Advances in risk parity portfolio optimization

by

Giorgio Costa Del Pozo

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Department of Mechanical and Industrial Engineering
University of Toronto

Abstract

Advances in risk parity portfolio optimization

Giorgio Costa Del Pozo

Doctor of Philosophy

Department of Mechanical and Industrial Engineering

University of Toronto

2021

Risk parity is an asset allocation strategy that seeks to equalize the risk contributions of the constituent assets in a portfolio. The resulting portfolio is fully diversified from a risk perspective. However, like other asset allocation strategies, risk parity is susceptible to estimation errors. Moreover, its mathematical formulation imposes some fundamental limitations. This thesis aims to modernize risk parity by addressing all of the aforementioned issues. We address the susceptibility to estimation errors through three different frameworks. First, we introduce a robust framework that quantifies estimation error and embeds this information during optimization to construct a robust risk parity portfolio. Our second framework takes a different approach, introducing robustness during the parameter estimation step. This is formulated as a game-theoretic minimax problem to make an optimal investment decision against the most adversarial estimate of our parameters. Our third framework improves the quality of our estimated parameters before optimization takes place. We posit that we can embed the cyclical information of financial markets directly into our estimates, resulting in risk parity portfolios aligned with the current market regime. The result is a Markov regime-switching factor model of asset returns from which we can naturally derive regime-dependent parameters for use during optimization. The final component of this thesis addresses the fundamental limitations of risk

parity: its lack of accountability for the investor's risk and reward appetite and its prohibition of short sales. We propose a generalized risk parity framework where the investor's risk and reward appetite define our objective, while still enforcing a desirable degree of risk-based diversification. Moreover, we propose an algorithm that allows us to consider portfolios with short positions. Thus, our generalized framework addresses the fundamental limitations of risk parity while retaining the desirable property of risk-based diversification. The frameworks proposed in this thesis can be used independently or in tandem, depending on the investor's needs and goals. The unifying subject of this thesis is to advance risk parity by addressing its fundamental weaknesses. This is achieved by proposing different frameworks and algorithms, with the overarching property of preserving the interpretability and computational tractability of our solutions.

To my parents, Marion and Luis, the companion of all of my childhood adventures, Antonella (Antonietta), and my partner in life, Lauren.

Acknowledgements

First and foremost, I would like to give my most profound thanks to my supervisor, Professor Roy H. Kwon. Watching him work has been an inspiration; his creativity, energy and passion for his work are contagious. My own work was reinvigorated after each and every one of our research meetings. However, beyond the traditional values that make him an excellent supervisor, I must add that I am eternally grateful to him for having faith in me and my work. His belief in me has, in turn, increased my own confidence. I am certain that I am leaving my doctoral studies as a better person altogether than when I started, and I could not have done it without him.

I would also like to thank my committee members: Professors Ken Jackson, Oleksandr Romanko, Yuri Lawryshyn, Luis Seco and David Saunders. Special thanks go to Ken and Oleksandr, who were part of my internal committee since the beginning of my doctoral studies: from my Qualifying Exam until my Final Oral Exam. I cannot thank them enough for their continuous support over the years. In addition, I would like to thank David for kindly agreeing to be my external examiner.

My time at the University of Toronto has been marvellous, and I owe much of that to the friends I made while working together at the Financial Optimization and Risk Management laboratory. Special thanks go to Vaughn Gambeta and Alexia Yeo for their companionship. In particular, I would like to thank my dear friend Hassan Anis for his support. I must say that many of the moments of clarity and breakthroughs in my research came from our many conversations over a cup of coffee. My appreciation of coffee also increased significantly from these 'hang outs'.

I would like to thank Professor Kwon a second time, as well as the MIE Department and the Engineering Science Division at the University of Toronto. As a student, I always looked at my instructors from a distance, marvelled by their level of confidence and their sheer amount of knowledge. I never imagined that my doctoral studies would grant me a chance to serve as a lecturer once, let alone six different times. Once again, Professor Kwon's faith in me lead to my first opportunity, and since then I have enjoyed every part of being a teacher. This has been one of the most rewarding opportunities I have ever had.

We often hear the saying about 'standing on the shoulders of giants' when making academic progress. I am sure this is true, but I like to think that I stand on the shoulders of my parents,

Luis and Marion, and my sister, Antonella. I owe to them all of my accomplishments in life. They have nurtured, educated and supported me my whole life, making me who I am today. More importantly, their love and encouragement have always kept me going forward. Although there is some geographical distance between us, I always carry them in my heart.

I have saved the very last part of my acknowledgements to the person whose encouragement made all of this possible. Lauren, I would not have had the courage to pursue my dreams if it were not for you. Being a graduate student is full of ups and downs, and moments of darkness and light. I was only able to endure and navigate these challenges because you were there with me every step of the way; you have been (and continue to be) my North Star. I would not have been able to find my way to this moment without you. More importantly, I look forward to the many more adventures you and I will have together in the future. Thank you, Lauren, for your unconditional love and never-ending support; I could not picture my life without you.

Contents

1 Introduction **1**

1.1 Thesis outline and contribution 4

1.1.1 Chapter 2: Modern portfolio theory and risk parity 4

1.1.2 Chapter 3: A robust framework for risk parity 5

1.1.3 Chapter 4: Distributionally robust risk parity 6

1.1.4 Chapter 5: A regime-switching factor model for risk parity 7

1.1.5 Chapter 6: A generalized risk parity framework 8

1.2 Notation 9

2 Modern portfolio theory and risk parity **12**

2.1 Portfolio optimization and factor models 13

2.1.1 Risk and reward 13

2.1.2 Factor models 14

2.1.3 Mean–Variance Optimization 17

2.2 Risk parity 18

2.2.1 Non-convexity of risk parity 20

2.2.2 Long-only convex risk parity problems 22

2.3 Measures of financial performance and risk concentration 23

3 A robust framework for risk parity **26**

3.1 Covariance uncertainty structure 28

3.1.1 Derivation of the uncertainty structure from factor models 29

3.2 Robust risk parity 30

3.3 Numerical experiments 33

3.3.1 Experiment with varying degrees of robustness 35

3.3.2 Experiment with different portfolio sizes . . . 41
3.4 Conclusion . . . 43

4 Distributionally robust risk parity **46**
4.1 Preliminaries . . . 50
4.1.1 Estimation of parameters . . . 50
4.1.2 Distributionally robust optimization . . . 53
4.1.3 Statistical distance measures . . . 55
4.2 Distributionally robust risk parity . . . 56
4.2.1 Probability distribution ambiguity set . . . 57
4.2.2 Minimax problem . . . 60
4.2.3 Projected gradient descent–ascent . . . 62
4.2.4 Sequential convex programming with projected gradient ascent . . . 65
4.3 Numerical Experiments . . . 69
4.3.1 Numerical performance and tractability . . . 69
4.3.2 In-sample experiment . . . 74
4.3.3 Out-of-sample experiment . . . 78
4.4 Conclusion . . . 79

