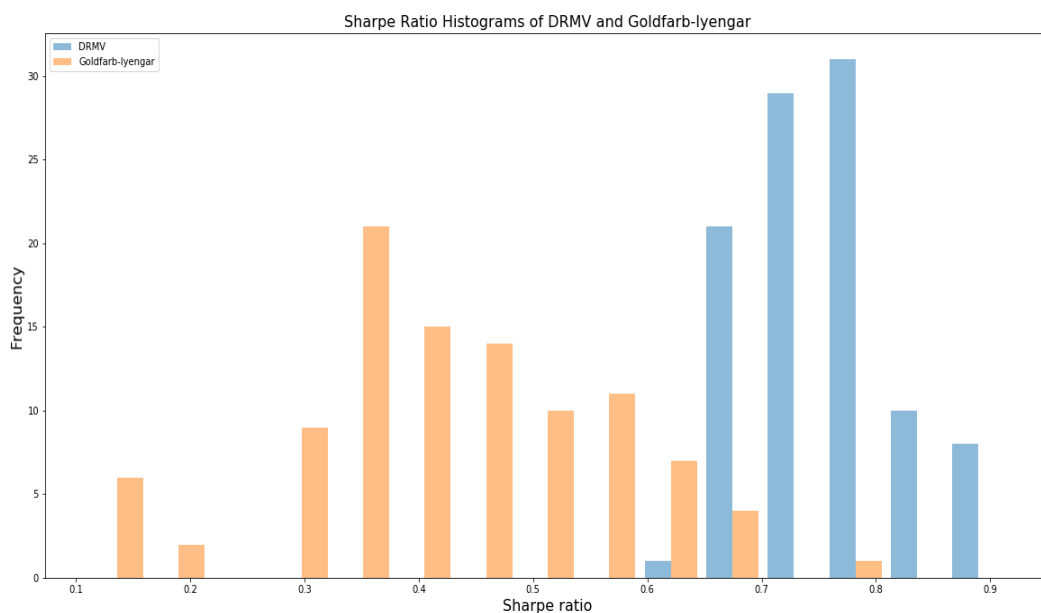


Figure 3.10: This graph presents the histograms of the Sharpe ratios of the 100 different experiments on the DRMV (blue) and Goldfarb–Iyengar (orange) portfolios. The x -axis represents the Sharpe ratios and the y -axis represents the number of Sharpe ratios.

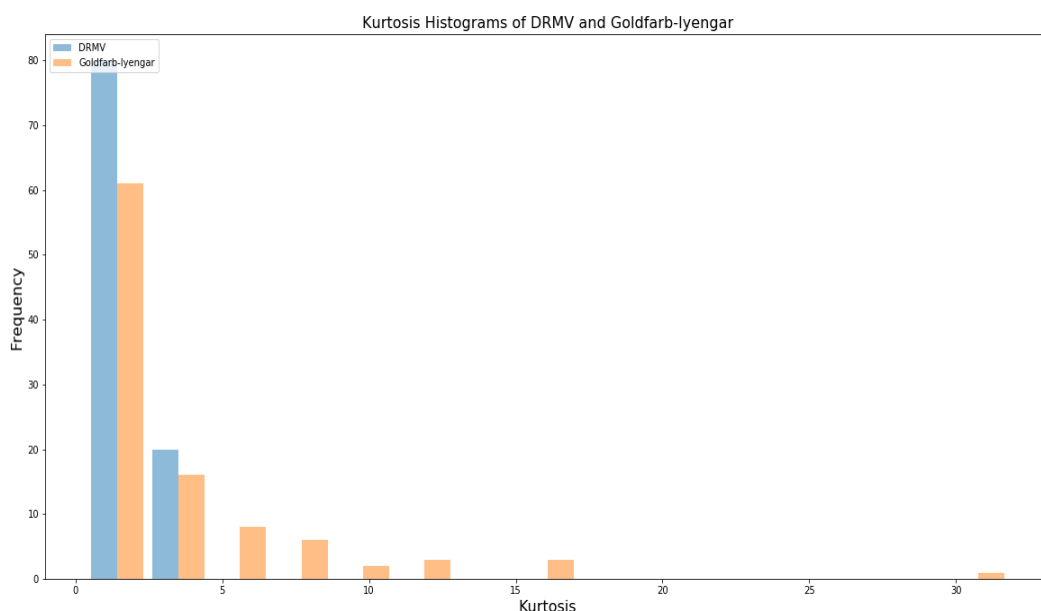


Figure 3.11: This graph presents the histograms of the kurtosises of the 100 different experiments on the DRMV (blue) and Goldfarb–Iyengar (orange) portfolios. The x -axis represents the kurtosis and the y -axis represents the number of kurtosises.

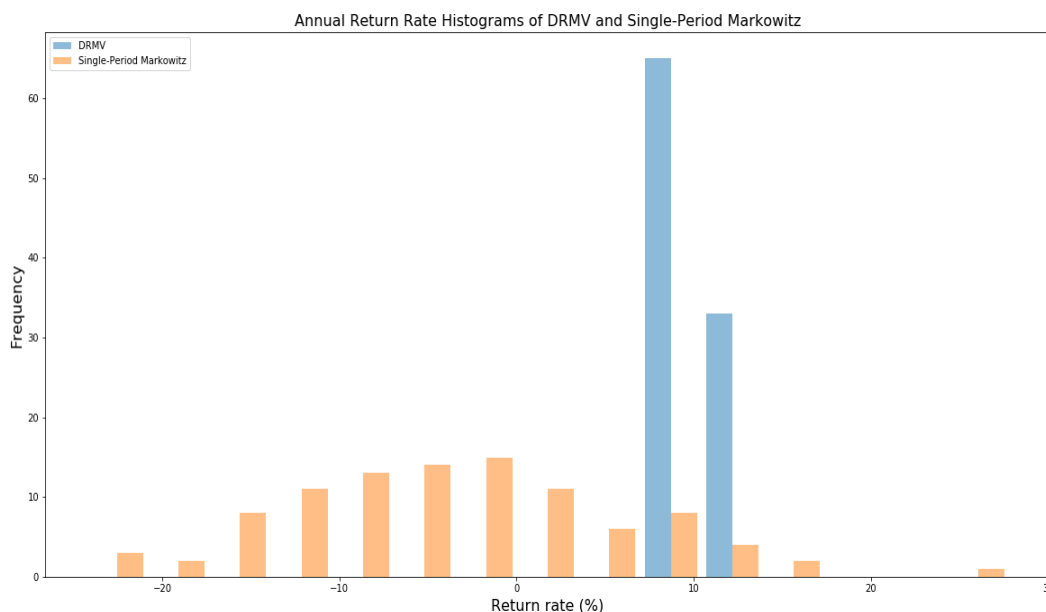


Figure 3.12: This graph presents the histograms of the annualized returns of the 100 different experiments on the DRMV (blue) and single-period Markowitz (orange) portfolios. The x -axis represents the annualized returns and the y -axis represents the number of returns.

Different targeted returns

We have also tested different (plausible/reasonable) values of the targeted return $\rho = 5\%$, 15% , 20% in addition to $\rho = 10\%$, and found that DRMV maintains the same outperformance with respect to other models and, indeed, some other models perform less well under higher targets. Figure 3.19 plots DRMV's average wealth processes under these four values of ρ when $d = 100$.¹² It is evident that the average performance is very robust with different ρ 's, although the 10% strategy slightly outperforms the others (which was the reason why we chose $\rho = 10\%$ for our main experiments). This, in turn, suggests that the choice of a specific value of ρ is unimportant for DRMV so long as it is in the reasonable range of $[5\%, 20\%]$, thereby releasing us from the need to tune or calibrate of this parameter.

¹²The plots under $\rho = 5\%$, 15% , 20% are almost identical so one could probably see only two plots in Figure 3.19.

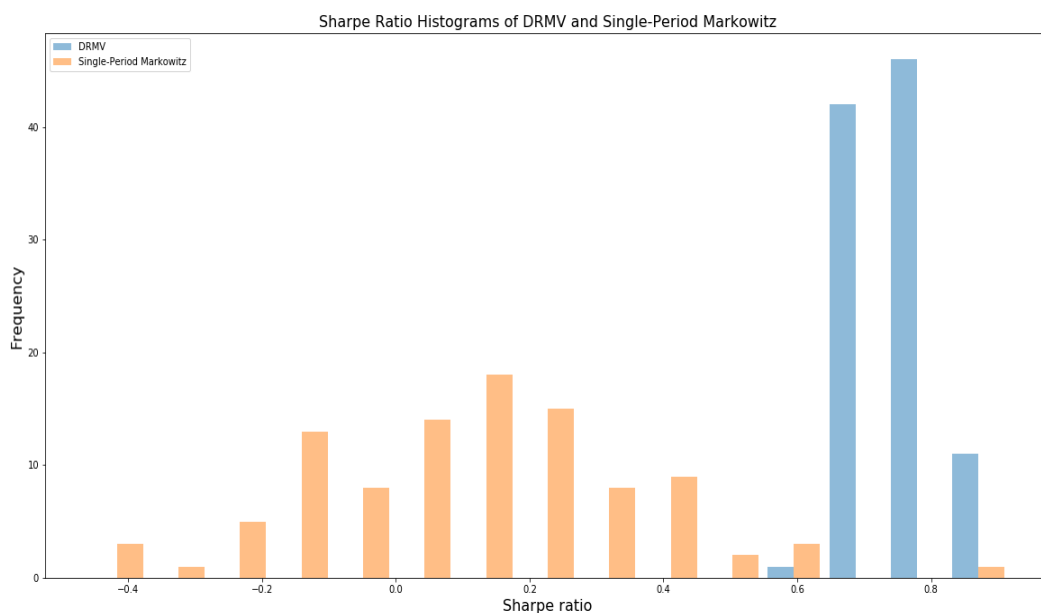


Figure 3.13: This graph presents the histograms of the Sharpe ratios of the 100 different experiments on the DRMV (blue) and single-period Markowitz (orange) portfolios. The x -axis represents the Sharpe ratios and the y -axis represents the number of Sharpe ratios.

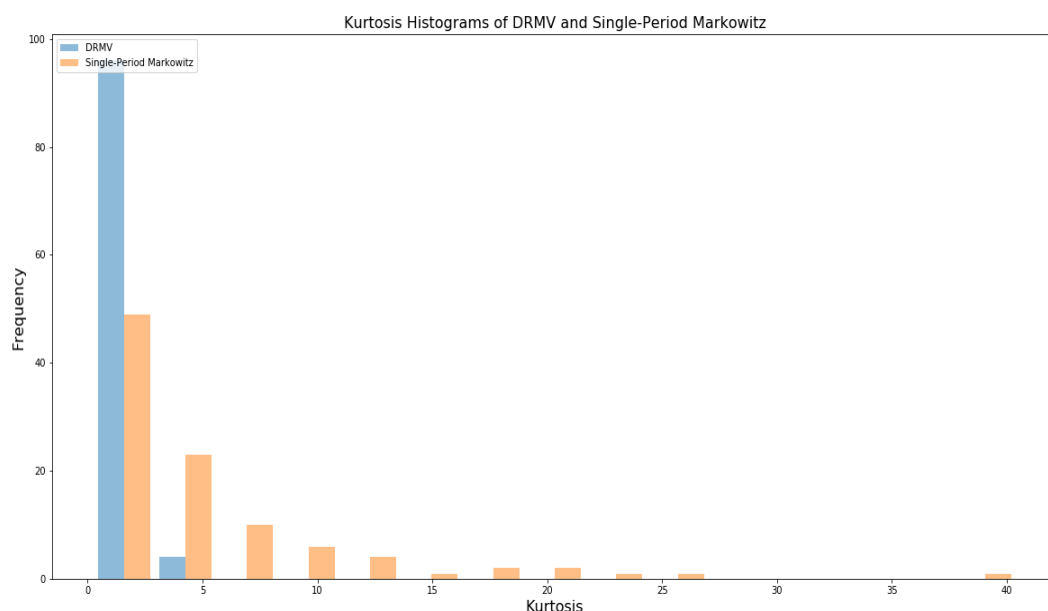


Figure 3.14: This graph presents the histograms of the kurtosis values of the 100 different experiments on the DRMV (blue) and single-period Markowitz (orange) portfolios. The x -axis represents the kurtosis and the y -axis represents the number of kurtosis values.

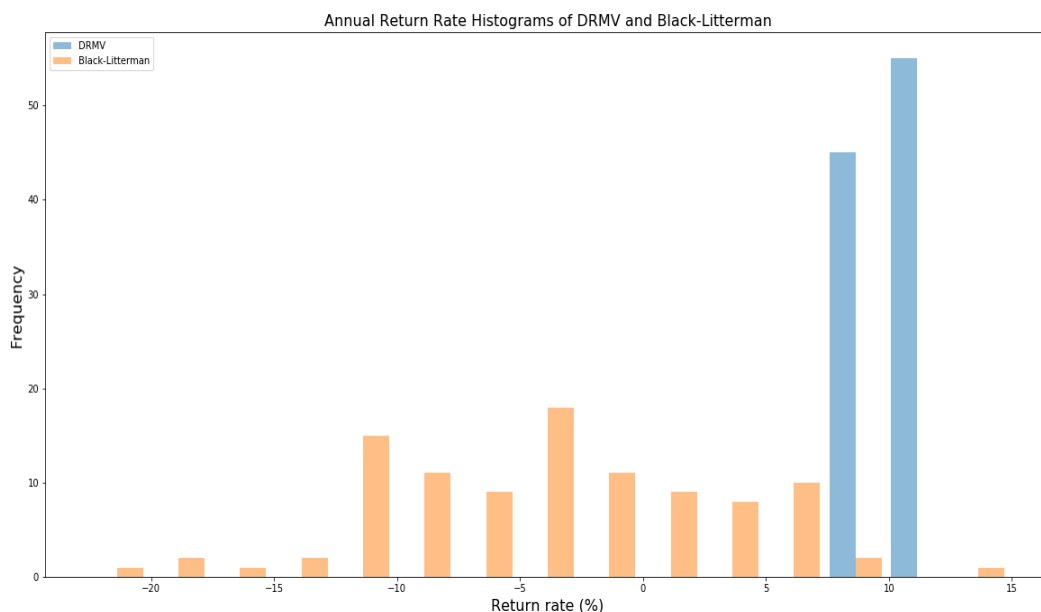


Figure 3.15: This graph presents the histograms of the annualized returns of the 100 different experiments on the DRMV (blue) and Black-Litterman (orange) portfolios. The x -axis represents the annualized returns and the y -axis represents the number of returns.

Turnover rates and transactions costs

Fabozzi et al. (2007) observe empirically that robust portfolios have low turnover rates. We reexamine this with our distributionally robust strategies. Figure 3.20 presents the histogram of the turnover rates (including buy and sell) with 100 experiments for $d = 100$ stocks.

Clearly, our result reconciles with the finding of Fabozzi et al. (2007). Indeed, most of the monthly turnover rates are lower than 5%, which is considered to be very good and reasonable in practice. If we take the average monthly turnover rate to be 5% (definitely an upper bound for the average), then the annual turnover rate is around 60%. Now, assuming that the proportional transaction cost is 0.02%, then each year the transaction cost is around $60\% \times 0.02\% = 0.012\%$, which is very small and ignorable.

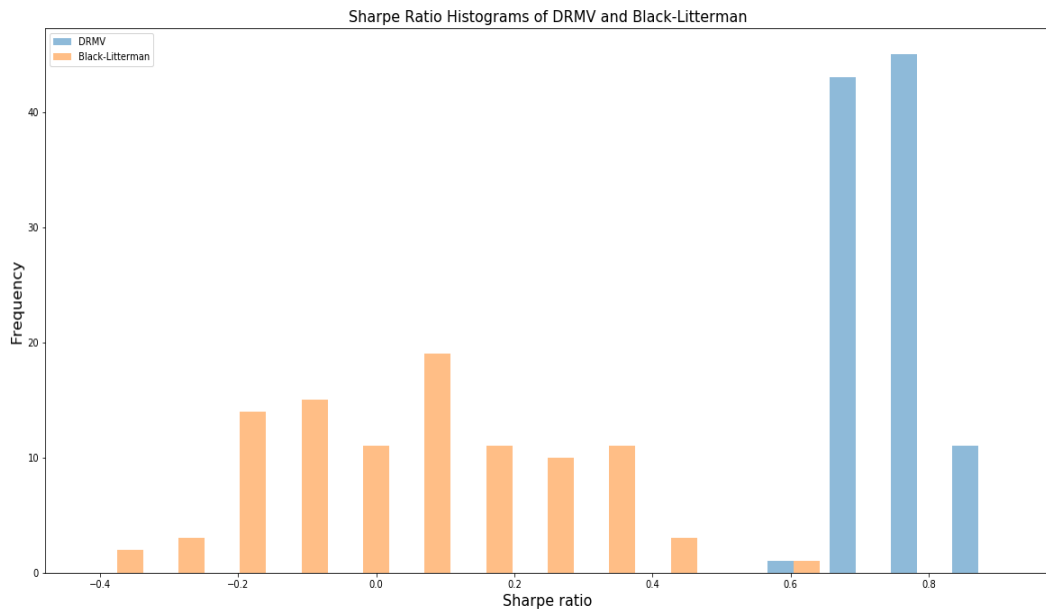


Figure 3.16: This graph presents the histograms of the Sharpe ratios of the 100 different experiments on the DRMV (blue) and Black-Litterman (orange) portfolios. The x -axis represents the Sharpe ratios and the y -axis represents the number of Sharpe ratios.

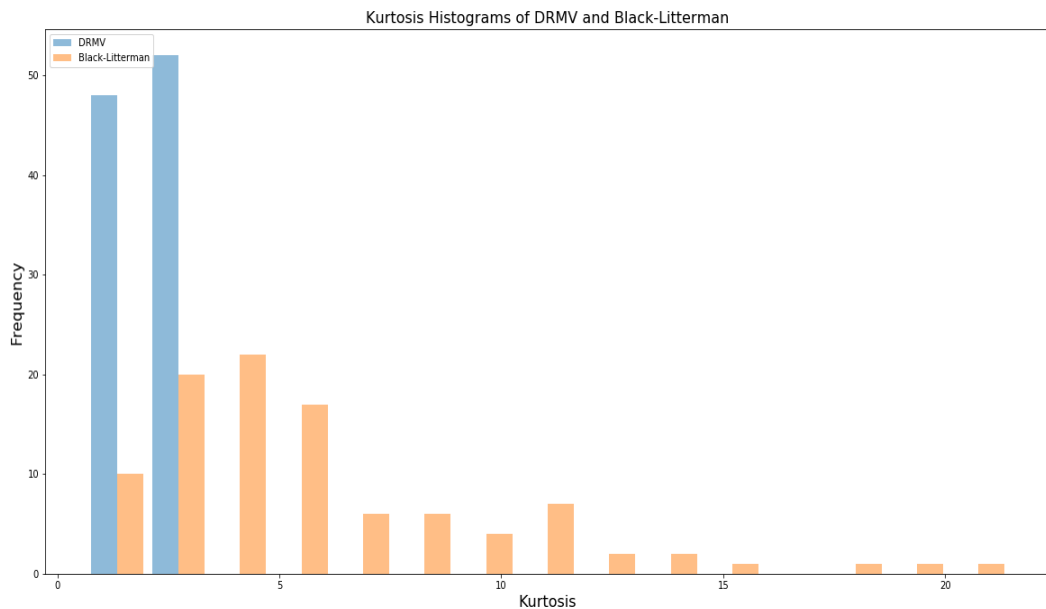


Figure 3.17: This graph presents the histograms of the kurtosises of the 100 different experiments on the DRMV (blue) and Black-Litterman (orange) portfolios. The x -axis represents the kurtosis and the y -axis represents the number of kurtosises.