5 A regime-switching factor model for risk parity **83**
5.1 Regime-switching factor model . . . 85
5.1.1 Selection of market regimes . . . 87
5.2 Regime-dependency in risk parity . . . 89
5.3 Numerical experiments . . . 90
5.4 Conclusion . . . 96

6 A generalized risk parity framework **98**
6.1 Generalized risk parity . . . 100
6.1.1 Robust generalized risk parity . . . 101
6.1.2 Lowest-variance risk parity . . . 103
6.2 Sequential tightening via ADMM . . . 104
6.2.1 SDP relaxation . . . 104
6.2.2 Problem decomposition and various ADMM steps . . . 105

6.2.3 ADMM algorithm . . . 110
6.3 Numerical experiments . . . 112
6.3.1 In-sample experiments . . . 114
6.3.2 Out-of-sample experiment . . . 123
6.4 Conclusion . . . 126

7 Conclusion and future research **129**
7.1 Future research . . . 132

Appendix A Numerical implementation of statistical distances **134**

Bibliography **136**

List of Tables

3.1 List of assets 34
3.2 Summary of financial performance of 1,000 trials with $n = 25$ 38
3.3 Sharpe ratio analysis of 1,000 trials with $n = 25$ 39
3.4 Summary of risk concentration measures of 1,000 trials with $n = 25$ 40
3.5 Summary of financial performance of 1,000 trials with $n = 15,\ 50,\ 75,\ 100$ . . . 42
3.6 Sharpe ratio analysis of 1,000 trials with $n = 15,\ 50,\ 75,\ 100$ 43
3.7 Summary of risk concentration measures of 1,000 trials with $n = 15,\ 50,\ 75,\ 100$ 44

4.1 List of assets 69
4.2 Comparison of numerical performance between the PGDA algorithm (A.1) and the SCP–PGA algorithm (A.2) 71
4.3 Numerical performance of the SCP–PGA algorithm 73
4.4 Portfolio variance and CV based on the nominal and worst-case estimates of the asset covariance matrix 77
4.5 Summary of financial performance of the risk parity portfolios over the periods 2000–2016 and 2007–2011 81

5.1 BIC values 89
5.2 List of assets 91
5.3 Summary of results 96

6.1 List of assets 115
6.2 Summary of in-sample results for the GRP framework 117
6.3 Summary of in-sample results for the robust GRP framework 120
6.4 Summary of in-sample results for the LVRP framework 122
6.5 Summary of out-of-sample results 127

List of Figures

3.1 Wealth evolution of 1,000 individual trials with 25 assets per portfolio over the period 2000–2016 36
3.2 Average wealth evolution of robust portfolios relative to the nominal portfolio for several values of ω 37

4.1 Asset weights of the nominal and DRRP portfolios with $\delta = 0.3$ 75
4.2 Risk contributions per asset of the nominal and DRRP portfolios with $\delta = 0.3$ . . 76
4.3 Wealth evolution for DRRP portfolios 80

5.1 Estimated regime switches from two-regime (top), three-regime (center), and four-regime (bottom) Markov models 90
5.2 Estimated regime switches 93
5.3 Portfolio wealth evolution 94

6.1 Convergence plots for the GRP framework 118
6.2 Convergence plots for the robust GRP framework 120
6.3 Convergence plots for the LVRP framework 122
6.4 Portfolio wealth evolution from 07-Jan-2000 to 30-Dec-2016 with 6-month rebalancing 125

Chapter 1

Introduction

Every asset manager, ranging from large institutions down to the individual investor, faces the challenge of making an investment decision with incomplete information. Let us decompose the previous statement. Information is 'incomplete' because it is marred by the uncertainty of future outcomes for which we only have forecasts. In other words, we must make an investment decision today based on predictions of some future outcome, with the hope that this will bring forth a positive return on our investment. Thus, the challenge is the following: how can we make the best investment decision today given that future returns are uncertain?

The topic of optimal asset allocation originated from Markowitz's seminal work, where he proposed the mathematical framework known as modern portfolio theory (MPT) [71]. MPT allows an investor to treat asset allocation and portfolio construction as a mathematical optimization problem based on forecasts of the portfolio expected return (mean) and risk (variance), also known as mean–variance optimization (MVO). However, the resulting optimal portfolios are still entirely based on forecasts, meaning they are susceptible to uncertainty and estimation error. Thus, ever since Markowitz's work, there has been a flurry of literature in the subject of optimal asset allocation and uncertainty.

In the context of MVO, optimality is based on the trade-off between portfolio risk and expected return. However, return forecasts are often unreliable; and although it may sound counterintuitive, this problem is only aggravated by optimization. In fact, the impact of estimation errors and forecasting has led to what is sometimes referred to as 'error maximization' [25, 74]. This weakness of MVO is so pervasive that it has spurred multiple advances in the literature to reduce the effects of estimation error and uncertainty [32, 49, 52, 69, 75, 96].

More importantly, this has motivated the introduction of alternative optimal asset allocation strategies.

'Risk parity' is one such alternative strategy and it has become increasingly popular among academics and practitioners in recent times. Simply put, risk parity addresses the issue of uncertainty in the return forecasts by discarding them entirely from the optimization problem. Thus, instead of targeting an optimal trade-off between risk and return, risk parity accepts that we can mitigate losses by improving diversification in our portfolio. Promoting diversification is considered as a good investment practice, and a general 'rule of thumb' to construct a diversified portfolio is to distribute wealth evenly between all constituent assets. The resulting portfolio, sometimes known as the '$1/n$' portfolio, is perfectly diversified from a wealth perspective. Although the $1/n$ portfolio is diverse, this approach ignores the risk and return forecasts of our assets. Thus, it fails to acknowledge that certain assets are riskier than others, meaning it achieves diversity naively and without regard for the possibility of the over-concentration of risk.

Risk parity is grounded in the same principle of diversity as the $1/n$ portfolio, except it seeks to be perfectly diversified from a risk perspective. As an asset allocation strategy, it achieves this goal by distributing wealth such that the risk contribution of every asset in the portfolio is equalized (i.e., each constituent in the portfolio contributes the same amount of risk towards the portfolio). Although the maxim of risk-based diversification sounds straightforward, the fact that the assets are correlated means that we must formulate risk parity as an optimization problem. Thus, the nominal risk parity portfolio optimization problem is designed such that, at optimality, we attain a portfolio where risk is equalized between the assets. Although risk parity ignores return forecasts during optimization, the resulting risk-diverse portfolio helps to mitigate losses, which in turn helps drive the return of the portfolio in the long-run [17].

As an optimal asset allocation strategy, we can see that risk parity inherits some of the weaknesses of MVO; and, we will see that it also creates a host of new issues stemming from its design. Risk parity is able to ignore the negative impact of uncertainty from return forecasts, but it is still sensitive to uncertainty in the portfolio risk measure. Thus, risk parity inherits the weakness of parameter uncertainty that permeates every aspect of portfolio optimization. Moreover, an investor may not be satisfied by the premise that we must ignore return forecasts during optimization because they are unreliable. More generally, an investor would likely prefer to have some degree of control over the overall risk and return of their portfolio, which is the

premise that originally motivated MVO and that we ignore with risk parity. Finally, the mathematical formulation of the risk parity optimization problem does not permit short sales, further restricting the financial flexibility of the investor.

Specifically, this thesis addresses the aforementioned risk parity weaknesses through the following contributions to the field:

- First, we address uncertainty in the estimated portfolio risk measure from three different perspectives.
  - We design a robust risk parity problem, where the uncertainty structure and subsequent robust reformulation are tailored to target the nuances of risk parity. Specifically, we target the estimation error in the individual asset risk contributions, as opposed to typical robust portfolio optimization problems that focus solely on the overall portfolio risk.
  - We design a distributionally robust risk parity problem that targets the estimation of the risk measure from a finite dataset. This is a fully data-driven approach where robustness in risk parity is attained directly from the discrete set of scenarios in our data.
  - We propose a regime-switching factor model to explain our asset returns. Unlike the previous two approaches, which focus on quantifying uncertainty and embedding it into the optimization problem, this one simply attempts to improve the quality of the estimated risk measure by reading latent signals from financial data. Thus, we operate under the premise that improving the quality of the estimated parameters should improve the resulting risk parity portfolio.
- Second, we reimagine the risk parity optimization problem to allow an investor to explicitly control their estimated portfolio risk and return, as well as allowing the investor to take short positions, while still promoting the risk parity goal of having a risk-diverse portfolio.

The main contribution of this thesis is to address the aforementioned weaknesses of the nominal risk parity portfolio optimization problem while adhering to the following maxim: our proposed solutions should attempt to retain the interpretability and computational tractability of the nominal problem. Thus, our proposed solutions, or 'advances' in risk parity, should not only remedy its weaknesses, but should do so in an intuitive fashion and should not come at

a prohibitively expensive computational cost. Moreover, we emphasize data-driven solutions, relieving the burden on an investor of having to rely on expert judgement. Instead, we attempt to extract this information directly from raw data. Thus, our overarching goal is to develop robust decision-making algorithms under uncertainty for a risk parity asset allocation strategy. We complete this section by presenting a brief summary of each chapter.

1.1 Thesis outline and contribution

1.1.1 Chapter 2: Modern portfolio theory and risk parity

The purpose of this chapter is to introduce and establish the preliminary content required for the development of subsequent chapters. Thus, its purpose is to conduct a literature review on the subject of portfolio optimization and risk parity, and to establish the mathematical groundwork required for the development of this thesis. Specifically, we discuss MPT [71] and mean–variance analysis. This is followed by an introduction to risk parity portfolio optimization and an analysis of its theoretical properties.

We begin by discussing the measures of financial risk and reward for portfolios and their constituent assets. In particular, we discuss how to estimate these measures from raw financial data, and this is followed by the introduction of factor models. Factor models are linear regression models where the explanatory variables are the 'factors' that explain our sources of financial risk. In a financial context, these linear models allow us to attribute an asset's systematic risk to its exposure to a basket of financial factors. Factor models are relevant to academics and practitioners alike. Additionally, factor models allow us to estimate the measures of risk and reward that serve as our inputs for portfolio optimization.

We proceed to discuss MPT and the MVO problem introduced by Markowitz [71]. We present the optimization problem and discuss relevant literature on the subject of MVO, in particular its strengths and weaknesses.

In particular, some of the weaknesses of MVO motivate the introduction of the nominal risk parity problem. The mathematical derivation of the risk parity problem is presented in detail, beginning with the partitioning of the risk measure that allows us to quantify the risk contribution per asset towards the overall portfolio risk. This is followed by the presentation of the nominal risk parity problem.

This is followed by a discussion of the fundamental limitations of risk parity. The problem in

itself is non-convex, and typically relies on the imposition of a 'long-only' condition. The long-only condition means that we are not allowed to take short positions in our assets. Doing this allows us to reformulate the nominal risk parity problem into a convex problem. Specifically, we discuss two convex reformulations of the nominal problem.

Finally, we use the remainder of this chapter to discuss useful measures of financial performance and risk concentration that will be used in subsequent chapters of this thesis. These measures will allow us to evaluate the impact of our proposed advances to the risk parity framework.

1.1.2 Chapter 3: A robust framework for risk parity

This chapter presents a robust formulation of the traditional risk parity problem, introducing an uncertainty structure specifically tailored to capture the intricacies of risk parity. Generic mean–variance portfolios attempt to introduce robustness by computing the worst-case estimate of the risk measure, which does not necessarily align with the objective of risk parity: to attain full risk-based diversification.

Instead, our motivation is to shield our risk parity portfolio against estimation error in the asset risk contributions. The risk contributions quantify the proportion of risk that each asset contributes towards the overall portfolio risk level, and are more meaningful to the risk parity problem than measuring the overall portfolio risk. Thus, this chapter develops a novel robust risk parity framework that introduces robustness around both the overall portfolio risk and the assets' risk contributions.

The robust risk parity problem is highly tractable and is able to retain the same level of complexity as the nominal risk parity problem. We provide a general procedure to derive an uncertainty structure from data. This uncertainty structure is modelled as a perturbation of the nominal covariance estimate, allowing us to intuitively embed robustness during optimization. In addition, we provide a specific example on how to derive the uncertainty structure from a factor model of asset returns.

The corresponding numerical experiments show that the robust risk parity problem can yield a higher risk-adjusted rate of return when compared to the nominal problem while retaining a sufficiently risk-diverse structure.

The contribution of this chapter towards the advancement of risk parity is the introduction of a robust framework specifically designed to address uncertainty in the estimates of the asset risk

contributions. In other words, we target the uncertainty arising from the individual partitions of the risk measure, as opposed to traditional robust portfolio optimization problems that simply target the portfolio risk measure as a whole. The contributions and findings from this research were published in [30].