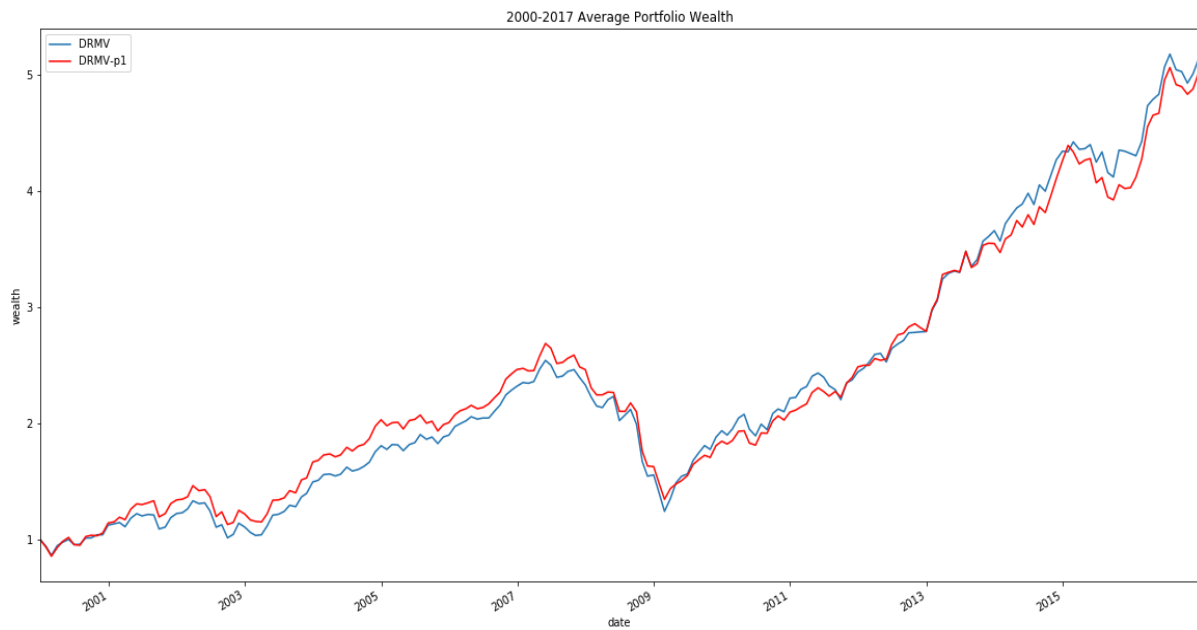


Figure 3.18: This graph presents the portfolio average wealth processes of DRMV ($p = 2$) and DRMV-p1 ($p = 1$) from January 2000 to December 2016. The averages are calculated over 100 numerical experiments. The x -axis indicates the time and the y -axis indicates the portfolio wealth. Initial wealth is set at 1.

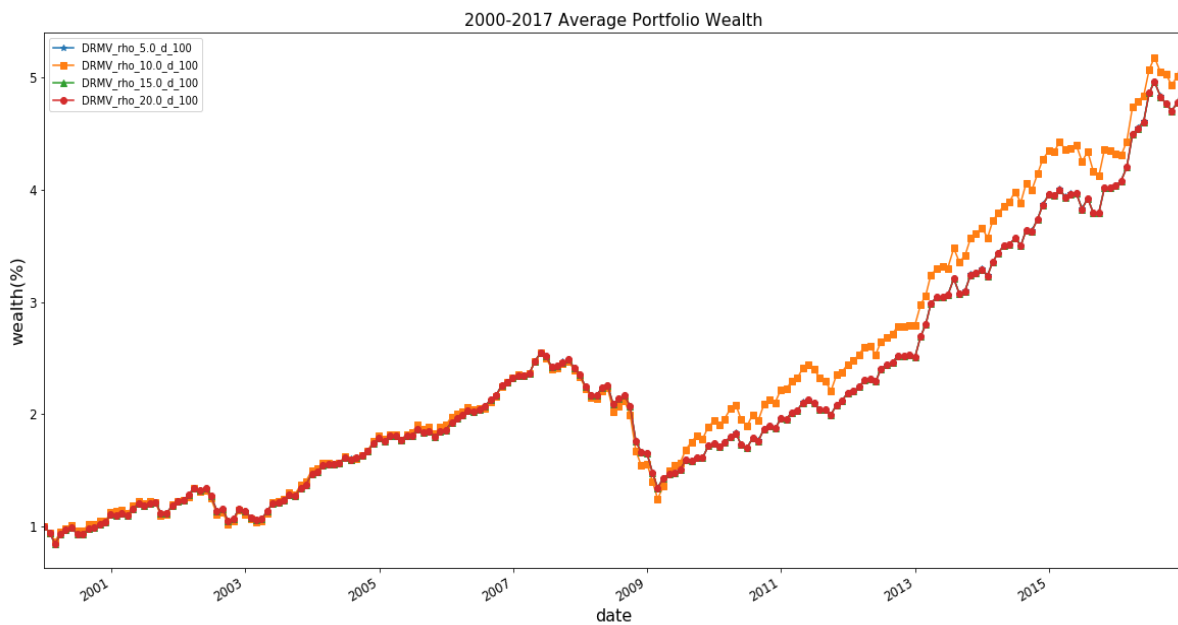


Figure 3.19: This graph presents DRMV's average wealth processes from January 2000 to December 2016 with different values of ρ . The averages are calculated over 100 numerical experiments. The x -axis indicates the time and the y -axis indicates the portfolio wealth. Initial wealth is set at 1.

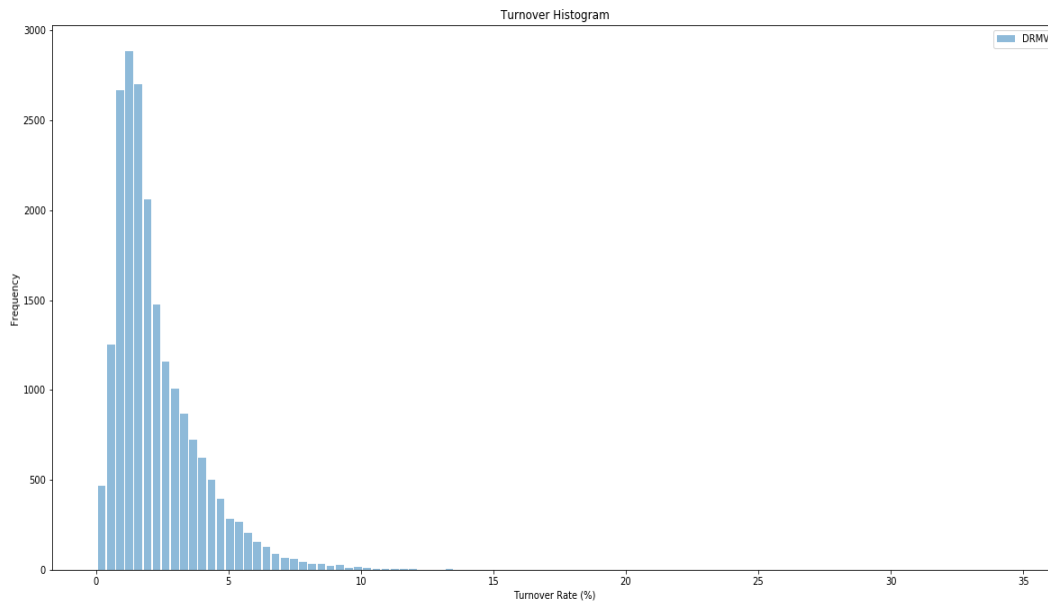


Figure 3.20: This graph presents the histogram of the DRMV monthly turnover rates of 100 experiments. The x -axis represents the turnover rate (%) and the y -axis represents the numbers of turnover rates.

Shrinkage estimators

In all of the experiments reported so far, sample covariance was used for the single-period Markovitz model as well as for the Fama–French and Black–Litterman models. We tested using the shrinkage covariance matrices for these three models.

We used the shrinkage estimator of Ollila and Raninen (2018):

$$\Sigma_{\alpha,\beta} = \beta\Sigma_n + \alpha I_n, \quad (3.40)$$

where Σ_n is the sample covariance matrix and I_n is the $n \times n$ identity matrix. The parameters α and β are estimated by the following:

$$\hat{\alpha} = (1 - \hat{\beta})\hat{\eta}, \quad \hat{\beta} = \frac{(\hat{\gamma} - 1)}{(\hat{\gamma} - 1) + \hat{\kappa}(2\hat{\gamma} + d)/n + (\hat{\gamma} + d)/(n - 1)},$$

where d is the number of stocks, $\hat{\eta} = \frac{\text{tr}(\Sigma_n)}{d}$, $\hat{\gamma} = \frac{nd}{n-1}[\text{tr}(\Sigma_{sgn}^2) - \frac{1}{n}]$ with $\Sigma_{sgn} =$

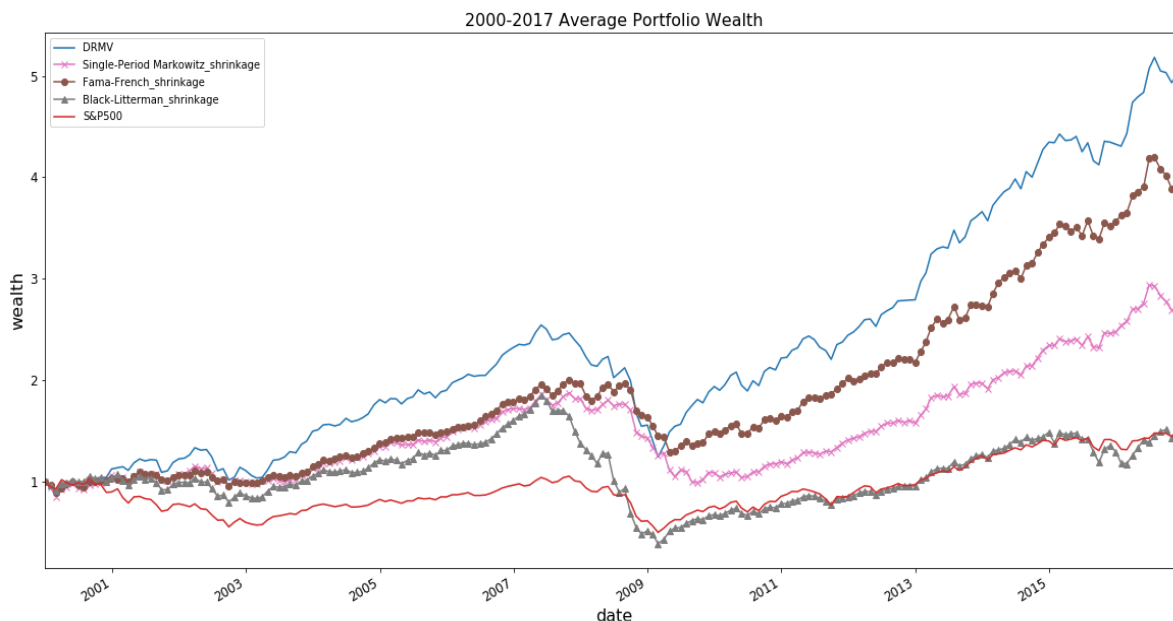


Figure 3.21: This graph presents portfolios’ average wealth processes using shrinkage estimators from January 2000 to December 2016. All the portfolios except S&P 500 consist of 100 stocks and the averages are calculated over 100 numerical experiments. The x -axis indicates the time in months (from 1 to 204) and the y -axis indicates the portfolio wealth. Initial wealth is set to be 1.

$$\frac{1}{n} \sum_{i=1}^n \frac{(R_i - \hat{\mu})(R_i - \hat{\mu})^T}{\|R_i - \hat{\mu}\|}, \text{ and } \hat{\mu} = \arg \min_{\mu} \sum_{i=1}^n \|R_i - \mu\|, \hat{\kappa} = \max \left(-\frac{2}{d+2}, \frac{1}{3d} \sum_{j=1}^d \hat{K}_j \right)$$

with $\hat{K}_j = \frac{n-1}{(n-2)(n-3)} [(n+1)\hat{k}_j + 6]$ and $\hat{k}_j = \frac{m_j^{(4)}}{(m_j^{(2)})^2} - 3$, $m_j^{(q)}$ denoting the q th order sample moment.

The three models, single-period Markowitz, Fama–French, and Black–Litterman all improved with the shrinkage estimators. Figure 3.21 presents the comparison of the average wealth processes among these three models, DRMV, and the S&P 500. Notably, the Black–Litterman model now outperforms the S&P 500 most of the time, as opposed to its underperformance without shrinkage (see Figure 3.2). However, DRMV still dominates all the three models. Histograms of return distributions and Sharpe ratio distributions show the superiority of DRMV over the other models, similar to the case without shrinkage.¹³

¹³As these histograms are similar, we do not present them here.

3.5 Concluding Remarks

We have provided a data-driven distributionally robust theory for Markowitz's mean–variance portfolio selection. The robust model can be solved via a non-robust one based on the empirical probability measure with an additional regularization term. The size of the distributional uncertainty region is not exogenously given; rather, it is informed by the return data in a scheme that we have developed in this chapter.

Our results may be generalized in several directions. We have chosen the l_q norm in defining our Wasserstein distance due to its popularity in regularization, but other transportation costs can be used. For example, one may consider the type of transportation cost related to adaptive regularization which has been studied by Blanchet et al. (2017), or the one related to industry cluster, as in Blanchet and Kang (2017). Another significant direction is a dynamic (discrete-time or continuous-time) version of the DRMV model.

Chapter 4

A Distributionally Robust Sharpe Ratio Model

4.1 Model Formulation

In this section, we formulate our distributionally robust Sharpe ratio (DRSR) model while reviewing some useful concepts.

The background setting in this chapter is similar to that of Chapter 3. For the readers' convenience, we first revisit the related notations and definitions. Let $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ be the space of all Borel probability measures supported on $\mathbb{R}^d \times \mathbb{R}^d$. A given element $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ can be assumed to be the joint distribution of a random vector (U, V) , where $U \in \mathbb{R}^d$ and $V \in \mathbb{R}^d$. We use π_U and π_V to denote the marginal distributions of U and V under π . In particular, $\pi_U(A) = \pi(A \times \mathbb{R}^d)$ and $\pi_V(A) = \pi(\mathbb{R}^d \times A)$ for every Borel set $A \subset \mathbb{R}^d$.

We start with a “cost” function $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, \infty]$, which we assume to be lower semicontinuous and such that $c(u, u) = 0$ for any $u \in \mathbb{R}^d$. For a given such cost function c , we introduce $D_c(\cdot, \cdot)$ to represent some “discrepancy” between two probability measures as follows:

$$D_c(\mathbb{P}, \mathbb{Q}) := \inf\{\mathbb{E}_\pi[c(U, V)] : \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), \pi_U = \mathbb{P}, \pi_V = \mathbb{Q}\}, \quad (4.1)$$

where \mathbb{P} and \mathbb{Q} are two probability measures supported on \mathbb{R}^d . This can be interpreted as the optimal (minimal) transportation cost (also known as the optimal transport discrepancy or the Wasserstein discrepancy) of moving the mass from \mathbb{P} into the mass of \mathbb{Q} under a cost $c(x, y)$ per unit of mass transported from x to y .

If for a given $p > 0$, $c^{1/p}(\cdot, \cdot)$ is a metric, then so is $D_c^{1/p}(\cdot, \cdot)$; see Villani (2003). Such a metric $D_c^{1/p}(\cdot, \cdot)$ is known as a Wasserstein distance of the order p . For the most part in this chapter, we choose the cost function

$$c(u, v) = \|u - v\|_q^2 \quad (4.2)$$

where $q \geq 1$ is fixed, which leads to a Wasserstein distance of the order 2.

Recall that R is the d -dimensional vector of random returns of the d stocks. Let \mathbb{P}_n be the empirical probability measure on \mathbb{R}^d with a sample size n , i.e.,

$$\mathbb{P}_n(dx) = \frac{1}{n} \sum_{i=1}^n \delta_{R_i}(dx),$$

where R_i ($i = 1, 2, \dots, n$) are realizations of R and $\delta_{R_i}(\cdot)$ is the indicator function. Let us first review the formulation for the classical Sharpe ratio portfolio selection model.

It is formulated as

$$\min_{\phi \in \mathcal{F}} \left\{ \frac{\sqrt{\phi^\top \text{Var}_{\mathbb{P}_*}(R) \phi}}{\phi^\top \mathbb{E}_{\mathbb{P}_*}[R]} : \phi^\top \mathbf{1} = 1 \right\}, \quad (4.3)$$

where R is the d -dimensional vector of random returns of the stocks; \mathbb{P}_* is the probability measure underlying the distribution of R ; $\mathbb{E}_{\mathbb{P}_*}$ and $\text{Var}_{\mathbb{P}_*}$ are respectively the expectation and variance under \mathbb{P}_* ; and the feasible region is $\mathcal{F} = \{\phi : \phi \in \mathbb{R}^d, \phi^\top \mathbb{E}_{\mathbb{P}_*}[R] > 0\}$. Note that the objective function in (4.3) is the inverse of the portfolio Sharpe ratio and that the problem is an equivalent formulation of the Sharpe ratio maximization problem. The formulation (4.3) will be more convenient for us to drive the tractability of DRSR, which is formulated later in this section. We have a positive return constraint in the feasible region because we want to focus solely on the port-

folios whose returns are positive. If the portfolio return is negative, it is not ideal to maximize the Sharpe ratio since this boils down to maximizing the portfolio variance, which is certainly not reasonable.

Define the ambiguity set as

$$\mathcal{U}_\delta(\mathbb{P}_n) = \{\mathbb{P} : D_c(\mathbb{P}, \mathbb{P}_n) \leq \delta\}$$

and the feasible region of portfolios as

$$\mathcal{F}(n) = \{\phi : \phi^\top \mathbf{1} = 1, \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] > 0\}.$$

The DRSR is formulated as follows:

$$\min_{\phi \in \mathcal{F}(n)} \max_{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)} \left\{ \frac{\sqrt{\phi^\top \text{Var}_{\mathbb{P}}(R) \phi}}{\phi^\top \mathbb{E}_{\mathbb{P}}[R]} \right\}. \quad (4.4)$$

The inner maximization part represents the inverse of the worst case Sharpe ratio. The objective is to choose a portfolio $\phi \in \mathcal{F}(n)$ that achieves the optimal worst case value in (4.4).