1.1.3 Chapter 4: Distributionally robust risk parity

This chapter introduces a distributionally robust formulation of the traditional risk parity problem. Unlike typical distributionally robust optimization (DRO) approaches to portfolio optimization that targets moment uncertainty, we focus on the estimation procedure that parametrizes these moments.

Traditionally, it is assumed that each scenario in a dataset is equally likely during parameter estimation (i.e., this implicitly assumes that there exists a discrete probability distribution that attaches a nominal probability to each scenario, and these probabilities are equal to each other).

Instead, we break away from this assumption and consider alternative instances of this discrete probability distribution (i.e., we do not assume that scenarios are equally likely, and instead consider an ambiguity set that allows us to find the most adversarial probability distribution that fits the data). In this sense, we emphasize the most robust estimate that we can derive from a raw data-driven approach.

Based on an investor's confidence level, we consider a bounded ambiguity set constrained by some predetermined maximum distance from the nominal 'equally-likely' distribution. This allows us to formulate the distributionally robust risk parity problem as a constrained convex–concave minimax problem. Our approach is also financially meaningful, and unlike Chapter 3, it aligns with more traditional views of robustness in portfolio optimization. This risk parity portfolio seeks to be fully risk diverse with respect to the worst-case estimate of the portfolio risk measure.

We propose a novel algorithmic approach to solve the constrained convex–concave minimax problem. This algorithm blends projected gradient ascent with sequential convex programming. By design, the algorithm is highly flexible and allows the investor to choose among several alternative statistical distance measures to define the ambiguity set, with the only condition that the statistical distance measure can be modelled as a convex function.

The resulting algorithm is highly tractable and scalable, and our numerical experiments suggest that a distributionally robust risk parity portfolio can yield a higher risk-adjusted rate

of return when compared against its nominal counterpart.

The contributions of this chapter towards the advancement of risk parity are twofold. First, we introduce a fully data-driven distributionally robust framework for risk parity. Second, we propose an algorithm that efficiently solves the resulting constrained minimax problem. The contributions and findings from this research have been compiled into a manuscript that has been submitted for publication.

1.1.4 Chapter 5: A regime-switching factor model for risk parity

This chapter takes a step back from optimization theory, and instead focuses on improving the estimated parameters used during optimization. The intent is to produce higher fidelity estimates that are aligned with the current market environment, under the assumption that markets are cyclical. Specifically, this chapter introduces a novel Markov regime-switching factor model to describe the cyclical nature of asset returns in modern financial markets.

Maintaining a linear factor model structure provides two advantages. First, calibrating the factor model is a straightforward process and can be done in closed-form through ordinary least squares regression, which aligns with current best practices in the financial industry. Second, the regime-switching factor model can be used to intuitively derive the regime-dependent systematic and idiosyncratic risk and return measures of our financial assets. In particular, we can use the regime-switching factor model to estimate the asset expected returns and covariance matrix, which are the main input parameters involved in traditional portfolio optimization.

By design, the resulting asset expected returns and covariance matrix are implicitly aligned with the current market regime. In turn, we can use these parameters as inputs during portfolio optimization to construct portfolios adapted to the current market regime. In alignment with this thesis, we use the estimated covariance matrix to construct regime-specific risk parity portfolios.

The simplicity of the factor model structure and subsequent derivation of input parameters for optimization translates into a framework that allows for the construction of large, realistic portfolios. In addition, given that the covariance matrix embeds all the pertinent regime-dependent information, the portfolio optimization step comes at no additional computational cost. In other words, the complexity of optimizing our regime-specific portfolios is the same as that of a nominal risk parity portfolio. Moreover, the viability of our regime-switching framework can be significantly improved by periodically rebalancing the portfolio, ensuring

proper alignment between the estimated parameters and the transient market regimes over time.

An out-of-sample computational experiment over a long investment horizon shows that the proposed regime-dependent risk parity portfolios are better aligned with the market environment, yielding a higher ex-post rate of return and lower volatility, even when compared against competing portfolios. Additionally, since the regime information is embedded within the input parameters, we can easily incorporate this framework into other problems. For example, our experiments demonstrate that we can combine this regime-switching framework with the robust risk parity problem from Chapter 3 to construct robust regime-switching risk parity portfolios.

The contribution of this chapter towards the advancement of risk parity stems from the ability to improve the quality of our estimated parameters, which directly impacts the subsequent optimization step. In essence, a better estimate of the risk measure translates to a better diversification of the portfolio risk among the assets. Although regime-switching models are not a new development, our proposed model greatly enhances their tractability and reconciles this model with the static nature of traditional asset management, where portfolios are constructed and held over some period of time. Additionally, this framework allows us to construct large portfolios in a computationally-efficient fashion. The contributions and findings from this research were published as two articles [26, 29].

1.1.5 Chapter 6: A generalized risk parity framework

This chapter proposes a new variant of portfolios that promote risk-based diversification while retaining the desirable properties of mean–variance analysis. Given the extended flexibility of this framework, we consider it a 'generalized' approach to risk parity.

The main objectives of this framework are threefold. First, we define an objective that seeks to maximize the portfolio expected return while minimizing the portfolio risk. Second, we relax the risk parity condition and instead bound the risk dispersion of the constituents within a predefined limit. This allows an investor to prescribe a desired risk dispersion range, yielding a portfolio with an optimal risk–return profile that is still well-diversified from a risk-based standpoint. Finally, this framework addresses a fundamental limitation of the nominal risk parity problem: our generalized risk parity problem allows for short sales.

The latter contribution is of particular importance because the allowance of short sales leads to a non-convex optimization problem. However, we propose an approach that relaxes

our generalized risk parity model into a convex semidefinite program. We proceed to tighten this relaxation sequentially through the alternating direction method of multipliers. This procedure iterates between the convex optimization problem and the non-convex problem with a rank constraint. In addition, we can exploit this structure to solve the non-convex problem analytically and efficiently during every iteration. However, given the non-convex nature of the original problem, we note that our algorithm does not guarantee global optimality.

We proceed to expand our framework and introduce two variants of the problem. The first variant addresses the impact of estimation error by introducing robustness through an ellipsoidal uncertainty structure placed around our estimated asset expected returns. We focus on the asset expected returns because their estimation errors are understood to have an impact an order of magnitude larger in portfolio optimization when compared to estimation errors in the covariance matrix [25].

The second variant addresses another fundamental limitation of risk parity. If we allow for short sales, we may have up to 2^{n-1} risk parity portfolios for a single estimate of the covariance matrix. These portfolios correspond to every possible long–short combination between the constituent assets, and they all satisfy the risk parity condition. Thus, from an optimization perspective, these portfolios are equivalent. However, in reality, these risk parity portfolios may have different levels of risk. In turn, this allows for the possibility of searching for the risk parity portfolio with the lowest risk. Our generalized framework is able to define the risk parity problem such that we search for a risk parity portfolio with the lowest risk.