Here, it is important to emphasize that, in the definition of the feasible region, there is no ambiguity set involved. Specifically, the target constraint in $\mathcal{F}(n)$ is $\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] > 0$ instead of $\min_{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)} \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] > 0$. This is due to the fact that if for some portfolio ϕ , $\min_{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)} \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] \leq 0$, the optimal value of the inner maximization part in (4.4) will be $+\infty$, which means that this ϕ cannot be the optimal solution for (4.4) and will be eliminated directly. Also, if the empirical portfolio return is positive, the optimal value of the inner maximization problem (4.4) will never be negative. Note that, there is no target return ρ in the formulation of (4.4), which is different from that of the DRMV. In DRSR, we do not need to worry about what the optimal choice of target return should be, it is decided automatically by the data.

4.2 Transformations and Tractability

Problem (4.4) appears to be complicated. Similar to (3.3) in Chapter 3, the inner maximization problem of (4.4) is over a set of probability measures, which means that it is an infinite dimensional optimization problem. Also, it is not clear whether the outer minimization problem, while finite dimensional, is convex. In this section, we reformulate (4.4) such that it becomes computationally tractable.

First, by fixing $\mathbb{E}_{\mathbb{P}}(\phi^{\top} R) = \alpha > 0$ in the inner maximization part of problem (4.4) we obtain the following equivalent formulation:

$$\min_{\substack{\phi: \phi^{\top} \mathbf{1}=1, \\ \phi^{\top} \mathbb{E}_{\mathbb{P}_n}[R] > 0}} \max_{\substack{\mathbb{P}: \mathbb{P} \in \mathcal{U}_{\delta}(\mathbb{P}_n), \\ \phi^{\top} \mathbb{E}_{\mathbb{P}}[R] = \alpha, \\ \alpha: \alpha > 0}} \frac{\sqrt{\text{Var}_{\mathbb{P}}(\phi^{\top} R)}}{\alpha}. \quad (4.5)$$

Note that, with respect to the inner maximization problem, we only consider $\alpha > 0$. This is due to the constraint $\phi^{\top} \mathbb{E}_{\mathbb{P}_n}[R] > 0$ in the outer minimization problem. To be more precise, this constraint means that there is at least one probability measure (e.g., \mathbb{P}_n) in the ambiguity set $\mathcal{U}_{\delta}(\mathbb{P}_n)$, such that $\phi^{\top} \mathbb{E}_{\mathbb{P}}[R] > 0$. Any probability measure satisfying $\phi^{\top} \mathbb{E}_{\mathbb{P}}[R] \leq 0$ is not a candidate for the optimal solution. Therefore, we can skip the case of $\alpha \leq 0$. Then, the inner maximization problem (4.5) is equivalent to the following problem:

$$\max_{\alpha: \alpha > 0} \max_{\substack{\mathbb{P}: \mathbb{P} \in \mathcal{U}_{\delta}(\mathbb{P}_n), \\ \phi^{\top} \mathbb{E}_{\mathbb{P}}[R] = \alpha}} \frac{\sqrt{\text{Var}_{\mathbb{P}}(\phi^{\top} R)}}{\alpha}. \quad (4.6)$$

In the above problem, we first optimize over \mathbb{P} . This can be done by means of the following proposition, which is based on Proposition 7 in Chapter 3.

Proposition 9 *The formulation given in (4.6) is equivalent to*

$$\max_{\alpha \in \mathcal{H}(\delta, n)} \frac{1}{\alpha} (\sqrt{\phi^{\top} \text{Var}_{\mathbb{P}_n}(R) \phi} + \sqrt{\delta \|\phi\|_p^2 - (\alpha - \phi^{\top} \mathbb{E}_{\mathbb{P}_n}[R])^2}), \quad (4.7)$$

where $\mathcal{H}(\delta, n) := \{\alpha : \alpha > 0, \delta \|\phi\|_p^2 - (\alpha - \phi^{\top} \mathbb{E}_{\mathbb{P}_n}[R])^2 \geq 0\}$, in the sense that the two

problems have the same optimal solution and optimal value.

Proof Formulation (4.6) is equivalent to

$$\max_{\alpha > 0} \max_{\substack{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n), \\ \phi^\top \mathbb{E}_{\mathbb{P}}[R] = \alpha}} \frac{\sqrt{\phi^\top \mathbb{E}_{\mathbb{P}}[RR^\top] \phi - \alpha^2}}{\alpha}. \quad (4.8)$$

By Proposition 7 in Chapter 3, we know that (4.6) is equivalent to

$$\max_{\alpha \in \mathcal{H}(\delta, n)} \frac{\sqrt{h(\alpha, \phi) - \alpha^2}}{\alpha}, \quad (4.9)$$

where

$$\begin{aligned} h(\alpha, \phi) &:= \mathbb{E}_{\mathbb{P}_n} [(\phi^\top R)^2] + 2(\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] + \delta \|\phi\|_p^2 \\ &\quad + 2\sqrt{\delta \|\phi\|_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2} \sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R) \phi}. \end{aligned}$$

However,

$$\begin{aligned} &h(\alpha, \phi) - \alpha^2 \\ &= \mathbb{E}_{\mathbb{P}_n} [(\phi^\top R)^2] + 2(\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] - \alpha^2 + \delta \|\phi\|_p^2 \\ &\quad + 2\sqrt{\delta \|\phi\|_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2} \sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R) \phi} \\ &= \phi^\top \text{Var}_{\mathbb{P}_n}(R) \phi + \{\delta \|\phi\|_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2\} \\ &\quad + 2\sqrt{\delta \|\phi\|_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2} \sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R) \phi} \\ &= \left(\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R) \phi} + \sqrt{\delta \|\phi\|_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2} \right)^2. \end{aligned}$$

As a result, (4.9) is equivalent to (4.7). ■

With the help of Proposition 9, we are able to solve the inner maximization problem in (4.4). This is given in the following theorem:

Theorem 4 *The primal formulation given in (4.4) is equivalent to the following dual*

problem

$$\min_{\phi \in \mathcal{F}_\delta(n)} \frac{\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi} \sqrt{\phi^\top \mathbb{E}_{\mathbb{P}_n}[RR^\top]\phi - \delta \|\phi\|_p^2} + \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] \sqrt{\delta} \|\phi\|_p}{\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] \sqrt{\phi^\top \mathbb{E}_{\mathbb{P}_n}[RR^\top]\phi - \delta \|\phi\|_p^2} - \sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi} \sqrt{\delta} \|\phi\|_p} \quad (4.10)$$

in the sense that the two problems have the same optimal solutions and optimal value, where the feasible region $\mathcal{F}(n)$ is modified as

$$\mathcal{F}_\delta(n) = \{\phi : \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] > \sqrt{\delta} \|\phi\|_p, \phi^\top \mathbf{1} = 1\}. \quad (4.11)$$

Proof Due to the constraint $\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] > 0$, problem (4.4) is equivalent to the following problem:

$$\min_{\phi: \phi^\top \mathbf{1}=1, \phi^\top \mathbb{E}_{\mathbb{P}_n}[R]>0} \max_{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)} \left\{ \frac{\sqrt{\phi^\top \text{Var}_{\mathbb{P}}(R)\phi}}{\phi^\top \mathbb{E}_{\mathbb{P}}[R]} \right\}. \quad (4.12)$$

We can transform (4.12) into the following problem:

$$\min_{\substack{\phi: \phi^\top \mathbf{1}=1, \\ \phi^\top \mathbb{E}_{\mathbb{P}_n}[R]>0}} \max_{\substack{\mathbb{P}: \mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n, \alpha), \\ \alpha: \alpha > 0}} \frac{\sqrt{\text{Var}_{\mathbb{P}}(\phi^\top R)}}{\alpha}. \quad (4.13)$$

By proposition 9, the above formulation is equivalent to the following problem:

$$\min_{\phi} \max_{\alpha \in \mathcal{H}(\delta, n)} \frac{1}{\alpha} (\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi} + \sqrt{\delta \|\phi\|_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2}), \quad (4.14)$$

where $\mathcal{H}(\delta, n) := \{\alpha : \alpha > 0, \delta \|\phi\|_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2 \geq 0\}$.

Define

$$g(\alpha) = \frac{1}{\alpha} (\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi} + f(\alpha)),$$

where

$$f(\alpha) = \sqrt{\delta \|\phi\|_p^2 - (\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2}.$$

Solving (4.14) is equivalent to finding the optimal value for $g(\alpha)$. First, let us

consider the derivatives of $f(\alpha)$:

$$f'(\alpha) = -\frac{\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R]}{f(\alpha)}, \quad (4.15)$$

$$\begin{aligned} f''(\alpha) &= -\frac{1}{f(\alpha)} + \frac{\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R]}{f(\alpha)^2} f'(\alpha) \\ &= -\frac{1}{f(\alpha)} - \frac{(\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2}{f(\alpha)^3} < 0. \end{aligned} \quad (4.16)$$

With the help of (4.15) and (4.16), we are able to calculate $g'(\alpha)$:

$$\begin{aligned} g'(\alpha) &= -\frac{\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi}}{\alpha^2} - \frac{(\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])}{\alpha f(\alpha)} - \frac{f(\alpha)}{\alpha^2} \\ &= -\frac{1}{\alpha^2 f(\alpha)} [\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi} f(\alpha) + \alpha(\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R]) + f(\alpha)^2] \\ &= -\frac{1}{\alpha^2 f(\alpha)} [\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi} f(\alpha) + \phi^\top \mathbb{E}_{\mathbb{P}_n}[R](\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R]) + \delta \|\phi\|_p^2]. \end{aligned} \quad (4.17)$$

By (4.16) we know that $f(\alpha)$ is a concave function in α ; Hence, so is $\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi} f(\alpha) + \phi^\top \mathbb{E}_{\mathbb{P}_n}[R](\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])$. In problem (4.14), the range of α is positive and satisfies the following inequalities:

$$\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] - \sqrt{\delta} \|\phi\|_p \leq \alpha \leq \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] + \sqrt{\delta} \|\phi\|_p. \quad (4.18)$$

Define $\alpha_1 = \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] - \sqrt{\delta} \|\phi\|_p$, $\alpha_2 = \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] + \sqrt{\delta} \|\phi\|_p$. Then, we obtain:

$$g'(\alpha_1) = -\frac{\delta \|\phi\|_p^2 - \sqrt{\delta} \|\phi\|_p \phi^\top \mathbb{E}_{\mathbb{P}_n}[R]}{\alpha^2 f(\alpha)}, \quad g'(\alpha_2) = -\frac{\delta \|\phi\|_p^2 + \sqrt{\delta} \|\phi\|_p \phi^\top \mathbb{E}_{\mathbb{P}_n}[R]}{\alpha^2 f(\alpha)}.$$

Next, we discuss two cases according to the relationship between $\sqrt{\delta} \|\phi\|_p$ and $\phi^\top \mathbb{E}_{\mathbb{P}_n}[R]$.

Case 1: $\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] > \sqrt{\delta} \|\phi\|_p$. In this case, $0 < \alpha_1 \leq \alpha \leq \alpha_2$, $g'(\alpha_1) \geq 0$ and $g'(\alpha_2) \leq 0$. Therefore, the optimal α^* is in $[\alpha_1, \alpha_2]$. By the concavity of $g'(\alpha)$, there is a unique α^* such that $g'(\alpha^*) = 0$.

Case 2: $0 < \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] \leq \sqrt{\delta} \|\phi\|_p$. In this case, $\alpha_1 \leq 0 \leq \alpha \leq \alpha_2$ and the feasible

region of α is $0 \leq \alpha \leq \alpha_2$. Therefore, $g(\alpha^*) \geq g(0) = \infty$.

Only the first case will occur because the outer optimization problem in (4.12) is minimization over ϕ . Thus, the optimal solution ϕ will be such that $\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] > \sqrt{\delta} \|\phi\|_p$. Because $\sqrt{\delta} \|\phi\|_p > 0$, the feasible region for ϕ can be replaced by $\{\phi : \phi^\top \mathbf{1} = 1, \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] > \sqrt{\delta} \|\phi\|_p\}$.

Now, we need only to find the root α^* of the following equation:

$$\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi} f(\alpha) + \phi^\top \mathbb{E}_{\mathbb{P}_n}[R](\alpha - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R]) + \delta \|\phi\|_p^2 = 0. \quad (4.19)$$

First, let us simplify the notation. Denote $c := \sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi}$, $b := \phi^\top \mathbb{E}_{\mathbb{P}_n}[R]$, $d := \delta \|\phi\|_p^2$. Thus, equation (4.19) is:

$$cf(\alpha) + b(\alpha - b) + d = 0, \quad (4.20)$$

which is equivalent to:

$$cf(\alpha) = -b(\alpha - b) - d.$$

By taking the squares of the two sides, we obtain:

$$c^2 f^2(\alpha) = b^2(\alpha - b)^2 + d^2 + 2bd(\alpha - b).$$

By plugging in the definition of $f(\alpha)$, we get:

$$c^2(d - (\alpha - b)^2) = b^2(\alpha - b)^2 + d^2 + 2bd(\alpha - b),$$

which is equivalent to

$$(c^2 + b^2)(\alpha - b)^2 + 2bd(\alpha - b) + d^2 - c^2d = 0.$$

The left hand side of the above is a quadratic equation in α . The corresponding

discriminant Δ is:

$$\begin{aligned}\Delta &= 4b^2d^2 - 4(c^2 + b^2)(d^2 - c^2d) \\ &= 4(b^2c^2d + c^4d - c^2d^2) \\ &= 4c^2d(b^2 + c^2 - d) > 0,\end{aligned}\tag{4.21}$$

as $b^2 > d$ due to $\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] > \sqrt{\delta} \|\phi\|_p$. Then we obtain

$$\begin{aligned}\alpha^* - b &= \frac{-2bd \pm \sqrt{\Delta}}{2(c^2 + b^2)} \\ &= \frac{-bd \pm c\sqrt{d}\sqrt{b^2 + c^2 - d}}{(c^2 + b^2)}.\end{aligned}\tag{4.22}$$

We have two roots for the above quadratic equation, and we need to decide which is the correct α^* . Note that α^* satisfies

$$cf(\alpha^*) = -b(\alpha^* - b) - d,\tag{4.23}$$

and

$$\begin{aligned}-b(\alpha^* - b) - d &= \frac{b^2d \mp cb\sqrt{d}\sqrt{b^2 + c^2 - d} - dc^2 - db^2}{(c^2 + b^2)} \\ &= \frac{\mp cb\sqrt{d}\sqrt{b^2 + c^2 - d} - dc^2}{(c^2 + b^2)}.\end{aligned}\tag{4.24}$$

By (4.23), we know that $-b(\alpha - b) - d = cf(\alpha) > 0$. Also, all of c , b , and d are positive. This means that we should only choose the positive sign in (4.24). As a result, α^* should satisfy the following equation:

$$\alpha^* - b = \frac{-bd - c\sqrt{d}\sqrt{b^2 + c^2 - d}}{(c^2 + b^2)},\tag{4.25}$$

or

$$\alpha^* = \frac{b(b^2 + c^2 - d) - c\sqrt{d}\sqrt{b^2 + c^2 - d}}{(c^2 + b^2)}.\tag{4.26}$$

Then, we plug α^* into $g(\alpha)$ to obtain:

$$\begin{aligned}
g(\alpha^*) &= \frac{1}{\alpha^*} (\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi} + \sqrt{\delta \|\phi\|_p^2 - (\alpha^* - \phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2}) \\
&= \frac{1}{\alpha^*} (c + f(\alpha^*)) \\
&= \frac{1}{\alpha^*} (c - \frac{b(\alpha^* - b) + d}{c}) \\
&= \frac{1}{\alpha^*} (c - \frac{cd - b\sqrt{d}\sqrt{b^2 + c^2 - d}}{c^2 + b^2}) \\
&= \frac{1}{\alpha^*} \left(\frac{c(b^2 + c^2 - d) + b\sqrt{d}\sqrt{b^2 + c^2 - d}}{c^2 + b^2} \right).
\end{aligned} \tag{4.27}$$

By defining $e := \sqrt{b^2 + c^2 - d} = \sqrt{(\phi^\top \mathbb{E}_{\mathbb{P}_n}[R])^2 + \phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi - \delta \|\phi\|_p^2} = \sqrt{\phi^\top \mathbb{E}_{\mathbb{P}_n}[RR^\top]\phi - \delta \|\phi\|_p^2}$, we obtain $\alpha^* = \frac{be^2 - c\sqrt{de}}{c^2 + b^2}$ and

$$\begin{aligned}
g(\alpha^*) &= \frac{1}{\alpha^*} \left(\frac{c(b^2 + c^2 - d) + b\sqrt{d}\sqrt{b^2 + c^2 - d}}{c^2 + b^2} \right) \\
&= \frac{1}{\alpha^*} \left(\frac{ce^2 + b\sqrt{de}}{c^2 + b^2} \right) \\
&= \frac{ce^2 + b\sqrt{de}}{be^2 - c\sqrt{de}} \\
&= \frac{ce + b\sqrt{d}}{be - c\sqrt{d}}.
\end{aligned} \tag{4.28}$$

Thus, the original problem (4.14) becomes

$$\min_{\phi: \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] > \sqrt{\delta} \|\phi\|_p} \frac{\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi} \sqrt{\phi^\top \mathbb{E}_{\mathbb{P}_n}[RR^\top]\phi - \delta \|\phi\|_p^2} + \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] \sqrt{\delta} \|\phi\|_p}{\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] \sqrt{\phi^\top \mathbb{E}_{\mathbb{P}_n}[RR^\top]\phi - \delta \|\phi\|_p^2} - \sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi} \sqrt{\delta} \|\phi\|_p}.$$

The proof is complete. ■

Note that one constraint of the feasible region has been changed from $\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] > 0$ to $\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] > \sqrt{\delta} \|\phi\|_p$. Due to Proposition 5 in Chapter 3, when $\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)$, $\phi^\top \mathbb{E}_{\mathbb{P}}[R]$ can take any value in the interval $[\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] - \sqrt{\delta} \|\phi\|_p, \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] + \sqrt{\delta} \|\phi\|_p]$. If $\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] \leq \sqrt{\delta} \|\phi\|_p$, $\phi^\top \mathbb{E}_{\mathbb{P}}[R]$ can be 0, which means that the optimal value of the inner maximization problem is $+\infty$. Therefore, the outer minimization problem will

never consider such a ϕ . For this reason, we can modify the feasible region to (4.11).

The above theorem is not a trivial application of the results presented in Gao and Kleywegt (2016) or in Chapter 3. First of all, problem (4.4) is not linear in probability measure \mathbb{P} . Second, the DRSR model involves a ratio of terms that are functions of the probability measure \mathbb{P} . This makes the derivation of Theorem 4 much more difficult than that of DRMV. Also, the objective function of the dual form (4.10) is no longer a simple regularized term of the sample variance in DRMV. Although δ still plays an important role in the problem, its effect becomes much more complicated and can no longer be described simply as “regularization” or a “penalty”.