Numerical results suggest that our algorithm converges to a higher quality optimal solution when compared to the competing non-convex problem, and can also yield a higher ex post risk-adjusted rate of return.

The contribution of this chapter towards the advancement of risk parity is the introduction of a generalized risk parity framework that blends the desirable properties of mean–variance analysis while still retaining the risk diversification properties of risk parity. Moreover, this chapter presents an algorithm to address the non-convexity of this generalized problem. The contributions and findings from this research were published in [28].

1.2 Notation

A general description of the notation follows. We denote a real space of dimension n by $\mathbb{R}^n$ and the corresponding non-negative orthant by $\mathbb{R}^n_+$. Moreover, symmetric matrices of dimension n with real-valued elements are denoted by $\mathbb{S}^n$, while the subset of positive semi-definite (PSD) matrices is denoted by $\mathbb{S}^n_+$.

Bold lowercase letters refer to vectors, while bold uppercase letters refer to matrices. A bold number refers to a vector of appropriate dimension (e.g., $\mathbf{1}$ refers to a vector where each element is equal to one). This 'appropriate dimension' is based on the context of some reference vector. For example, if we have $\mathbf{1}^\top \boldsymbol{z} = b$, then the vector $\mathbf{1}$ has the same dimension as the vector $\boldsymbol{z}$. Moreover, as shown in the previous example, the superscript '$\top$' indicates the transpose of a vector or matrix. If we need to reference some specific element i within a vector, we do this by using the subscript i. For example, if we wish to reference the i^{th} element of vector $\boldsymbol{z}$, we denote this as z_i. In certain circumstances, if we define a vector as the product between a matrix and a vector, and we wish to reference its i^{th} element; then, we use square brackets and the subscript i to do so. For example, if we have some matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and some vector $\boldsymbol{z} \in \mathbb{R}^n$, and the resulting product between them is the vector $\boldsymbol{Az} \in \mathbb{R}^m$; then, its i^{th} element is denoted as $[\boldsymbol{Az}]_i$. Additionally, when relational operators are used in the context of vectors, they are applied in an element-wise fashion (e.g., $\boldsymbol{z} \geq \mathbf{0}$ implies that every element of the vector $\boldsymbol{z}$ is greater than or equal to zero). Element-wise operations, also known as Hadamard operations, are denoted by the symbol '$\circ$'. For example, the element-wise product between two matrices of the same dimension $\boldsymbol{A}$ and $\boldsymbol{B}$ is shown as $[\boldsymbol{A} \circ \boldsymbol{B}]_{ij} = A_{ij}B_{ij} \ \forall \ i, j$. Other element-wise operations, such as the element-wise square of each element in a matrix or vector, is shown as $[\boldsymbol{A}^{\circ 2}]_{ij} = A^2_{ij} \ \forall \ i, j$.

The ℓ_p-norm of an arbitrary vector $\boldsymbol{v} \in \mathbb{R}^n$ is denoted by $\|\cdot\|_p$, where $\|\boldsymbol{v}\|_p \triangleq \left(\sum_{i=1}^n |v_i|^p \right)^{1/p}$. In the case of matrix norms, we specifically refer to the Frobenius norm, which is a natural extension of the ℓ_2-norm for matrices. Specifically, the Frobenius norm of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is defined as $\|\boldsymbol{A}\|_F \triangleq \left(\sum_{i=1}^m \sum_{j=1}^n |A_{ij}|^2 \right)^{1/2}$.

As exemplified in the previous paragraph, we also use the notation '$\triangleq$' to mean 'equal by definition'. Some definitions will be expressed as functions of the form $\boldsymbol{a}(\boldsymbol{b}, \boldsymbol{c}) \in \mathbb{R}^n$, where, in this example, $\boldsymbol{b} \in \mathbb{R}^l$ and $\boldsymbol{c} \in \mathbb{R}^m$ are arguments and $\boldsymbol{a} : \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}^n$. Note that neither the arguments nor the output need to be vectors, and this will be indicated by their bold font type

when appropriate. Moreover, functions may serve to denote a specific parameter, such as the previous example that defined some parameter $\boldsymbol{a}(\cdot) \in \mathbb{R}^n$.

Sets are denoted by uppercase letters using a calligraphic font (e.g., $\mathcal{X}$). The expectation operator is denoted as $\mathbb{E}[\cdot]$. In the case where multiple variables are specified within the expectation operator, we indicate which variable the operator is acting upon with a corresponding subscript (e.g., $\mathbb{E}_{\boldsymbol{\zeta}}[f(\boldsymbol{z}, \boldsymbol{\zeta})]$ for some variables $\boldsymbol{z}$ and $\boldsymbol{\zeta}$ and some function $f(\boldsymbol{z}, \boldsymbol{\zeta})$).

Finally, we denote the optimal solution to an optimization problem with an asterisk superscript (e.g., $\boldsymbol{z}^*$ is the optimal solution corresponding to some decision variable $\boldsymbol{z}$).

Chapter 2

Modern portfolio theory and risk parity

This chapter briefly discusses MPT, portfolio optimization, and risk parity. Specifically, we focus on the estimation procedure of the parameters that quantify the two main features of MPT: risk and reward. Any practical implementation of MPT requires the estimation of risk and reward measures from data. As we will see, we can produce these estimates directly from data, or through the application of factor models.

We then proceed to introduce risk parity. This includes a careful derivation of the nominal risk parity problem. As we will see, we can formulate the risk parity problem as multiple alternative optimization problems, all of which achieve the same objective: a portfolio where the asset risk contributions are equalized. In addition, we discuss the theoretical limitations and deficiencies of risk parity that motivated this thesis.

Finally, we discuss some general measures of financial performance and risk concentration. These are assessment tools to evaluate the performance of the frameworks we will propose in this thesis. Thus, the measures of financial performance and risk concentration will be used consistently in subsequent chapters for their respective numerical experiments.

2.1 Portfolio optimization and factor models

2.1.1 Risk and reward

We define a portfolio as a collection of n assets in which we invest our available wealth. Mathematically, our portfolio is represented by a vector of asset weights $\boldsymbol{x} \in \mathbb{R}^n$, where x_i represents the proportion of wealth invested in asset i. From an asset management perspective, $\boldsymbol{x}$ is our vector of decision variables that represents our asset allocation strategy. A negative asset weight, $x_i < 0$, implies we have a short position on that asset, i.e., we will benefit if the asset's value falls.