Theorem 4 transforms problem (4.4) into an equivalent finite dimensional minimization problem. However, there is still one unsolved problem: the objective function of (4.10) may not be convex. This means that the resulting minimization problem may still be tractable. However, we can prove that the square of the objective function is convex. To simplify the notation, define the objective function of (4.10) as $G(\phi)$, i.e.,:

$$G(\phi) = \frac{\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi} \sqrt{\phi^\top \mathbb{E}_{\mathbb{P}_n}[RR^\top]\phi - \delta \|\phi\|_p^2 + \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] \sqrt{\delta} \|\phi\|_p}{\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] \sqrt{\phi^\top \mathbb{E}_{\mathbb{P}_n}[RR^\top]\phi - \delta \|\phi\|_p^2} - \sqrt{\phi^\top \text{Var}_{\mathbb{P}_n}(R)\phi} \sqrt{\delta} \|\phi\|_p}. \quad (4.29)$$

Proposition 10 $G^2(\phi)$ is convex in $\phi \in \mathcal{F}_\delta(n)$.

Proof We first prove two lemmas.

Lemma 3 Consider the following two problems:

$$\max_{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)} \left\{ \frac{\sqrt{\phi^\top \text{Var}_{\mathbb{P}}(R)\phi}}{\phi^\top \mathbb{E}_{\mathbb{P}}[R]} \right\} \quad (4.30)$$

and

$$\max_{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)} \left\{ \frac{\phi^\top \text{Var}_{\mathbb{P}}(R)\phi}{(\phi^\top \mathbb{E}_{\mathbb{P}}[R])^2} \right\}. \quad (4.31)$$

When $\phi \in \mathcal{F}_\delta(n) = \{\phi : \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] > \sqrt{\delta} \|\phi\|_p\}$, the two problems have the same

optimal solution \mathbb{P}^* .

Proof Note that the objective function of (4.31) is the square of that of (4.30).

In order to prove that they have the same optimal solution, it suffices to show that

$$\forall \mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n), \frac{\sqrt{\phi^\top \text{Var}_{\mathbb{P}}(R)\phi}}{\phi^\top \mathbb{E}_{\mathbb{P}}[R]} > 0.$$

By Proposition 5 in Chapter 3, we know:

$$\min_{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)} \mathbb{E}_{\mathbb{P}}[\phi^\top R] = \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] - \sqrt{\delta} \|\phi\|_p.$$

Therefore, when $\phi \in \mathcal{F}_\delta(n)$, $\min_{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)} \mathbb{E}_{\mathbb{P}}[\phi^\top R] > 0$. This means that $\forall \mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)$,

$$\frac{\sqrt{\phi^\top \text{Var}_{\mathbb{P}}(R)\phi}}{\phi^\top \mathbb{E}_{\mathbb{P}}[R]} > 0. \quad \blacksquare$$

Lemma 4 $\forall \mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)$, $W(\phi) := \frac{\phi^\top \text{Var}_{\mathbb{P}}(R)\phi}{(\phi^\top \mathbb{E}_{\mathbb{P}}[R])^2}$ is a convex function of ϕ .

Proof For notational simplicity, we define $\Sigma := \text{Var}_{\mathbb{P}}(R)$ and $\mu := \mathbb{E}_{\mathbb{P}_n}(R)$. The Hessian matrix of $W(\phi)$ is

$$\frac{\partial^2 W(\phi)}{\partial \phi^2} = \frac{2}{(\phi^\top \mu)^4} [(\phi^\top \mu)^2 \Sigma - 2(\phi^\top \mu)(\Sigma \phi)\mu^\top + 3(\phi^\top \Sigma \phi)(\mu \mu^\top)]. \quad (4.32)$$

Next, $\forall x \in \mathbb{R}^d$,

$$\begin{aligned} x^\top \frac{\partial^2 W(\phi)}{\partial \phi^2} x &= \frac{2}{(\phi^\top \mu)^4} [(\phi^\top \mu)^2 (x^\top \Sigma x) - 2(\phi^\top \mu)(x^\top \Sigma \phi)(\mu^\top x) + 3(\phi^\top \Sigma \phi)(\mu^\top x)^2] \\ &= \frac{2}{(\phi^\top \mu)^4} [(\phi^\top \mu)^2 (x^\top \Sigma x) - 2(\phi^\top \mu)(x^\top \Sigma \phi)(\mu^\top x) + (\phi^\top \Sigma \phi)(\mu^\top x)^2 + 2(\phi^\top \Sigma \phi)(\mu^\top x)^2]. \end{aligned} \quad (4.33)$$

Consider

$$\begin{aligned} &(\phi^\top \mu)^2 (x^\top \Sigma x) - 2(\phi^\top \mu)(x^\top \Sigma \phi)(\mu^\top x) + (\phi^\top \Sigma \phi)(\mu^\top x)^2 \\ &\geq 2|(\phi^\top \mu)(\sqrt{\phi^\top \Sigma \phi} \sqrt{x^\top \Sigma x})(\mu^\top x)| - 2(\phi^\top \mu)(x^\top \Sigma \phi)(\mu^\top x) \\ &\geq 2|(\phi^\top \mu)(\sqrt{\phi^\top \Sigma \phi} \sqrt{x^\top \Sigma x})(\mu^\top x)| - 2|(\phi^\top \mu)(x^\top \Sigma \phi)(\mu^\top x)| \\ &= 2|(\phi^\top \mu)(\mu^\top x)| (|\sqrt{\phi^\top \Sigma \phi} \sqrt{x^\top \Sigma x}| - |x^\top \Sigma \phi|) \geq 0, \end{aligned} \quad (4.34)$$

where the last inequality holds because of Cauchy-Schwarz's inequality. Therefore, (4.33) becomes

$$\begin{aligned} x^\top \frac{\partial^2 W(\phi)}{\partial \phi^2} x &= \frac{2}{(\phi^\top \mu)^4} [(\phi^\top \mu)^2 (x^\top \Sigma x) - 2(\phi^\top \mu)(x^\top \Sigma \phi)(\mu^\top x) \\ &\quad + (\phi^\top \Sigma \phi)(\mu^\top x)^2 + 2(\phi^\top \Sigma \phi)(\mu^\top x)^2] \\ &\geq 2(\phi^\top \Sigma \phi)(\mu^\top x)^2 \geq 0. \end{aligned} \quad (4.35)$$

Thus, $W(\phi)$ is convex in ϕ . ■

From the proof of Theorem 4, we know that when $\phi \in \mathcal{F}_\delta(n) = \{\phi : \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] > \sqrt{\delta} \|\phi\|_p\}$, the following equation holds:

$$G(\phi) = \max_{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)} \left\{ \frac{\sqrt{\phi^\top \text{Var}_{\mathbb{P}}(R)\phi}}{\phi^\top \mathbb{E}_{\mathbb{P}}[R]} \right\}. \quad (4.36)$$

By Lemma 3, when $\phi \in \mathcal{F}_\delta(n)$,

$$G^2(\phi) = \max_{\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)} \left\{ \frac{\phi^\top \text{Var}_{\mathbb{P}}(R)\phi}{(\phi^\top \mathbb{E}_{\mathbb{P}}[R])^2} \right\}. \quad (4.37)$$

It follows from Lemma 4 that $\frac{\phi^\top \text{Var}_{\mathbb{P}}(R)\phi}{(\phi^\top \mathbb{E}_{\mathbb{P}}[R])^2}$ is convex. Then, $G^2(\cdot)$ is the maximum of a set of convex functions, which is also convex. ■

With Theorem 4 and Proposition 10, we are able to transform problem (4.4) into an equivalent tractable problem.

Corollary 11 *The primal formulation given in (4.4) is equivalent to the following problem*

$$\min_{\phi \in \mathcal{F}_\delta(n)} G^2(\phi), \quad (4.38)$$

where

$$G(\phi) = \frac{\sqrt{\phi^\top \text{Var}_{\mathbb{P}_n} \phi} \sqrt{\phi^\top \mathbb{E}_{\mathbb{P}_n}[RR^\top] \phi - \delta \|\phi\|_p^2 + \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] \sqrt{\delta} \|\phi\|_p}}{\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] \sqrt{\phi^\top \mathbb{E}_{\mathbb{P}_n}[RR^\top] \phi - \delta \|\phi\|_p^2 - \sqrt{\phi^\top \text{Var}_{\mathbb{P}_n} \phi} \sqrt{\delta} \|\phi\|_p}} \quad (4.39)$$

and

$$\mathcal{F}_\delta(n) = \{\phi : \phi^\top \mathbb{E}_{\mathbb{P}_n}[R] \geq \sqrt{\delta} \|\phi\|_p, \phi^\top \mathbf{1} = 1\}, \quad (4.40)$$

in the sense that the two problems have the same optimal solution and optimal value.

It is clear that the feasible region $\mathcal{F}_\delta(n)$ is convex. So (4.38) is a convex optimization problem, and thus it is tractable.

4.3 Choice of Model Parameter δ

There is only one undefined parameter δ in DRSR, which is different from DRMV in Chapter 3 in that there is no equality constraint for expected portfolio return in the Sharpe ratio optimization problem (4.3). The requirement for expected return has been incorporated into the optimization of the Sharpe ratio. Thanks to this formulation, we do not need to specify the target return, which makes the model more data driven.

The choice of δ is not only curious in theory, but also crucial in practical implementation and for the success of our algorithm. Similar to Chapter 3, the idea is that the choice of this parameter should be informed by the data (i.e., in a data-driven way) based on some statistical principles, rather than being arbitrarily exogenous. Specifically, we define the distributional uncertainty region just large enough that the correct optimal portfolio (the one that we would apply if the underlying distribution were known) becomes a plausible choice with a sufficiently high confidence level. We use the same RWPI approach as in Chapter 3 to find the value of δ , however, in this case the RWP function and its asymptotic distribution become more elegant, which is further elaborated in remainder of this section.

Before deciding how to choose δ , we need to impose several statistical assumptions.

A1) The underlying return time series $(R_k : k \geq 0)$ is a stationary, ergodic process satisfying $\mathbb{E}_{\mathbb{P}_*}(\|R_k\|_2^4) < \infty$ for each $k \geq 0$. Moreover, for each measurable $g(\cdot)$ such

that $|g(x)| \leq c(1 + \|x\|_2^2)$ for some $c > 0$, the limit

$$\Upsilon_g := \lim_{n \rightarrow \infty} \text{Var}_{\mathbb{P}_*} \left(n^{-1/2} \sum_{k=1}^n g(R_k) \right)$$

exists and the central limit theorem holds:

$$n^{1/2} [\mathbb{E}_{\mathbb{P}_n}(g(R)) - \mathbb{E}_{\mathbb{P}_*}(g(R))] \Rightarrow N(0, \Upsilon_g),$$

where (and henceforth) “ \Rightarrow ” denotes weak convergence.

A2) For any matrix $\Lambda \in \mathbb{R}^{d \times d}$ and any vector $\zeta \in \mathbb{R}^d$ such that either $\Lambda \neq 0$ or $\zeta \neq 0$,

$$\mathbb{P}_*(\|\Lambda R + \zeta\|_2 > 0) > 0.$$

A3) For the empirical probability measure \mathbb{P}_n and the true probability measure \mathbb{P}_* , we have $1^\top \mathbb{E}_{\mathbb{P}_n}[RR^\top]^{-1} \mathbb{E}_{\mathbb{P}_n}[R] > 0$, $1^\top \mathbb{E}_{\mathbb{P}_*}[RR^\top]^{-1} \mathbb{E}_{\mathbb{P}_*}[R] > 0$.

The first two assumptions are the same as those with respect to DRMV in Chapter 3. Assumption **A3)**, however, is different. If Assumption **A3)** does not hold, the classical and empirical Sharpe Ratio portfolio selection problem will not have any finite optimal solution. To be more precise, if $1^\top \mathbb{E}_{\mathbb{P}_*}[RR^\top]^{-1} \mathbb{E}_{\mathbb{P}_*}[R] \leq 0$, the optimal portfolio ϕ_* will satisfy $\mathbb{E}_{\mathbb{P}_*}[\phi_*^\top R] = \infty$, which means that ϕ_* is not a finite solution. In practice, it is impossible to implement such a portfolio. Further details are provided in the proof of Proposition 12.

As stated in Chapter 3, the “optimal” order of δ_n should be $O(n^{-1})$. In order to choose an appropriate δ_n , we follow here the idea behind the RWPI approach presented in Chapter 3. Intuitively, δ should be chosen such that the set $\mathcal{U}_\delta(\mathbb{P}_n) = \{\mathbb{P} : D_c(\mathbb{P}, \mathbb{P}_n) \leq \delta\}$ contains all of the probability measures that are plausible variations

of the data represented by \mathbb{P}_n . Denote by $\mathcal{Q}(\mathbb{P})$ the Sharpe ratio portfolio selection problem assuming that \mathbb{P} is the underlying true probability measure:

$$\min_{\phi: 1^\top \phi = 1, \mathbb{E}_{\mathbb{P}}[\phi^\top R] > 0} \frac{\sqrt{\text{Var}_{\mathbb{P}}(\phi^\top R)}}{\mathbb{E}_{\mathbb{P}}[\phi^\top R]}, \quad (4.41)$$

by $\phi_{\mathbb{P}}$ a solution to $\mathcal{Q}(\mathbb{P})$ and by $\Phi_{\mathbb{P}}$ the set of all such solutions. Note that above problem will always be feasible.

Now, when δ is suitably chosen so that $\mathcal{U}_{\delta}(\mathbb{P}_n)$ constitutes the models that are plausible variations of \mathbb{P}_n , any $\phi_{\mathbb{P}}$ with $\mathbb{P} \in \mathcal{U}_{\delta}(\mathbb{P}_n)$ is a plausible estimate of ϕ_* . This intuition motivates the definition of the following set:

$$\Lambda_{\delta}(\mathbb{P}_n) = \cup_{\mathbb{P} \in \mathcal{U}_{\delta}(\mathbb{P}_n)} \Phi_{\mathbb{P}},$$

which corresponds to all of the plausible estimates of ϕ_* . As a result, $\Lambda_{\delta}(\mathbb{P}_n)$ is a natural confidence region for ϕ_* and, therefore, δ should be chosen as the smallest number δ_n^* such that ϕ_* belongs to this region with a given confidence level. Namely,

$$\delta_n^* = \min\{\delta > 0 : \mathbb{P}^*(\phi_* \in \Lambda_{\delta}(\mathbb{P}_n)) \geq 1 - \delta_0\},$$

where $1 - \delta_0$ is a user-defined confidence level (typically 95%).

However, by mere definition, it is difficult to compute δ_n^* . We now provide a simpler representation for δ_n^* via the robust Wasserstein profile (RWP) function.

Before introducing the formal definition of this function, we characterize the condition that $\phi_* \in \Lambda_{\delta}(\mathbb{P}_n)$.

Proposition 12 *If Assumption **A3**) is satisfied, then $\phi_* \in \Lambda_{\delta}(\mathbb{P}_n)$ if and only if there exist $\mathbb{P} \in \mathcal{U}_{\delta}(\mathbb{P}_n)$ and a positive constant c such that*

$$\mathbb{E}_{\mathbb{P}}[RR^\top \phi_* - cR] = 0.$$

Proof We first study the conditions for any given ϕ to be a solution to problem $\mathcal{Q}(\mathbb{P})$.

We transform (4.41) into

$$\min_{\phi, \alpha: \phi^\top \mathbf{1} = 1, \mathbb{E}_{\mathbb{P}}[\phi^\top R] = \alpha, \alpha > 0} \frac{\sqrt{\text{Var}_{\mathbb{P}}(\phi^\top R)}}{\alpha}. \quad (4.42)$$

Due to Lemma 3, the solution of (4.42) is the same as the following problem:

$$\min_{\phi, \alpha: \phi^\top \mathbf{1} = 1, \mathbb{E}_{\mathbb{P}}[\phi^\top R] = \alpha, \alpha > 0} \frac{\phi^\top \mathbb{E}_{\mathbb{P}}(RR^\top)\phi}{\alpha^2}. \quad (4.43)$$

The Lagrangian is

$$J(\phi, \lambda_1, \lambda_2) = \frac{\phi^\top \mathbb{E}_{\mathbb{P}}(RR^\top)\phi}{\alpha^2} - \lambda_1(\mathbb{E}_{\mathbb{P}}[R]^\top \phi - \alpha) - \lambda_2(\phi^\top \mathbf{1} - 1). \quad (4.44)$$

It follows from $\frac{\partial J}{\partial \phi} = 0$ that

$$\frac{2\mathbb{E}_{\mathbb{P}}(RR^\top)\phi}{\alpha^2} - \lambda_1 \mathbb{E}_{\mathbb{P}}[R] - \lambda_2 \mathbf{1} = 0. \quad (4.45)$$

Solving for ϕ , we obtain

$$\phi = \frac{\alpha^2}{2} \mathbb{E}_{\mathbb{P}}(RR^\top)^{-1} [\lambda_1 \mathbb{E}_{\mathbb{P}}[R] + \lambda_2 \mathbf{1}]. \quad (4.46)$$

For notational simplicity, denote $a := \mathbb{E}_{\mathbb{P}}[R]^\top \mathbb{E}_{\mathbb{P}}(RR^\top)^{-1} \mathbb{E}_{\mathbb{P}}[R]$, $b := \mathbb{E}_{\mathbb{P}}[R]^\top \mathbb{E}_{\mathbb{P}}(RR^\top)^{-1} \mathbf{1}$, $c := \mathbf{1}^\top \mathbb{E}_{\mathbb{P}}(RR^\top)^{-1} \mathbf{1}$. Then, λ_1 and λ_2 can be solved as:

$$\lambda_1 = \left[\frac{c - \frac{1}{\alpha} b}{ac - b^2} \right] \frac{2}{\alpha}, \quad \lambda_2 = \left[\frac{b - \frac{1}{\alpha} a}{b^2 - ac} \right] \frac{2}{\alpha}.$$

Plugging in the original problem, we get:

$$\begin{aligned}
\frac{\phi^\top \mathbb{E}_{\mathbb{P}}[RR^\top] \phi}{\alpha^2} &= \frac{\alpha^2}{4} [\lambda_1^2 a + 2\lambda_1 \lambda_2 b + \lambda_2^2 c] \\
&= \frac{a(c - \frac{1}{\alpha}b)^2}{(ac - b^2)^2} - \frac{2(c - \frac{1}{\alpha}b)(b - \frac{1}{\alpha}a)b}{(ac - b^2)^2} + \frac{(b - \frac{1}{\alpha}a)^2 c}{(ac - b^2)^2} \\
&= \frac{1}{(ac - b^2)^2} [(c - \frac{1}{\alpha}b)(ac - b^2) + (b - \frac{1}{\alpha}a)(\frac{b^2 - ac}{\alpha})] \\
&= \frac{1}{(ac - b^2)} [(c - \frac{1}{\alpha}b) - (\frac{b}{\alpha} - \frac{1}{\alpha^2}a)] \\
&= \frac{1}{(ac - b^2)} [c + \frac{a}{\alpha^2} - \frac{2b}{\alpha}].
\end{aligned} \tag{4.47}$$

If the following constraint is satisfied:

$$b = \mathbb{E}_{\mathbb{P}}[R]^\top \mathbb{E}_{\mathbb{P}}(RR^\top)^{-1} \mathbf{1} > 0, \tag{4.48}$$

then the optimal $\alpha^* = \frac{a}{b}$ and the optimal value of (4.47) is:

$$\begin{aligned}
\frac{\phi^\top \mathbb{E}_{\mathbb{P}}[RR^\top] \phi}{\alpha^2} &= \frac{1}{(ac - b^2)} [c + \frac{a}{\alpha^2} - \frac{2b}{\alpha}] \\
&= \frac{1}{(ac - b^2)} [c + \frac{b^2}{a} - \frac{2b^2}{a}] \\
&= \frac{1}{(ac - b^2)} [c - \frac{b^2}{a}] \\
&= \frac{1}{a}.
\end{aligned} \tag{4.49}$$

Thus, in order for ϕ to be the optimal solution of $Q(\mathbb{P})$, \mathbb{P} should satisfy (4.48) and (4.49). Then, we obtain

$$\frac{\mathbb{E}_{\mathbb{P}}^2[\phi^\top R]}{\phi^\top \mathbb{E}_{\mathbb{P}}(RR^\top) \phi} = a = \mathbb{E}_{\mathbb{P}}[R]^\top \mathbb{E}_{\mathbb{P}}(RR^\top)^{-1} \mathbb{E}_{\mathbb{P}}[R]. \tag{4.50}$$

Therefore,

$$(\phi^\top \mathbb{E}_{\mathbb{P}}[R])^2 = (\mathbb{E}_{\mathbb{P}}[R]^\top \mathbb{E}_{\mathbb{P}}(RR^\top)^{-1} \mathbb{E}_{\mathbb{P}}[R]) (\phi^\top \mathbb{E}_{\mathbb{P}}(RR^\top) \phi). \tag{4.51}$$

By letting $\tilde{\phi} := \mathbb{E}_{\mathbb{P}}(RR^{\top})^{\frac{1}{2}}\phi$, $\tilde{\mu} := \mathbb{E}_{\mathbb{P}}(RR^{\top})^{-\frac{1}{2}}\mathbb{E}_{\mathbb{P}}[R]$, we obtain:

$$(\tilde{\phi}^{\top}\tilde{\mu})^2 = \|\tilde{\mu}\|_2^2\|\tilde{\phi}\|_2^2. \quad (4.52)$$

By Cauchy-Schwarz's inequality, for any constant c ,

$$\tilde{\phi} = c\tilde{\mu},$$

i.e.,

$$\mathbb{E}_{\mathbb{P}}(RR^{\top})^{\frac{1}{2}}\phi = c\mathbb{E}_{\mathbb{P}}(RR^{\top})^{-\frac{1}{2}}\mathbb{E}_{\mathbb{P}}[R].$$

Hence,

$$\mathbb{E}_{\mathbb{P}}(RR^{\top})\phi = c\mathbb{E}_{\mathbb{P}}[R]. \quad (4.53)$$

From the above discussion, if a probability measure \mathbb{P} satisfies $\mathbb{E}_{\mathbb{P}}[R]^{\top}\mathbb{E}_{\mathbb{P}}(RR^{\top})^{-1}\mathbf{1} > 0$, then the solutions $\Phi_{\mathbb{P}}$ for $\mathcal{Q}(\mathbb{P})$ must be finite. Therefore, by Assumption **A3**, ϕ_* is finite.

Now, if $b = \mathbb{E}_{\mathbb{P}}[R]^{\top}\mathbb{E}_{\mathbb{P}}(RR^{\top})^{-1}\mathbf{1} \leq 0$, then it follows from (4.48) that the optimal $\alpha^* = \infty$, which is impossible for any fixed finite ϕ . Thus, ϕ_* will never be in the set $\Phi_{\mathbb{P}}$. So this case can be ignored in our discussion.

In conclusion, in order to make ϕ_* be one of the optimal solutions for $\mathcal{Q}(\mathbb{P})$, the probability measure \mathbb{P} has to satisfy:

$$\begin{cases} \mathbb{E}_{\mathbb{P}}[R]^{\top}\mathbb{E}_{\mathbb{P}}(RR^{\top})^{-1}\mathbf{1} > 0 \\ \exists c \in \mathbb{R}, \mathbb{E}_{\mathbb{P}}(RR^{\top})\phi_* = c\mathbb{E}_{\mathbb{P}}[R]. \end{cases} \quad (4.54)$$

A simple transformation on (4.53) yields

$$\mathbf{1} = \mathbf{1}^{\top}\phi_* = c(\mathbf{1}^{\top}\mathbb{E}_{\mathbb{P}}[RR^{\top}]^{-1}\mathbb{E}_{\mathbb{P}}[R]) = cb. \quad (4.55)$$

Therefore, $b = \frac{1}{c}$ and $b > 0$ is equivalent to $c > 0$. Then, conditions (4.54) can be

simplified as

$$\exists c \in \mathbb{R}^+, \mathbb{E}_{\mathbb{P}}(RR^\top)\phi_* = c\mathbb{E}_{\mathbb{P}}[R]. \quad (4.56)$$

Thus, $\phi \in \Lambda_\delta(\mathbb{P}_n)$ if and only if there exist $\mathbb{P} \in \mathcal{U}_\delta(\mathbb{P}_n)$ and a constant $c > 0$ such that

$$\mathbb{E}_{\mathbb{P}}[RR^\top\phi_* - cR] = 0.$$

■

Proposition 12 gives us an equivalent condition to present $\phi_* \in \Lambda_\delta(\mathbb{P}_n)$. Then, we can define the RWP function as follows:

$$\bar{\mathcal{R}}_n(\phi, c) := \inf\{D_c(\mathbb{P}, \mathbb{P}_n) : \mathbb{E}_{\mathbb{P}}[RR^\top\phi - cR] = 0\}. \quad (4.57)$$

Note that, in the definition of $\bar{\mathcal{R}}_n(\phi, c)$, when \mathbb{P} is fixed, c is uniquely decided. Specifically, $c = c(\mathbb{P}) = \frac{1}{1^\top \mathbb{E}_{\mathbb{P}}[RR^\top]^{-1} \mathbb{E}_{\mathbb{P}}[R]}$. This is due to the fact that all of the feasible solution ϕ in problem (4.41) have to satisfy the budget constraint: $\phi^\top 1 = 1$. However, we choose not to define a RWP function as $\tilde{\mathcal{R}}_n(\phi) = \bar{\mathcal{R}}_n(\phi, c(\mathbb{P}))$. This is because if we plug $c = c(\mathbb{P}) = \frac{1}{1^\top \mathbb{E}_{\mathbb{P}}[RR^\top]^{-1} \mathbb{E}_{\mathbb{P}}[R]}$ into equation (4.57), the equation will be no longer be linear in \mathbb{P} . It would then be very difficult to find an explicit expression for the RWP function in this case. In the remaining part of this section, we further explain how this RWP function helps us to find an upper bound for δ_n^* .

We first introduce some notations. Define $\Sigma_n := \mathbb{E}_{\mathbb{P}_n}[RR^\top]$, $\mu_n := \mathbb{E}_{\mathbb{P}_n}[R]$, $c_n := \frac{1}{1^\top \Sigma_n^{-1} \mu_n}$ and $\phi_n := c_n \Sigma_n^{-1} \mu_n$. Parallely, we define $\Sigma_* := \mathbb{E}_{\mathbb{P}_*}[RR^\top]$, $\mu_* := \mathbb{E}_{\mathbb{P}_*}[R]$, $c_* := \frac{1}{1^\top \Sigma_*^{-1} \mu_*}$ and $\phi_* := c_* \Sigma_*^{-1} \mu_*$.

It is not difficult to verify that ϕ_n and ϕ_* are the optimal solution of problems $\mathcal{Q}(\mathbb{P}_n)$ and $\mathcal{Q}(\mathbb{P}_*)$, respectively. By definition (4.57) and Proposition 12, we conclude that

$$\bar{\mathcal{R}}_n(\phi_*, c_*) \leq \delta \implies \phi_* \in \Lambda_\delta(\mathbb{P}_n). \quad (4.58)$$

Note that, by Assumption **A3**), both c_n and c_* are positive. Thus, Proposition 12 can be applied to (4.58). Just as in Chapter 3, we would like to know what the asymptotic distribution of $n\bar{\mathcal{R}}_n(\phi_*, c_*)$ will be.

Theorem 5 *Assume that **A1**) and **A2**) hold and define $g(x) = xx^\top \phi_* - c_*x$. Then,*

$$n\bar{\mathcal{R}}_n(\phi^*, c^*) \Rightarrow L_0 := \sup_{\bar{\lambda} \in \mathbb{R}^r} \left\{ \bar{\lambda}^\top Z - \frac{1}{4} \mathbb{E}_{\mathbb{P}_*} \left\| (R^\top \phi_* - c_*) \bar{\lambda} + (\bar{\lambda}^\top R) \phi_* \right\|_p^2 \right\},$$

where $Z \sim N(0, \Upsilon_g)$. Moreover, if $p = 2$, then

$$L_0 = Z^\top A^{-1} Z,$$

where $A := \mathbb{E}_{\mathbb{P}_*} \left\{ [(R^\top \phi_* - c_*)I + \phi_* R^\top]^\top [(R^\top \phi_* - c_*)I + \phi_* R^\top] \right\}$.

Proof By proposition 1 of Blanchet et al. (2016), $\bar{\mathcal{R}}_n(\phi, c)$ can be expressed as

$$\bar{\mathcal{R}}_n(\phi, c) = \sup_{\lambda \in \mathbb{R}^r} \left\{ -\mathbb{E}_{\mathbb{P}_n} \left[\sup_{u \in \mathbb{R}^m} \left\{ \lambda^\top (uu^\top \phi - cu) - \|u - R_i\|_q^2 \right\} \right] \right\}. \quad (4.59)$$

Applying similar techniques to those used in Chapter 3, we obtain

$$\begin{aligned} \bar{\mathcal{R}}_n(\phi, c) &= \sup_{\lambda \in \mathbb{R}^r} \left\{ -\mathbb{E}_{\mathbb{P}_n} \left[\sup_{u \in \mathbb{R}^m} \left\{ \lambda^\top (uu^\top \phi - cu) - \|u - R\|_q^2 \right\} \right] \right\} \\ &= \sup_{\lambda \in \mathbb{R}^r} \left\{ -\mathbb{E}_{\mathbb{P}_n} \left[\sup_{\Delta \in \mathbb{R}^m} \left\{ \lambda^\top ((R + \Delta)(R + \Delta)^\top \phi - c(R + \Delta)) - \|\Delta\|_q^2 \right\} \right] \right\} \\ &= \sup_{\lambda \in \mathbb{R}^r} \left\{ -\mathbb{E}_{\mathbb{P}_n} \left[\lambda^\top (RR^\top \phi - cR) \right] - \right. \\ &\quad \left. \mathbb{E}_{\mathbb{P}_n} \left[\sup_{\Delta \in \mathbb{R}^m} \left\{ (R^\top \phi \lambda^\top + \lambda^\top R \phi^\top - c\lambda^\top) \Delta - \Delta^\top \lambda \phi^\top \Delta - \|\Delta\|_q^2 \right\} \right] \right\}. \end{aligned} \quad (4.60)$$

Introduce the scaling $\Delta = \bar{\Delta}/n^{1/2}$ and $\bar{\lambda} = \lambda n^{1/2}$. Then, we obtain

$$\begin{aligned} n\bar{\mathcal{R}}_n(\phi, c) &= \sup_{\bar{\lambda} \in \mathbb{R}^r} \left\{ -n^{1/2} \mathbb{E}_{\mathbb{P}_n} \left[\bar{\lambda}^\top (RR^\top \phi - cR) \right] - \right. \\ &\quad \left. \mathbb{E}_{\mathbb{P}_n} \left[\sup_{\bar{\Delta} \in \mathbb{R}^m} \left\{ (R_i^\top \phi \bar{\lambda}^\top + \bar{\lambda}^\top R_i \phi^\top - c\bar{\lambda}^\top) \bar{\Delta} - \bar{\Delta}^\top \bar{\lambda} \phi^\top \bar{\Delta} / n^{1/2} - \|\bar{\Delta}\|_q^2 \right\} \right] \right\}. \end{aligned}$$

In the proof of Proposition 3 in Blanchet et al. (2016), under Assumption **A2**), a technique is introduced to show that $\bar{\Delta}$ and $\bar{\lambda}$ can be restricted to compact sets with high probability, and therefore the term $\bar{\Delta}^\top \bar{\lambda} \phi^\top \bar{\Delta} / n^{1/2}$ is asymptotically negligible for each fixed ϕ . Then, we obtain:

$$\begin{aligned}
& \sup_{\bar{\Delta} \in \mathbb{R}^m} \{(R^\top \phi \bar{\lambda}^\top + \bar{\lambda}^\top R \phi^\top - c \bar{\lambda}^\top) \bar{\Delta} - \|\bar{\Delta}\|_q^2\} \\
&= \sup_{\bar{\Delta} \in \mathbb{R}^m} \{\|R^\top \phi \bar{\lambda} + \bar{\lambda}^\top R \phi - c \bar{\lambda}\|_p \|\bar{\Delta}\|_q - \|\bar{\Delta}\|_q^2\} \\
&= \frac{1}{4} \| (R^\top \phi) \bar{\lambda} + (\bar{\lambda}^\top R) \phi - c \bar{\lambda} \|_p^2 \\
&= \frac{1}{4} \| (R^\top \phi - c) \bar{\lambda} + (\bar{\lambda}^\top R) \phi \|_p^2.
\end{aligned}$$

Therefore, if

$$-n^{1/2} \mathbb{E}_{\mathbb{P}_n} [(RR^\top \phi - cR)] \Rightarrow -Z$$

for some Z (to be characterized momentarily), then we conclude

$$n\bar{\mathcal{R}}_n(\phi, c) \Rightarrow L_0 = \sup_{\bar{\lambda} \in \mathbb{R}^r} \{ \bar{\lambda}^\top Z - \frac{1}{4} \mathbb{E}_{\mathbb{P}_*} \| (R^\top \phi - c) \bar{\lambda} + (\bar{\lambda}^\top R) \phi \|_p^2 \}. \quad (4.61)$$

If $p = 2$, then we have

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_*} \| (R^\top \phi - c) \bar{\lambda} + (\bar{\lambda}^\top R) \phi \|_2^2 &= \mathbb{E}_{\mathbb{P}_*} \| [(R^\top \phi - c)I + \phi R^\top] \bar{\lambda} \|_2^2 \\
&= \bar{\lambda}^\top A \bar{\lambda},
\end{aligned} \quad (4.62)$$

where

$$A := \mathbb{E}_{\mathbb{P}_*} \{ [(R^\top \phi - c)I + \phi R^\top]^\top [(R^\top \phi - c)I + \phi R^\top] \}. \quad (4.63)$$

Therefore, it is easy to find that

$$L_0 = Z^\top A^{-1} Z. \quad (4.64)$$

The next problem is to find the distribution of Z . To this end,

$$\begin{aligned}
\sqrt{n}\mathbb{E}_{\mathbb{P}_n}[RR^\top\phi^* - c^*R] &= \sqrt{n}\mathbb{E}_{\mathbb{P}_n}[RR^\top]\phi^* - \sqrt{nc^*}\mathbb{E}_{\mathbb{P}_n}[R] \\
&= \sqrt{n}[\mathbb{E}_{\mathbb{P}_n}[RR^\top] - \mathbb{E}_{P^*}[RR^\top]]\phi^* - \sqrt{nc^*}[\mathbb{E}_{\mathbb{P}_n}[R] - \mathbb{E}_{P^*}[R]] \\
&\Rightarrow Z := X - Y,
\end{aligned}
\tag{4.65}$$

where X is the asymptotic distribution of $\sqrt{n}[\mathbb{E}_{\mathbb{P}_n}[RR^\top] - \mathbb{E}_{P^*}[RR^\top]]\phi^*$ and Y is the asymptotic distribution of $\sqrt{nc^*}[\mathbb{E}_{\mathbb{P}_n}[R] - \mathbb{E}_{P^*}[R]]$. By Assumption **A1**), we know that both X and Y are multivariate normal distribution and Z is also normally distributed with mean 0. Thus, the distribution of $Z^\top A^{-1}Z$ equals a linear combination of independent chi-square distributions whose coefficients are the eigenvalue of matrix $(\Sigma_Z^{1/2})^\top A^{-1}\Sigma_Z^{1/2}$, where Σ_Z is the covariance matrix of Z . ■

Note that L_0 has an explicit expression when $p = 2$. When $p \neq 2$, using the inequalities that $\|x\|_p^2 \geq \|x\|_2^2$ if $p < 2$ and $d^{(\frac{1}{2}-\frac{1}{p})}\|x\|_p^2 \geq \|x\|_2^2$ if $p > 2$, we can find a stochastic upper bound of L_0 that can be explicitly expressed. In this case we can obtain $\bar{\delta}_n^*$ in exactly the same way: namely, first compute the $1 - \delta_0$ quantile of L_0 and then let $\bar{\delta}_n^*$ be such a quantile multiplied by $1/n$. The distribution of L_0 can be calibrated using a natural plug-in estimator, leading to an asymptotically equivalent estimator of $\bar{\delta}_n^*$. The validity of this type of (plug-in) approach is explained in the following section in a slightly different setting, but the principle applies directly to the setting of L_0 . This approach is tested empirically in Section 4.4.

We now present a simple “menu” for estimating δ .

1. Collect return data $\{R_i\}_{i=1}^n$.
2. Use the sample mean $\mu_n = \mathbb{E}_{\mathbb{P}_n}(R)$ and the sample second-moment matrix $\Sigma_n = \mathbb{E}_{\mathbb{P}_n}(RR^\top)$ to approximate μ_* and Σ_* , respectively, appearing in Theorem 5.
3. Use the solution ϕ_n , which is the solution to problem $\mathcal{Q}(\mathbb{P}_n)$ (see (4.41)), to approximate ϕ^* in Theorem 3.

4. Apply Theorem 5 to determine $\delta = \bar{\delta}^*$ with the 95% confidence level.

4.4 Empirical Performance and Comparisons

In this section we report the results of our backtesting experiments on the S&P 500 constituents that compare the performance of our DRSR portfolios with those of the portfolios based on the following models: classical (non-robust) Sharpe ratio optimization, Fama-French, Black-Litterman, classical single-period Markowitz, DRMV and an equally weighted portfolio. The first four models are well-known and have been widely used in practice and the fifth one is the distributional robust model based on a Markowitz model, as is developed in Chapter 3 of the present thesis. The equally weighted strategy is actually an extreme outcome of the DRMV model when the uncertainty size $\delta = \infty$. We implement our experiments in the same environment as in Chapter 3 (Section 3.4), with some necessary adjustments.