The 'assets' are the financial instruments which we can buy and sell in the market in order to construct a portfolio. Over time, the price of an asset changes, and a relative change in price over a specific time period is referred to as the rate of return (or simply the 'return'). The asset returns are modelled as random variables which we define as the vector $\boldsymbol{\xi} \in \mathbb{R}^n$. In traditional asset management,[1] we typically measure the return over discrete time steps of daily, weekly, or monthly length. The random returns are governed by some joint probability distribution with first and second moments defined as the asset expected returns $\boldsymbol{\mu} \in \mathbb{R}^n$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{S}_+^n$, respectively. MPT [71] parametrizes the joint probability distribution of asset returns using the first two moments. Thus, this implicitly assumes that the asset returns are normally distributed, i.e., $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

At the asset-level, the financial reward is measured by $\boldsymbol{\mu}$ and the financial risk by $\boldsymbol{\Sigma}$. In particular, the asset return probability distribution, parametrized by its first two moments, is assumed to be latent and the moments must be estimated from data. In turn, this means that our estimates of the expected returns and covariance matrix are prone to suffer from estimation error [14, 20, 73].

At the portfolio-level, return is calculated as a weighted linear combination of the returns of the n constituent assets. In other words, the portfolio random return is $\pi = \boldsymbol{\xi}^\top \boldsymbol{x} \in \mathbb{R}$. The corresponding measures of portfolio reward and risk are

$$\mu_\pi \triangleq \boldsymbol{\mu}^\top \boldsymbol{x}, \tag{2.1}$$

$$\sigma_\pi^2 \triangleq \boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}, \tag{2.2}$$

[1] We purposely exclude high frequency trading from this thesis.

where the portfolio expected return is $\mu_\pi \in \mathbb{R}$, while the portfolio variance is $\sigma_\pi^2 \in \mathbb{R}_+$.

The first two moments of our asset returns, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, are typically estimated from data (e.g., historical scenarios of asset returns). Assume we have an asset return dataset consisting of T discrete scenarios for n assets (i.e., we have a dataset $\hat{\boldsymbol{\xi}} \in \mathbb{R}^{n \times T}$ consisting of scenarios, and these scenarios suffice to represent the possible outcomes of the random variable $\boldsymbol{\xi}$). Furthermore, if we assume each scenario is equally likely, we can statistically derive the estimates of the first two moments. Let $\hat{\boldsymbol{\xi}}^t \in \mathbb{R}^n$ be the t^{th} scenario of the dataset $\hat{\boldsymbol{\xi}}$. We can derive the estimates of the asset expected returns $\hat{\boldsymbol{\mu}} \in \mathbb{R}^n$ and covariance matrix $\hat{\boldsymbol{\Sigma}} \in \mathbb{S}_+^n$ as follows,

$$\hat{\boldsymbol{\mu}} \triangleq \frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{\xi}}^t, \tag{2.3}$$

$$\hat{\boldsymbol{\Sigma}} \triangleq \frac{1}{T-1} \sum_{t=1}^{T} (\hat{\boldsymbol{\xi}}^t - \hat{\boldsymbol{\mu}})(\hat{\boldsymbol{\xi}}^t - \hat{\boldsymbol{\mu}})^\top. \tag{2.4}$$

The definition of the covariance matrix in (2.4) corresponds to its unbiased estimator. In addition, we note that $\hat{\boldsymbol{\Sigma}}$ arises from the weighted sum of T rank-1 symmetric matrices, meaning $\hat{\boldsymbol{\Sigma}}$ is guaranteed to be a PSD matrix.

The estimates of the portfolio expected return and variance follow the same logic as (2.1) and (2.2), except we replace the latent parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with their corresponding estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$,

$$\hat{\mu}_\pi \triangleq \hat{\boldsymbol{\mu}}^\top \boldsymbol{x}, \tag{2.5}$$

$$\hat{\sigma}_\pi^2 \triangleq \boldsymbol{x}^\top \hat{\boldsymbol{\Sigma}} \boldsymbol{x}. \tag{2.6}$$

2.1.2 Factor models

Factor models attempt to explain the behaviour of a random variable either through a single factor, such as the capital asset pricing model [68, 77, 93], or through a combination of multiple factors, such as the Fama–French three-factor model [39]. Factor models are popular in finance due to the economic relevance of the factors, as well as their ability to explain and quantify different sources of an asset's systematic risk. One important application of these models is to estimate the asset expected returns and covariance matrix, where the inherent properties of the

factors systematically explain the covariance between the assets.[2]

Suppose that the random asset returns $\boldsymbol{\xi}$ can be explained through a linear combination of m factors. It follows that the random asset returns can be described as

$$\boldsymbol{\xi} = \boldsymbol{\mu} + \boldsymbol{V}^\top \boldsymbol{\phi} + \boldsymbol{\epsilon}, \tag{2.7}$$

where $\boldsymbol{\phi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{F}) \in \mathbb{R}^m$ is the vector of centred factor returns, $\boldsymbol{V} \in \mathbb{R}^{m \times n}$ is the matrix of factor loadings, and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}) \in \mathbb{R}^n$ are the residual returns. $\boldsymbol{F} \in \mathbb{S}_+^m$ and $\boldsymbol{D} \in \mathbb{S}_+^n$ denote the factor covariance matrix and the diagonal matrix of residual variance, respectively. We choose to use *centred* factor returns so that the factors serve only to explain the variability of the asset returns.

Therefore, if we take the expectation of (2.7), we can see that the intercept of regression corresponds to the asset expected returns, i.e.,

$$\mathbb{E}[\boldsymbol{\xi}] = \mathbb{E}\big[\boldsymbol{\mu} + \boldsymbol{V}^\top \boldsymbol{\phi} + \boldsymbol{\epsilon}\big] = \boldsymbol{\mu}.$$

Moreover, if we measure the variance and covariance terms from (2.7), we find that the asset covariance matrix can be calculated as follows,

$$\boldsymbol{\Sigma} = \boldsymbol{V}^\top \boldsymbol{F} \boldsymbol{V} + \boldsymbol{D}.$$

The factor model in (2.7) assumes that the residual returns are independent of one another, i.e. $\operatorname{cov}(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$. Additionally, the factor model assumes that the residual returns are independent of the factor returns, i.e. $\operatorname{cov}(\epsilon_i, \phi_j) = 0 \; \forall \; i, \; j$. However, the model does not assume that the factors are independent of one another, i.e. the factor covariance matrix, $\boldsymbol{F}$, is not required to be a diagonal matrix.