4.4.1 Experiment Design and Data Preparation

We backtested for the period January 2000–December 2016 with the training (estimation) period being January 1991–December 1999 (i.e., the previous ten years). All of the stock monthly price data were obtained from the database of Wharton Business School. At the beginning of 2000, we *randomly* chose 100 stocks from the constituents of the S&P 500 that have at least ten years' historical price data available. The basic period is set to be one year in all of the single-period models involved with target annual mean return rate fixed at $\rho = 10\%$ ¹. Then, we used the training period to estimate the out-of-sample parameters, namely the mean and the variance, to construct the optimal strategies of the various models tested. The classical single-period Markowitz model and the DRMV model are described in detail in Chapter 3.

¹In this experiment, only the classical Markowitz model and DRMV need to specify the target return parameter ρ . Here, we follow Section 3.4 in choosing $\rho = 10\%$.

DRSR model

Let us first describe the construction of the DRSR strategy for the selected 100 stocks. We generated this 17-year long strategy in an (overlapping) rolling horizon fashion, with each horizon being one month. Specifically, on the first trading day of January 2000, we solved our DRSR model to obtain a portfolio, denoted as ϕ_R . In doing so, we set $p = q = 2$ and obtained the parameter δ using the menu presented at the end of Section 4.3. We then substituted δ in the optimization problem described in Theorem 4 to obtain ϕ_R . We retained ϕ_R only until the first trading day of February 2000. At that point we re-estimated the parameter δ using the *immediate* previous ten-year (namely February 1991 – January 2000) price data, re-solved the DRSR model, and generated a new portfolio ϕ_R for February 2000, the second month in our backtesting period. We repeated the same steps for all of the subsequent months.

If at the beginning of a month some stocks in our portfolio had been removed from the S&P 500 during the previous month, then we removed them from our portfolio, replaced them with the same number of stocks that were randomly chosen from the S&P 500 (yet having at least 10 years' historical data), and then re-balanced based on our DRSR model. We still denoted by ϕ_R the overall portfolio for the 17-year period and kept track of the wealth process that had been updated at the end of each month.

In what follows we describe the implementation of the other models, mentioned at the beginning of this section, under comparison. Since all of the models are single-period, we applied the same monthly rolling horizon approach to build the respective strategies. Moreover, whenever stocks were dropped from S&P 500, we replaced them with exactly the same set of stocks as in the DRSR model so as to maintain consistency across the various models.

Classical Sharpe ratio model

For the classical Sharpe ratio optimization model, for each period (month) we used the sample mean and sample covariance matrix of the immediate previous ten-year return data to estimate the corresponding parameters in problem $\mathcal{Q}(\mathbb{P}_n)$. Then, we considered the following problem:

$$\min_{\phi: \mathbf{1}^\top \phi = 1, \mathbb{E}_{\mathbb{P}_n}[\phi^\top R] > 0} \frac{\sqrt{\text{Var}_{\mathbb{P}_n}(\phi^\top R)}}{\mathbb{E}_{\mathbb{P}_n}[\phi^\top R]}. \quad (4.66)$$

We generated the optimal portfolio of this model, ϕ_{CSR} , by solving the above problem, on exactly the same rolling horizon basis as for the DRSR model.

Fama-French model and Black-Litterman model under Sharpe ratio

Both the Fama–French model and the Black–Litterman model were developed to address the mean-blur problem: namely, the fact that compared with variance, it is much more difficult to estimate with workable accuracy the expected returns of stocks based purely on sample means. These models estimate the stock returns by their respective methods while keeping the sample covariance matrix and feed them into the classical Sharpe ratio optimization model to obtain the corresponding strategies.

Note that in Chapter 3 we studied the Fama-French and Black-Litterman models and compared them with the DRMV model. However, therein we fed the estimates of the mean vector of stock returns into the MV model (1.1), while here the estimates are to be used in the Sharpe ratio model (4.3). Clearly, it is fair to also compare DRSR with the other models under the Sharpe ratio setting.

In implementing the Fama–French model, we first downloaded the monthly data of the three factors (i.e., $R_m - R_f$, SMB, and HML)² from Kenneth French’s personal

²We assume that the factors have been processed according to the available papers (Fama and French (1992) and Fama and French (1993)).

website.³ Then, on the first trading day of each month during the period January 2000–December 2016, for each stock, we used its immediate prior ten-year history returns to fit the three-factor Fama–French model. Next, we plugged in the factors data available on that day to obtain an estimate of the stock’s return for the month. We used these estimates for all of the randomly chosen 100 stocks as the plugged mean vector in problem (4.3), and we used the sample covariance matrix as the estimator of $\text{Var}_{\mathbb{P}_*}(R)$. Next, by solving problem (4.3) with the above estimators, we obtained the portfolio denoted as ϕ_{FSR} . This process was then repeated for the subsequent months on a rolling horizon basis.

For the Black–Litterman model, on the first trading day of each month during the period January 2000–December 2016 we calculated the implied returns of all of the S&P 500 constituent stocks having at least ten years’ historical data using the following formula:

$$R_{implied} = \lambda \Sigma \phi_{market},$$

where $\lambda = 3.07$, Σ was the sample covariance matrix of the previous ten years’ returns of these stocks and ϕ_{market} was the corresponding market portfolio (i.e., ϕ_{market} is a vector whose components add up to 1 and are proportional to the capitalizations of the S&P 500 constituents having at least ten years’ historical data) at the closing prices of the previous trading day; see Idzorek (2002).⁴ Then, we picked from $R_{implied}$ the implied returns of the 100 stocks that had been randomly chosen. We inputted these returns and the sample covariance matrix into the Sharpe ratio optimization model (4.3) to obtain the portfolio ϕ_{BSR} . This process was repeated for subsequent months on a rolling horizon fashion.

³http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

⁴Here, we include only those having at least ten years’ historical data to be consistent with the other models.

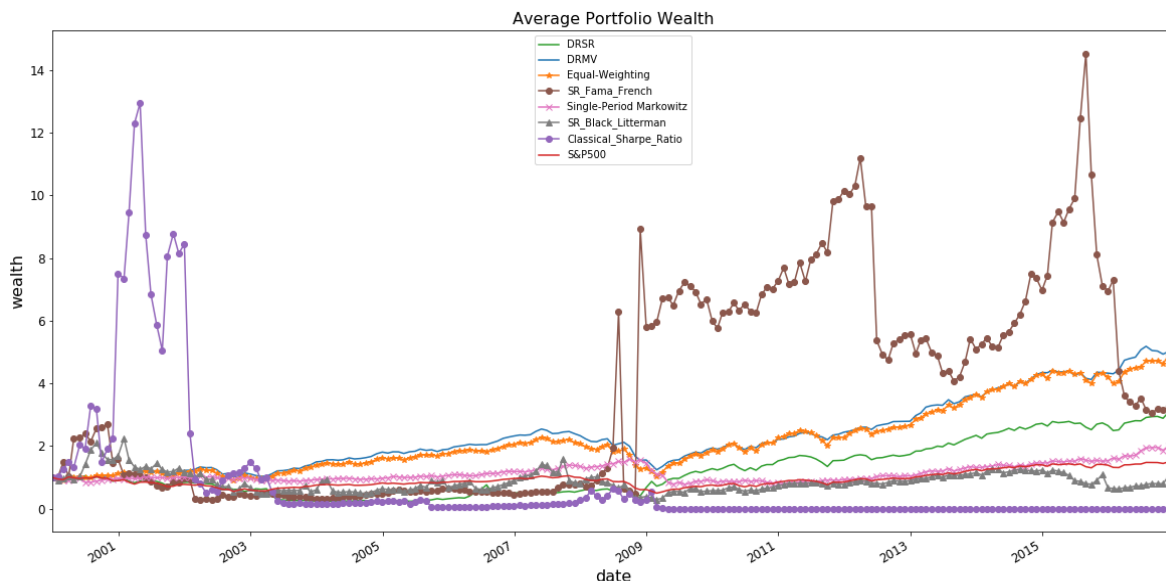


Figure 4.1: This graph presents the wealth processes of all portfolios (allow short-selling) and of the S&P 500 from January 2000 to December 2016. All of the portfolios except the S&P 500 consist of 100 stocks, and the averages are calculated over 100 numerical experiments. The x -axis indicates the time in months (from 1 to 204) and the y -axis indicates the portfolio wealth. Initial wealth is set at 1.

4.4.2 Comparisons

Assume that the initial wealth at the start of the backtesting period (i.e., January 2000) is 1. For each randomly selected set of 100 stocks, we generate the wealth process for the period 2000-2016 under each of the five models, as described in the previous subsection, as well as that under the equal weighting. Then, we repeat the experiments on 100 such sets of 100 stocks and obtain the *average* realized wealth process for each model. These processes, along with that of the S&P500 (normalized to start from 1 at the start of the testing period), are plotted in Figure 4.1.

From Figure 4.1, we can observe the extremely volatile behavior of the classical Sharpe ratio model and the Fama-French model. In order to visualize the comparison among other portfolios, it is necessary to remove the wealth curves of these two models, resulting in Figure 4.2. DRSR shows a much more stable performance than its non-robust counterpart. Also, in terms of final realized wealth, DRSR outperforms the single-period Markovitz model, the Black-Litterman model, and the S&P 500.

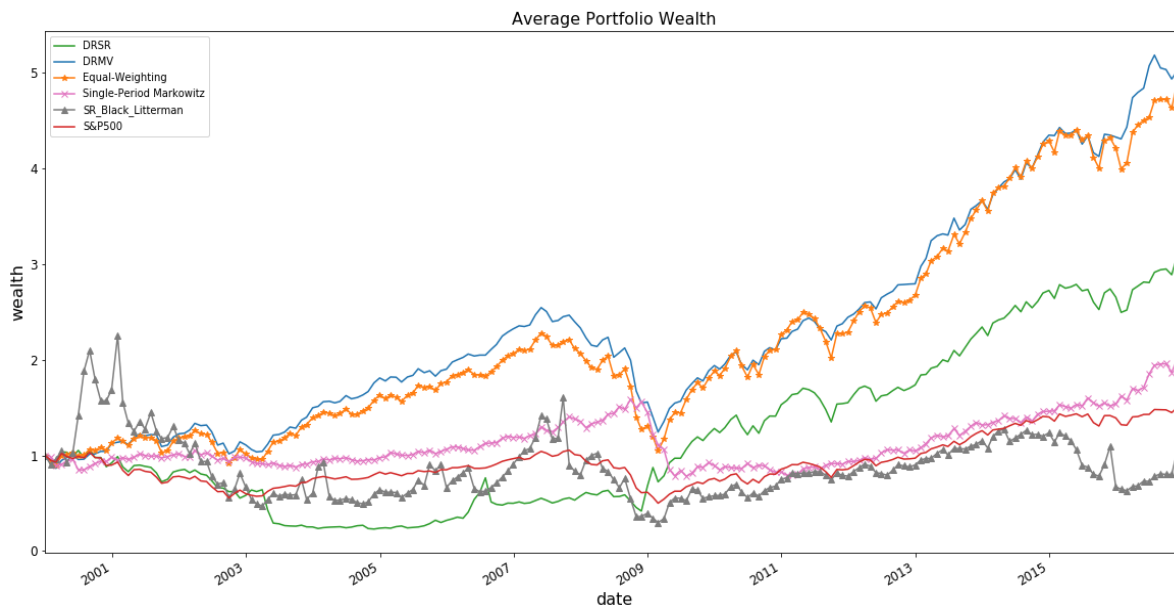


Figure 4.2: This graph presents the wealth processes of all portfolios (allow short-selling and without classical Sharpe ratio and Fama-French model) and of the S&P 500 from January 2000 to December 2016. All the portfolios except the S&P 500 consist of 100 stocks, and the averages are calculated over 100 numerical experiments. The x -axis indicates the time in months (from 1 to 204) and the y -axis indicates the portfolio wealth. Initial wealth is set at 1.

Thus, from Figure 4.1 and Figure 4.2, we can conclude that DRSR represents a significant improvement on the classical Sharpe ratio model.

However, DRSR is still worse than the equally weighted and DRMV portfolios. This is due to the fact that the classical Sharpe ratio optimization model (4.3) becomes very unstable when the expected portfolio return is close to 0. In order to avoid this disadvantage, DRSR adds a constraint to the portfolio, which is that $\phi^\top \mathbb{E}_{\mathbb{P}_n}[R] > \sqrt{\delta} \|\phi\|_p$ in the feasible region (4.11). This is a very aggressive constraint on portfolio return in practice. Specifically, using the optimal portfolio ϕ_R of the DRSR model to compute $\sqrt{\delta} \|\phi_R\|_2$, we usually observe that the obtained value is higher than 30%, which means that $\phi_R^\top \mathbb{E}_{\mathbb{P}_n}[R] > 30\%$. This is a very aggressive target for portfolio return in practice. Due to this requirement, the performance of DRSR is not very good, although it is much more robust than the classical Sharpe ratio model.

In the above discussion, all of the portfolios allow short-selling. If we do not al-

low short-selling, their performance becomes much more stable. This result is shown in Figure 4.3. With no short-selling, almost all of the portfolios have better performances. In terms of final realized wealth, DRMV and equal-weight outperform the other portfolios, and the rest of the portfolios have similar performance except Black-Litterman.

We can study more closely the variability of the performance and the overall return-risk efficiency of these models by examining their histograms of annualized returns (i.e., the distributions of the annualized returns of the 100 experiments) and those of their Sharpe ratios. These data are plotted in Figure 4.4 - 4.15.

We first compare DRSR with the classical Sharpe ratio model; see Figures 4.4 and 4.5. In both histograms, DRSR is more concentrated than the classical Sharpe ratio model, which shows that DRSR is more stable. Similar things happen when we compare DRSR with Fama-French (Figures 4.10 and 4.11). This again illustrates that, while DRSR has similar average performance to Fama-French, DRSR outperforms Fama-French in terms of variability. As for the Black-Litterman model, see Figures 4.8 and 4.9, DRSR wins in both average performance and robustness.

When it comes to the single-period Markovitz model, things are different. Although DRSR is more concentrated in both annualized returns and Sharpe ratios, the average Sharpe ratio of the Markovitz model is slightly higher than that of DRSR. It is interesting that, in order to optimize the Sharpe ratio, the Markovitz model may be a better choice than the Sharpe ratio optimization model. Again, one important reason may be owing to the unstable property of the Sharpe ratio when portfolio return is close to 0, which does not exist in the MV model.

Finally, we compare DRSR with DRMV and equally weighted portfolios; see Figure 4.12 to Figure 4.15. Both the equally weighted and the DRMV portfolios have superior performance to DRSR, which suggests that Sharpe ratio may not be a proper criterion for portfolio selection problems.

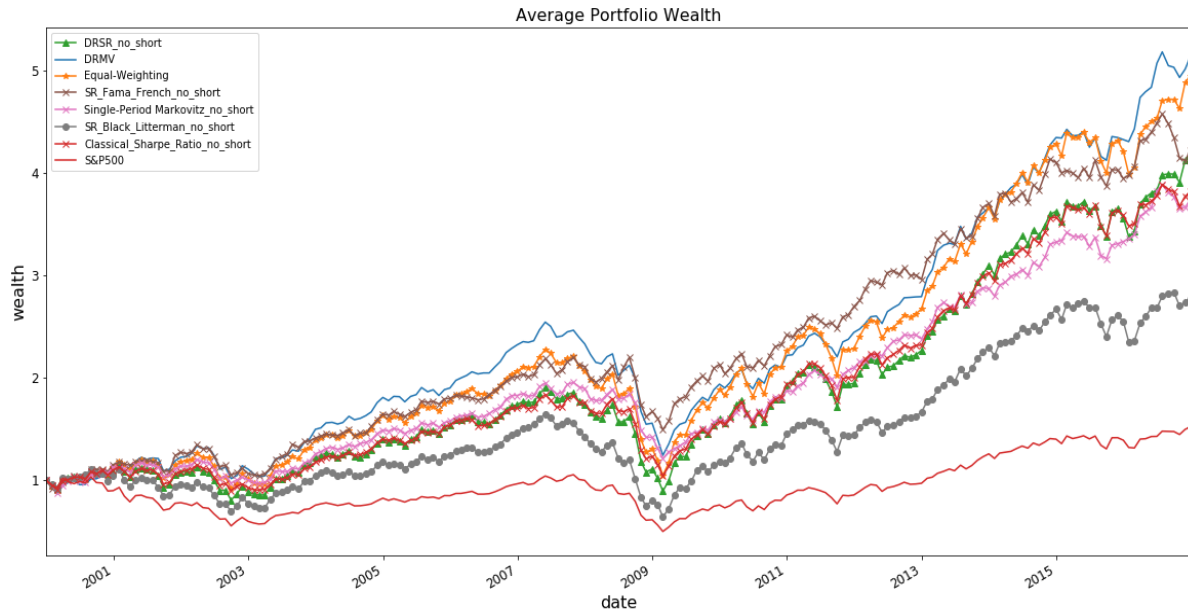


Figure 4.3: This graph presents the wealth processes of all portfolios (NOT allowing short-selling) and of the S&P 500 from January 2000 to December 2016. All the portfolios except S&P 500 consist of 100 stocks, and the averages are calculated over 100 numerical experiments. The x -axis indicates the time in months (from 1 to 204) and the y -axis indicates the portfolio wealth. Initial wealth is set at 1.

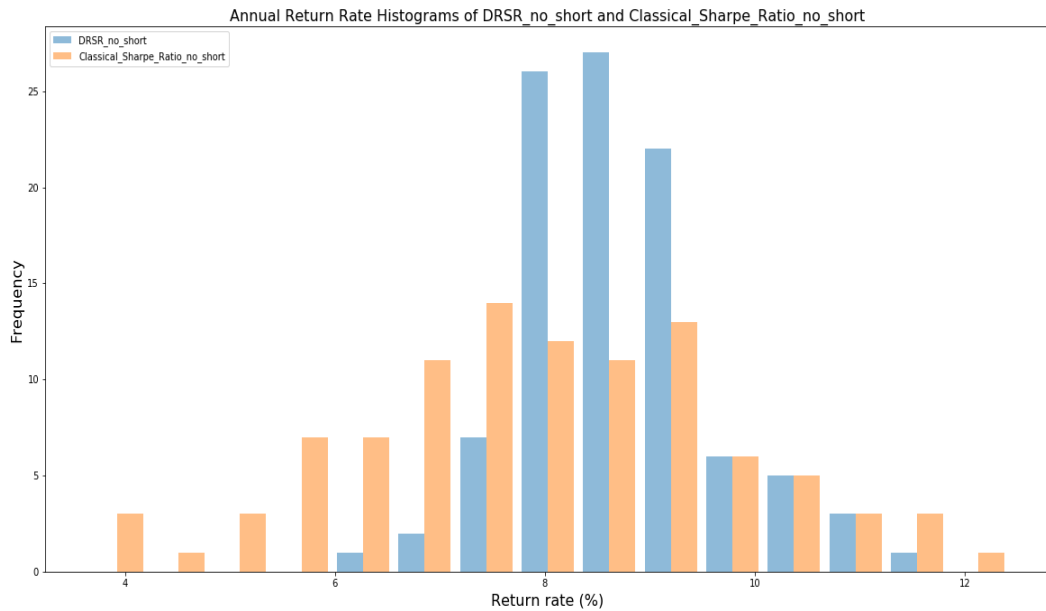


Figure 4.4: This graph presents the histograms of the annualized returns of the 100 different experiments on the DRSR (blue) and classical Sharpe ratio (orange) portfolios. The x -axis represents the annualized return and the y -axis represents the number of returns.

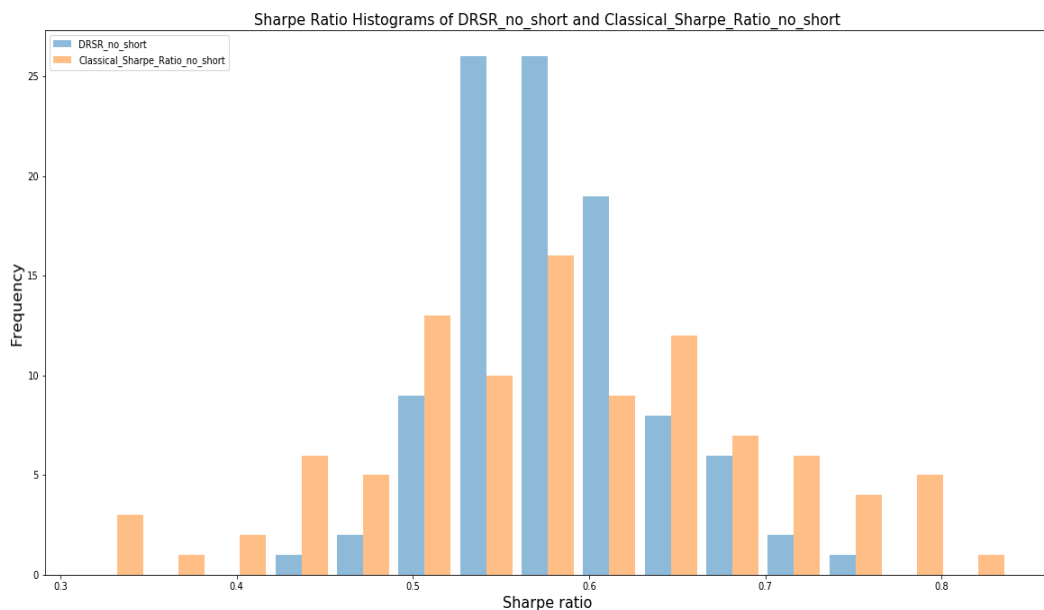


Figure 4.5: This graph presents the histograms of the Sharpe ratio of the 100 different experiments on the DRSR (blue) and classical Sharpe ratio (orange) portfolios. The x -axis represents the Sharpe ratio and the y -axis represents the number of Sharpe ratios.

4.5 Concluding Remarks

We have formulated a data-driven distributionally robust model for a Sharpe ratio optimization problem. Based a similar approach to that employed in Chapter 3, the newly developed model can be equivalently transformed into a convex optimization problem that is tractable. Unlike the elegant penalized formulation of DRMV’s objective function, the objective function of DRSR is much more complicated. However, in DRSR, we do not need to select the target return ρ , which makes our solution completely data-driven. Moreover, the RWP function to determine δ is more concise. In the numerical experiment, DRSR does show significant improvements over the classical Sharpe ratio model, as well as over some famous models like Fama-French and Black-Litterman. Nevertheless, DRSR performed worse than the equally weighted and DRMV strategies in both return and stability. The reason lies in the formulation of the Sharpe ratio and further research is needed to modify it so as to overcome its

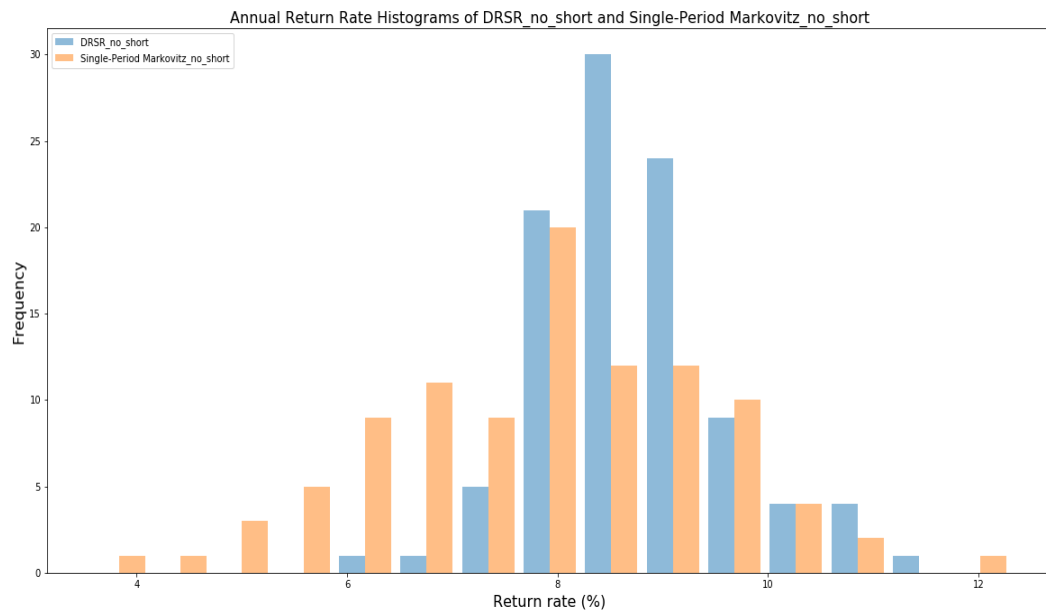


Figure 4.6: This graph presents the histograms of the annualized returns of the 100 different experiments on the DRSR (blue) and single-period Markovitz (orange) portfolios. The x -axis represents the annualized return and the y -axis represents the number of returns.

volatile behavior when portfolio return is close to 0.

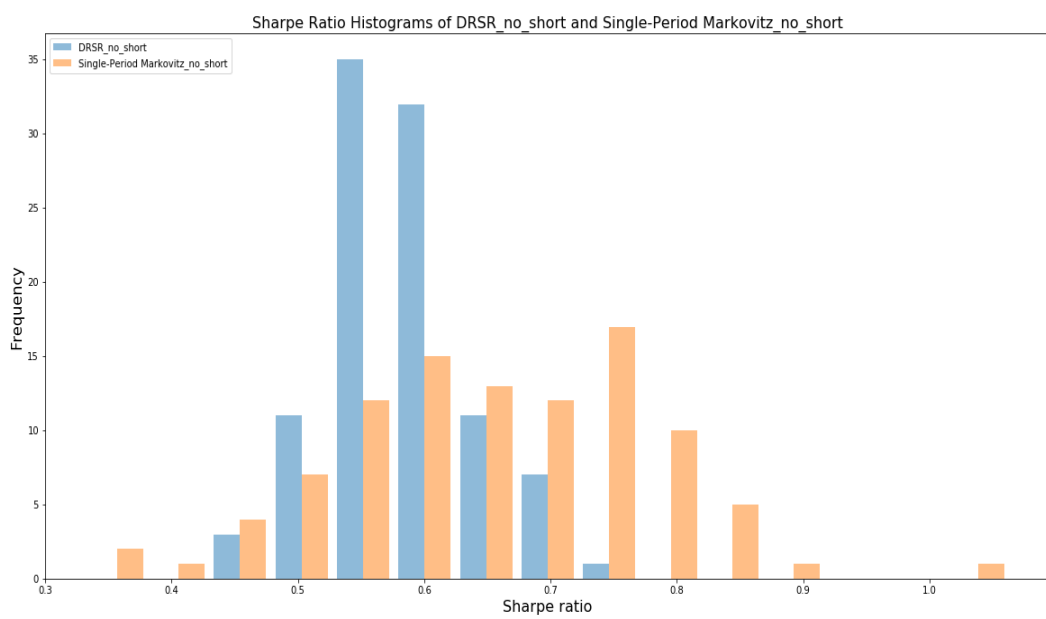


Figure 4.7: This graph presents the histograms of the Sharpe ratio of the 100 different experiments on the DRSR (blue) and single-period Markovitz (orange) portfolios. The x -axis represents the Sharpe ratio and the y -axis represents the number of Sharpe ratios.

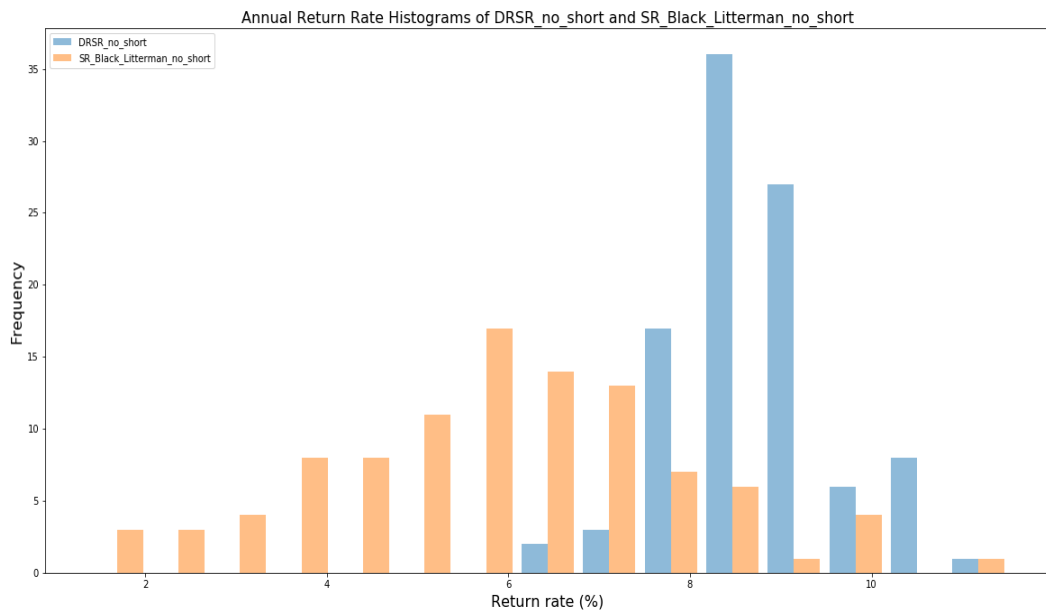


Figure 4.8: This graph presents the histograms of the annualized returns of the 100 different experiments on the DRSR (blue) and Black-Litterman (orange) portfolios. The x -axis represents the annualized return and the y -axis represents the number of returns.

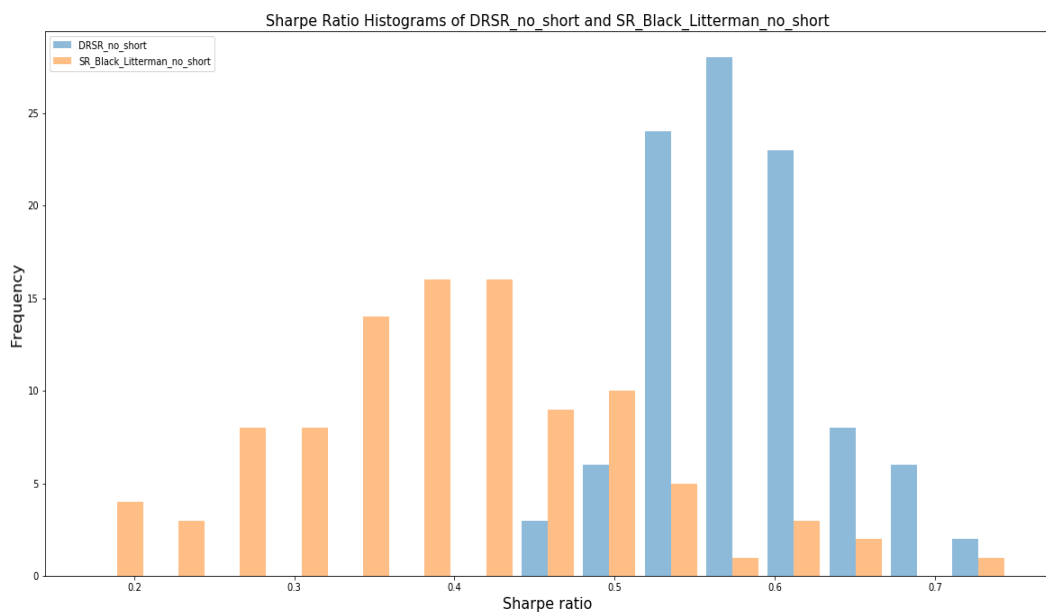


Figure 4.9: This graph presents the histograms of the Sharpe ratio of the 100 different experiments on the DRSR (blue) and Black-Litterman (orange) portfolios. The x -axis represents the Sharpe ratio and the y -axis represents the number of Sharpe ratios.

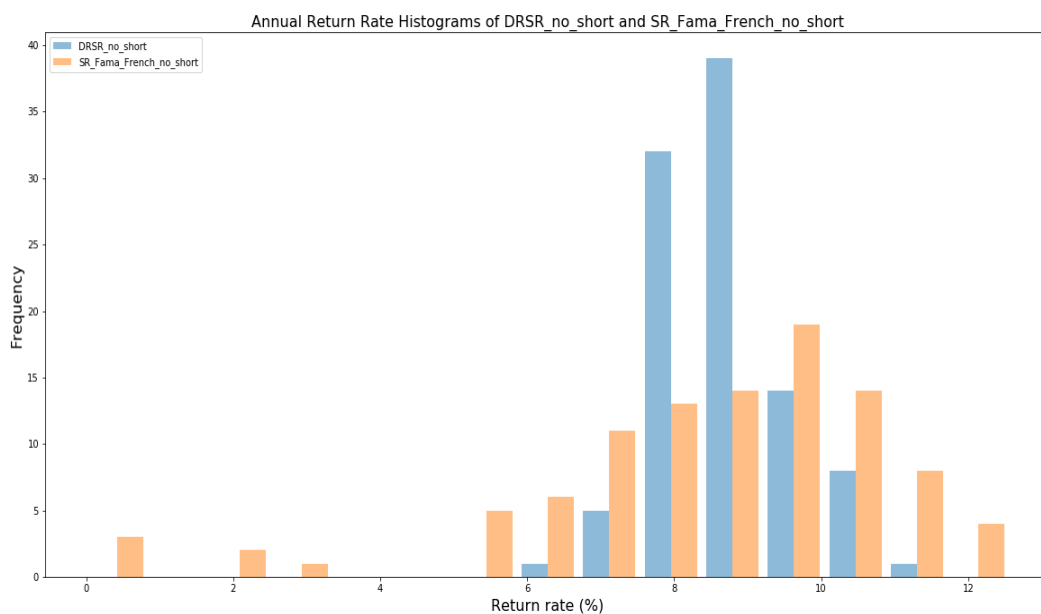


Figure 4.10: This graph presents the histograms of the annualized returns of the 100 different experiments on the DRSR (blue) and Fama-French (orange) portfolios. The x -axis represents the annualized return and the y -axis represents the number of returns.

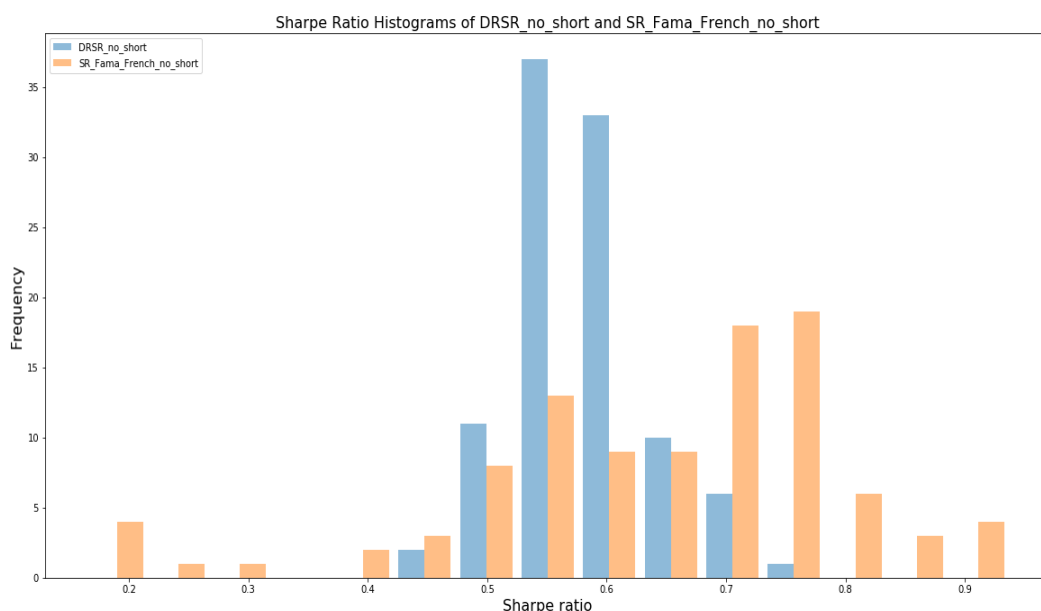


Figure 4.11: This graph presents the histograms of the Sharpe ratio of the 100 different experiments on the DRSR (blue) and Fama-French (orange) portfolios. The x -axis represents the Sharpe ratio and the y -axis represents the number of Sharpe ratios.

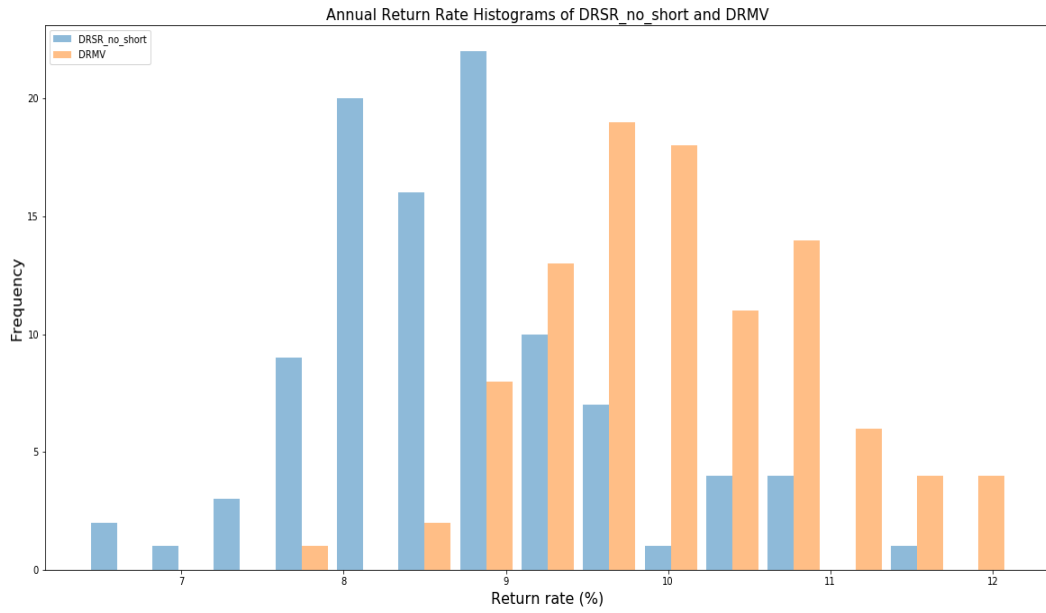


Figure 4.12: This graph presents the histograms of the annualized returns of the 100 different experiments on the DRSR (blue) and DRMV (orange) portfolios. The x -axis represents the annualized return and the y -axis represents the number of returns.