In practice, $\boldsymbol{\mu}$, $\boldsymbol{V}$ and $\boldsymbol{D}$ are estimated through ordinary least squares regression, while the factor covariance matrix $\boldsymbol{F}$ is estimated directly from raw factor returns data. The estimation of the factor covariance matrix follows the same process as the one shown in (2.4), but we explain it again for clarity. Assume our factor return data consist of T discrete scenarios for m factors, i.e., we have $\hat{\boldsymbol{\phi}} \in \mathbb{R}^{m \times T}$ scenarios where $\hat{\boldsymbol{\phi}}^t \in \mathbb{R}^m$ is the t^{th} scenario. Since the factors

[2]Measuring the covariance between the assets in a portfolio is paramount to the proper estimation of a portfolio's risk level.

are centred, the unbiased estimator of the factor covariance matrix is

$$\hat{\boldsymbol{F}} \triangleq \frac{1}{T-1}\sum_{t=1}^{T}\hat{\boldsymbol{\phi}}^t(\hat{\boldsymbol{\phi}}^t)^\top. \tag{2.8}$$

In addition, the estimates $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{V}}$ and $\hat{\boldsymbol{D}}$ are calculated as follows. First, we add a column of ones to the factor data to account for the intercept (i.e., the estimated asset expected returns), $\hat{\boldsymbol{A}} \triangleq [\mathbf{1}\ \hat{\boldsymbol{\phi}}^\top] \in \mathbb{R}^{T\times(m+1)}$. Next, using the asset data, we find the least squares estimates of the expected returns and factor loadings as follows

$$\hat{\boldsymbol{B}} \triangleq \left[\hat{\boldsymbol{\mu}}\ \hat{\boldsymbol{V}}^\top\right]^\top = (\hat{\boldsymbol{A}}^\top\hat{\boldsymbol{A}})^{-1}\hat{\boldsymbol{A}}^\top\hat{\boldsymbol{\xi}}^\top. \tag{2.9}$$

Accordingly, the matrix of scenario residuals, $\hat{\boldsymbol{\epsilon}} \in \mathbb{R}^{n\times T}$, is calculated as follows

$$\hat{\boldsymbol{\epsilon}} \triangleq \hat{\boldsymbol{\xi}} - \hat{\boldsymbol{B}}^\top\hat{\boldsymbol{A}}^\top.$$

The asset residual variances are estimated as the sum of squared residuals. In other words, the vector of the estimated residual variances, $\hat{\boldsymbol{s}}^2 \in \mathbb{R}^n$, is

$$\hat{\boldsymbol{s}}^2 \triangleq \frac{\hat{\boldsymbol{\epsilon}}^{\circ 2}\mathbf{1}}{T-m-1}. \tag{2.10}$$

Thus, the matrix $\hat{\boldsymbol{D}}$ is the diagonal matrix arising from the vector $\hat{\boldsymbol{s}}^2$, i.e., $\hat{D}_{ii} = \hat{s}_i^2$ for $i = 1,\ldots,n$ and $\hat{D}_{ii} = 0$ for $i \neq j$. Finally, the estimate of the asset covariance matrix stemming from a factor model structure is the following,

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{V}}^\top\hat{\boldsymbol{F}}\hat{\boldsymbol{V}} + \hat{\boldsymbol{D}}. \tag{2.11}$$

We can also use the residual variances to calculate the standard error of our estimated factor loadings. As we will see in Chapter 3, the standard errors of the factor loadings will allow us to naturally derive an uncertainty structure from a factor model, which in turn will allow us to introduce robustness into the optimization problem.

For now, we describe how to derive the standard errors of the factor loadings. The estimated factor loading corresponding to factor i and asset j is denoted as $\hat{V}_{ij}$. The corresponding

standard error, denoted as $\mathrm{SE}(\hat{V}_{ij})$, can be calculated as follows

$$\mathrm{SE}(\hat{V}_{ij}) = \sqrt{\hat{s}_j^2 \left[(\hat{\boldsymbol{\phi}}\hat{\boldsymbol{\phi}}^\top)^{-1}\right]_{ii}}. \tag{2.12}$$

As we proceed through this thesis, we will clearly indicate whether the estimated asset parameters $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are estimated directly from raw data as shown in (2.3) and (2.4), respectively, or through a factor model as shown in (2.9) and (2.11), respectively. In general, we will resort to a factor model for estimation only when it is useful for our model development. Specifically, factor models will be used in Chapters 3 and 5.

2.1.3 Mean–Variance Optimization

The MVO problem introduced by Markowitz [71] analyses the trade-off between risk and return to construct optimal portfolios. These portfolios provide an optimal balance between risk and return relative to an investor's risk appetite. However, while MVO has received widespread attention as a quantitative investment tool in asset management, it has also been criticized for its susceptibility to estimation error in its input parameters, namely the estimated expected returns and covariance matrix. The discrepancies between the estimated parameters and their ex-post realizations can severely hinder the performance of an 'optimal' portfolio, where optimality is determined solely from these estimates.

The sensitivity of MVO to its inputs has been extensively explored in the literature, with important examples shown in [14, 73, 20]. Indeed, when estimated parameters are noisy, MVO has sometimes been referred to as 'error maximization', yielding poor out-of-sample portfolio performance. However, the effect of estimation errors from the expected returns and the covariance matrix are not the same. A widely referenced conclusion found in [25] argues that errors in expected returns can have an impact an order of magnitude larger than errors in the covariance matrix.

Traditionally, estimation errors have been addressed by either of two methods. The most obvious remediation is to improve the quality of our estimated parameters. This is sometimes achieved by using factor models as shown in Section 2.1.2, or through Bayesian shrinkage [59, 65]. Alternatively, we can apply robust optimization techniques. Robust portfolio optimization methods mitigate the impact of uncertainty through the use of either box constraints or ellipsoidal constraints [69, 96]. A different approach incorporates the uncertainty of the

choice of probability distribution governing the asset returns, formulating a DRO problem [32]. We can also use robust optimization in tandem with factor models, where the uncertainty is derived from the estimation errors in the regression coefficients of the factor model [49].

We proceed by introducing the following version of the MVO problem, where the investor seeks to minimize the estimated portfolio variance, $\hat{\sigma}_\pi^2$, while simultaneously maximizing the estimated portfolio expected return, $\hat{\mu}_\pi$. The MVO problem is

$$\min_{\boldsymbol{x}} \quad \boldsymbol{x}^\top \hat{\boldsymbol{\Sigma}} \boldsymbol{x} - \lambda \hat{\boldsymbol{\mu}}^\top \boldsymbol{x} \tag{2.13a}$$

$$\text{s.t.} \quad \mathbf{1}^\top \boldsymbol{x} = 1, \tag{2.13b}$$

where the equality constraint (2.13b) ensures that the entirety of our available budget is invested in the assets. Colloquially, this constraint is often referred to as the 'budget constraint'. In the problem above, an investor's risk–return profile is defined by a predefined trade-off coefficient $\lambda \in \mathbb{R}_+$. Since $\hat{\boldsymbol{\Sigma}} \in \mathbb{S}_+^n$, the MVO problem is quadratic and convex. Additional constraints can be added to the MVO problem in (2.13) to satisfy the investor's design specifications. Provided these additional constraints define a convex feasible set, then the MVO problem will remain convex. In general, we will refer to the set defined by the constraints on $\boldsymbol{x}$ as the set of admissible portfolios.