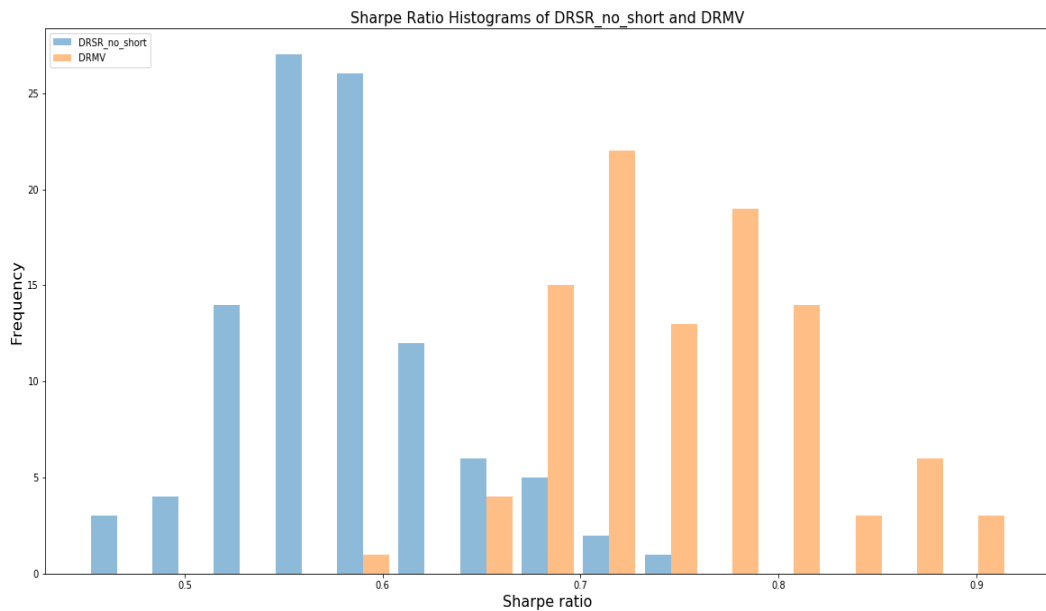


Figure 4.13: This graph presents the histograms of the Sharpe ratio of the 100 different experiments on the DRSR (blue) and DRMV (orange) portfolios. The x -axis represents the Sharpe ratio and the y -axis represents the number of Sharpe ratios.

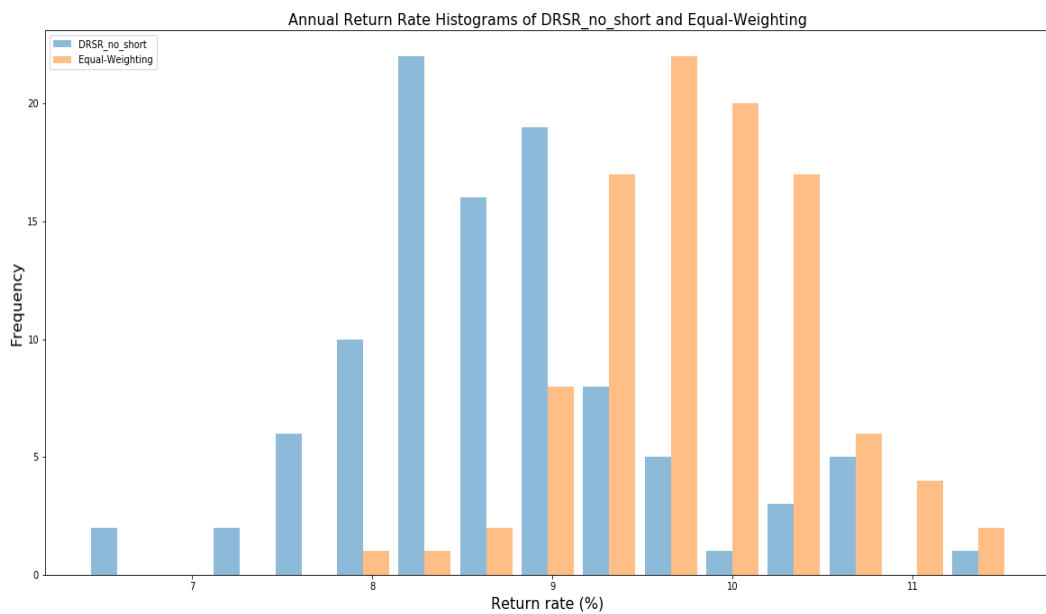


Figure 4.14: This graph presents the histograms of the annualized returns of the 100 different experiments on the DRSR (blue) and equal-weighting (orange) portfolios. The x -axis represents the annualized return and the y -axis represents the number of returns.

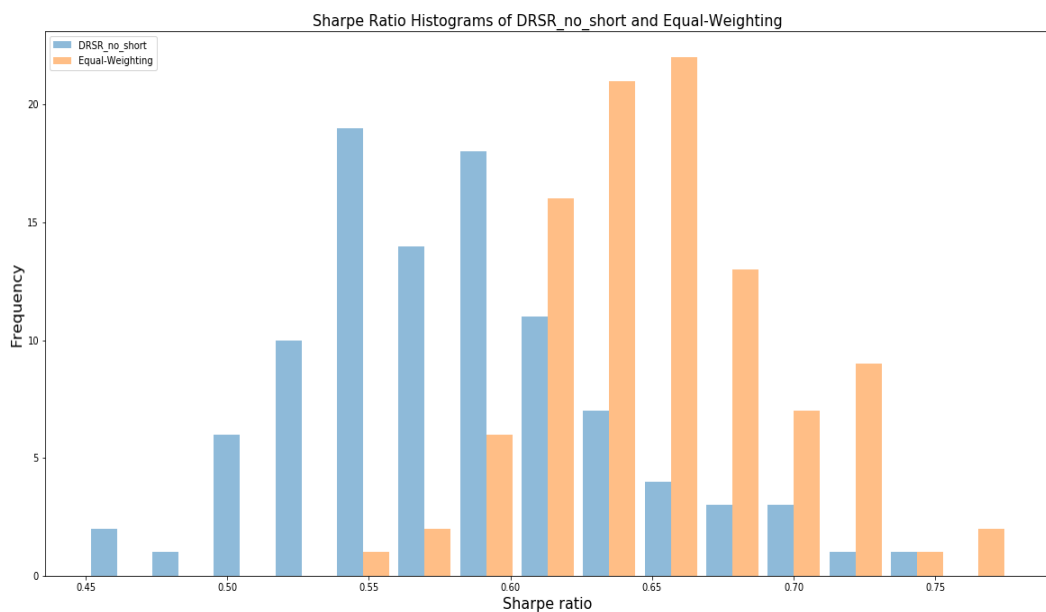


Figure 4.15: This graph presents the histograms of the Sharpe ratio of the 100 different experiments on the DRSR (blue) and equal-weighting (orange) portfolios. The x -axis represents the Sharpe ratio and the y -axis represents the number of Sharpe ratios.

Bibliography

- Basak, S. and Chabakauri, G. Dynamic mean-variance asset allocation. *Review of Financial Studies*, 23:2970–3016, 2010.
- Bayraksan, G. and Love, D. Data-driven stochastic programming using phi-divergences. *Tutorials in Operations Research*, 2015.
- Ben-Tal, A. and Nemirovski, A. Robust convex optimization. *Mathematics of Operations Research*, 23:769–805, 1998.
- Ben-Tal, A. and Nemirovski, A. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25:1–13, 1999.
- Bielecki, T., Jin, H., Pliska, S., and X., Z. Continuous-time mean-variance portfolio selection with bankruptcy prohibition. *Mathematical Finance*, 15:213–244, 2005a.
- Bielecki, T., Jin, H., Pliska, S., and Zhou, X. Continuous-time mean-variance portfolio selection with bankruptcy prohibition. *Mathematical Finance*, 15:213–244, 2005b.
- Bjork, T. and Murgoci, A. A general theory of markovian time inconsistent stochastic control problems. *working paper*, 2010.
- Bjork, T., Murgoci, A., and Zhou, X. Mean-variance portfolio optimization with state-dependent risk aversion. *Mathematics Finance*, 24:1–24, 2014.
- Blanchet, J. and Kang, Y. Distributionally robust groupwise regularization estimator. *arXiv:https://arxiv.org/abs/1705.04241*, 2017.

- Blanchet, J. and Murthy, K. Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, 44(2):565–600, 2019.
- Blanchet, J., Kang, Y., and Murthy, K. Robust wasserstein profile inference and applications to machine learning. *arXiv:https://arxiv.org/abs/1610.05627*, 2016.
- Blanchet, J., Kang, Y., Zhang, F., and Murthy, K. Data-driven optimal transport cost selection for distributionally robust optimization. *arXiv:https://arxiv.org/abs/1705.07152*, 2017.
- Blanchet, J., Murthy, K., and Si, N. Confidence regions in wasserstein distributionally robust estimation. *https://arxiv.org/abs/1906.01614*, 2019.
- Chen, A., Jen, F., and Zions, S. The optimal portfolio revision policy. *The Journal of Business*, 44:51–61, 1971.
- Chen, L. and Zhou, X. Naive strategies in a general linear quadratic model. *Working Paper*, 2020.
- Chen, W. and Sim, M. Goal-driven optimization. *Operations Research*, 57:342–357, 2009.
- Chen, X., Sim, M., and Sun, P. A robust optimization perspective on stochastic programming. *Operations Research*, 55:1058–1071, 2007.
- Chopra, V. and Ziemba, W. The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19:6–11, 1993.
- Cui, X., Gao, J., and Li, D. Continuous-time mean-variance portfolio selection with finite transactions. *Stochastic analysis and applications to finance*, 13:77–98, 2012.
- Dai, M., Xu, Z., and Zhou, X. Continuous-time markowitz’s model with transaction costs. *SIAM Journal on Financial Mathematics*, 1:96–125, 2010.

- Delage, E. and Ye, Y. Data-driven distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58: 595–612, 2010.
- Delage, E., Kuhn, D., and Wiesemann, W. Decision-making under uncertainty: When can a random decision reduce risk? *Management Science*, 65:3282–3301, 2019.
- DeMiguel, V. and Nogales, F. Portfolio selection with robust estimation. *Operations Research*, 57:1–18, 2009.
- Dupacova, J. The minimax approach to stochastic programming and an illustrative application. *Stochastics*, 20:73–88, 1987.
- El Ghaoui, L., Oks, M., and Oustry, F. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51:543–556, 2003.
- Elton, E. and Gruber, M. On the optimality of some some multiperiod portfolio selection criteria. *The Journal of Business*, 47:231–243, 1974a.
- Elton, E. and Gruber, M. The multi-period consumption investment problem and single period analysis. *Oxford Economic Papers*, 26:289–301, 1974b.
- Erdogan, E. and Iyengar, G. Ambiguous chance-constrained problems and robust optimization. *Mathematical Programming*, 107:37–61, 2006.
- Esfahani, P. and Kuhn, D. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):115–166, 2018.
- Fabozzi, F., Kolm, P., Pachamanova, D., and Focardi, S. Robust portfolio optimization. *Journal of Portfolio Management*, 33(3):40–48, 2007.

- Fama, E. F. and French, K. R. The cross-section of expected stock returns. *The Journal of Finance*, 47(2):427–465, 1992.
- Fama, E. F. and French, K. R. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56, 1993.
- Gao, R. and Kleywegt, A. Distributionally robust stochastic optimization with wasserstein distance. <https://arxiv.org/abs/1604.02199>, 2016.
- Gao, R., Chen, X., and Kleywegt, A. Wasserstein distributional robustness and regularization in statistical learning. <https://arxiv.org/abs/1712.06050>, 2017.
- Garlappi, L., Uppal, R., and Wang, T. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies*, 20:41–81, 2007.
- Glasserman, P. and Xu, X. Robust risk measurement and model risk. *Quantitative Finance*, 14:29–58, 2014.
- Goldfarb, D. and Iyengar, G. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38, 2003.
- Hakansson, N. Multi-period mean-variance analysis: Toward a general theory of portfolio choice. *The Journal of Finance*, 26:857–884, 1971.
- Hampel, F. *Contribution to the theory of robust estimation*. PhD thesis, University of California, Berkeley, CA., 1964.
- He, X. and Jiang, Z. Dynamic mean-risk asset allocation and myopic strategies: A universal portfolio rule. *To be appear*, 2017.
- Huber, P. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35:73–101, 1964.

- Idzorek, T. A step-by-step guide to the black-litterman model. *Technical Report*, 2002.
- Jiang, R. and Guan, Y. Data-drive chance constrained stochastic program. *Mathematical Programming*, 158:291–327, 2016.
- Jin, H. and Zhou, X. Continuous-time portfolio selection under ambiguity. *Mathematical Control and Related Fields*, 5:475–488, 2015.
- Kantorovich, L. On the transfer of masses (in russian). *Doklady Akademii Nauk*, 37:227–229, 1942.
- Karatzas, I. and Shreve, S. E. *Methods of Mathematical Finance*. Springer, 1998.
- Lagoa, C. and Barmish, B. Distributionally robust monte carlo simulation: A tutorial survey. *IFAC Proceedings Volumes*, 35:151–162, 2002.
- Li, D. and Ng, W. Optimal dynamic portfolio selection: Multiperiod mean-variance formulation. *Mathematical Finance*, 10:387–406, 2000.
- Lim, A. E. B. and Zhou, X. Quadratic hedging and mean-variance portfolio selection with random parameters in a complete market. *Mathematics of Operations Research*, 27:101–120, 2002.
- Lintner, J. Valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economic Studies*, 47:13–37, 1965.
- Lobo, M. and Boyd, S. The worst-case risk of a portfolio. https://web.stanford.edu/~boyd/papers/pdf/risk_bnd.pdf, 2000.
- Lopes, L. Between hope and fear: The psychology of risk. *Advances in Experimental Social Psychology*, 20:255–295, 1987.
- Lu, Z. Robust portfolio selection based on a joint ellipsoidal uncertainty set. *Optimization Methods and Software*, 26:89–104, 2011.

- Markowitz, H. Portfolio selection. *Journal of Finance*, 7:77–91, 1952.
- Markowitz, H. *Portfolio Selection: Efficient Diversification of Investment*. Wiley, New York, 1959.
- Michaud, R. The markowitz optimization enigma: Is optimized optimal? *Financial Analysts*, 45:31–42, 1989.
- Monge, G. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 1781.
- Mossin, J. Equilibrium in capital asset markets. *Econometrica*, 34:768–783, 1966.
- Natarajan, K. and Sim, M. Tractable robust expected utility and risk models for portfolio optimization. *Mathematical Finance*, 20:695–731, 2010.
- Ollila, E. and Raninen, E. Optimal shrinkage covariance matrix estimation under random sampling from elliptical distribution. <https://arxiv.org/abs/1808.10188>, 2018.
- Pendersen, J. L. and Peskir, G. Optimal mean-variance portfolio selection. *Mathematics and Financial Economics*, 11:137–160, 2017.
- Perret-Gentil, C. and Victoria-Feser, M. Robust mean-variance portfolio selection. *Available at SSRN 721509*, 2005.
- Pflug, G. and Wozabal, D. Ambiguity in portfolio selection. *Quantitative Finance*, 7: 435–442, 2007.
- Popescu, I. Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55:98–112, 2007.
- Richardson, H. R. A minimum variance result in continuous trading portfolio optimization. *Management Science*, 35:1045–1055, 1989.

- Samuelson, P. Lifetime portfolio selection by dynamic stochastic programming. *The review of economics and statistics*, 50:239–246, 1969.
- Scarf, H. A min-max solution of an inventory problem. *Stanford University Press, Stanford, CA*, pages 201–209, 1958.
- Shapiro, A. and Ahmed, S. On a class of minimax stochastic programs. *SIAM Journal on Optimization*, 14:1237–1249, 2004.
- Shapiro, A. and Kleywegt, A. Minimax analysis of stochastic programs. *Optimization Methods and Software*, 17:523–542, 2002.
- Sharpe, W. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19:425–442, 1964.
- Strotz, R. H. Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies*, 23:165–180, 1956.
- Tütüncü, R. and Koenig, M. Robust asset allocation. *Annals of Operations Research*, 132:157–187, 2004.
- Villani, C. *Topics in optimal transportation*, volume 58. Graduate Studies in Mathematics, American Mathematics Society, Providence, RI., 2003.
- Wang, Z., Glynn, P., and Ye, Y. Likelihood robust optimization for data-driven problems. *Computational Management Science*, 13:241–261, 2016.
- Welsch, R. and Zhou, X. Application of robust statistics to asset allocation models. *Revstat, Statistical Journal*, 5:97–114, 2007.
- Wozabal, D. A framework for optimization under ambiguity. *Annals of Operations Research*, 193:21–47, 2012.
- Xia, J. Mean-variance portfolio choice: Quadratic partial hedging. *Mathematical Finance*, 15:533–538, 2005.

- Yaari, M. E. The dual theory of choice under risk. *Econometrica*, 55(1):95–115, 1987.
- Ye, K., Parpas, P., and Rustem, B. Robust portfolio optimization: A conic programming approach. *Computational Optimization and Applications*, 52:463–481, 2012.
- Yin, G. and Zhou, X. Markowitz’s mean-variance portfolio selection with regime switching: From discrete-time models to their continuous-time limits. *IEEE Transactions on automatic control*, 49:349–360, 2004.
- Zackova, J. On minimax solution of stochastic linear programming problems. *Casopis pro Pestovani Matematiky*, 91:423–430, 1966.
- Zhao, C. and Guan, Y. Data-driven risk-averse stochastic optimization with wasserstein metric. *Operations Research Letters*, 46(2):262–267, 2018.
- Zhou, X. and Li, D. Continuous-time mean-variance portfolio selection: a stochastic lq framework. *Applied Mathematics*, 42:19–33, 2000.
- Zymler, S., Kuhn, D., and Rustem, B. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137:167–198, 2013.