The MVO problem can be modified to explicitly target a desired attribute. For example, we can remove the portfolio expected return from the objective function and instead impose a minimum target return constraint. Thus, this version of MVO would find the portfolio with the lowest variance that satisfies the desired target return. Conversely, we could maximize the portfolio expected return while constraining the portfolio variance.

2.2 Risk parity

The American investment management firm Bridgewater Associates is often cited as the first to pioneer the risk parity portfolio in 1996, calling it an 'all-weather' fund. The name stems from its design, which was meant to be fully diversified from a risk perspective in order to avoid the negative impacts associated with risk concentration. Qian [83] was the first to refer to this particular asset allocation strategy as 'risk parity', and this terminology has since been widely adopted.

Risk parity aims to construct a portfolio where every asset has the same level of risk contribution towards the overall portfolio risk. Conceptually, it is similar to the popular '$1/n$' portfolio, except that instead of being perfectly diversified from a wealth perspective, risk parity aims to be fully diversified from a risk perspective. By design, we are solely concerned with the portfolio risk, which in our case is measured as a function of the asset covariance matrix $\mathbf{\Sigma}$. Thus, the risk parity portfolio optimization problem does not require a reward measure, i.e., we do not require the estimated asset expected returns as an input.

As shown in [70], we can measure the individual risk contribution of each asset by applying Euler's homogeneous function theorem to partition the portfolio risk measure. For now, assume we have perfect knowledge of the distribution of the random asset returns (i.e., we have knowledge of the true covariance matrix $\mathbf{\Sigma}$). The portfolio standard deviation can be found by taking the square root of Equation (2.2). Applying Euler's theorem, the portfolio standard deviation can be partitioned as follows

$$\sigma_\pi = \sqrt{\boldsymbol{x}^\top \mathbf{\Sigma} \boldsymbol{x}} = \sum_{i=1}^{n} x_i \frac{\partial \sigma_\pi}{\partial x_i} = \sum_{i=1}^{n} x_i \frac{[\mathbf{\Sigma}\boldsymbol{x}]_i}{\sqrt{\boldsymbol{x}^\top \mathbf{\Sigma} \boldsymbol{x}}}. \tag{2.14}$$

The latter part of (2.14) shows the partitions of the portfolio standard deviation for each asset $i = 1, \ldots, n$. Note that the denominator in this expression is consistent for all partitions. In addition, the denominator is equal to the portfolio standard deviation. Thus, as shown in [28], we can rearrange this expression to partition the portfolio variance instead. We can express the portfolio variance as the sum of n components,

$$\sigma_\pi^2 = \boldsymbol{x}^\top \mathbf{\Sigma} \boldsymbol{x} = \sum_{i=1}^{n} x_i [\mathbf{\Sigma}\boldsymbol{x}]_i = \sum_{i=1}^{n} r_i, \tag{2.15}$$

where $r_i \triangleq x_i[\mathbf{\Sigma}\boldsymbol{x}]_i$ is the individual risk contribution of asset i.

Now that we are able to measure the individual risk contributions, we can formulate an optimization problem to construct a risk parity portfolio such that $r_i = r_j \; \forall \; i, j$. As shown in

[70], this can be achieved through a least squares approach,

$$\min_{\boldsymbol{x}} \quad \sum_{i=1}^{n}\sum_{j=1}^{n}\left(x_i[\boldsymbol{\Sigma x}]_i - x_j[\boldsymbol{\Sigma x}]_j\right)^2 \tag{2.16a}$$

$$\text{s.t.} \quad \mathbf{1}^T\boldsymbol{x} = 1, \tag{2.16b}$$

$$\boldsymbol{x} \geq \mathbf{0}. \tag{2.16c}$$

The objective function in (2.16a) aims to minimize the difference between the asset risk contributions. The budget constraint in (2.16b) ensures that all available wealth is invested. Finally, the non-negativity constraint in (2.16c) prohibits short selling. Both constraints are affine and create a convex feasible set.

However, partitioning the portfolio variance into the individual risk contributions leads to a non-convex objective function. The issue of non-convexity is prevalent in any risk parity formulation where the individual risk contributions are employed explicitly. In fact, we are forced to impose the non-negativity constraint in (2.16c) to guarantee the uniqueness of the optimal solution to (2.16). In other words, this non-negativity constraint is not financially motivated, but rather it stems from a fundamental limitation of the nominal risk parity problem.

2.2.1 Non-convexity of risk parity

The non-convexity of (2.16) becomes apparent if we inspect the individual risk contributions, r_i. We begin by recasting the asset risk contribution r_i in standard quadratic notation

$$r_i = x_i[\boldsymbol{\Sigma x}]_i = \boldsymbol{x}^\top \boldsymbol{R}^i \boldsymbol{x},$$

where $\boldsymbol{R}^i \in \mathbb{S}^n$ serve to capture the individual risk contribution of asset i. The symmetric matrices $\boldsymbol{R}^i$ are composed of the superposition of the i^{th} row and i^{th} column of the covariance matrix $\boldsymbol{\Sigma}$ multiplied by one half, with all other elements in the matrix equal to zero, i.e.,

$$\boldsymbol{R}^i = \frac{1}{2}(\boldsymbol{e}_i\boldsymbol{e}_i^\top\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\boldsymbol{e}_i\boldsymbol{e}_i^\top), \tag{2.17}$$

where $\boldsymbol{e}_i \in \mathbb{R}^n$ denotes the i^{th} column of the identity matrix. Thus, by design, $\boldsymbol{\Sigma} = \sum_{i=1}^{n} \boldsymbol{R}^i$. Inspecting the sparse matrices $\boldsymbol{R}^i$ reveals these are indefinite, each having a single positive eigenvalue and a single negative eigenvalue, with all other eigenvalues equal to zero. Thus, any

optimization problem that employs the risk contributions r_i will be non-convex.

Nevertheless, the problem in (2.16) is numerically efficient and is able to consistently attain a unique global solution within the feasible set of long-only portfolios [5, 70]. As we previously stated, imposing a long-only constraint is not financially motivated, but rather it is a necessity arising from the non-convexity of the optimization problem itself. A unique global solution, where the asset risk contributions are equalized, is only guaranteed if we shrink the feasible set to require non-negative values of $\boldsymbol{x}$.

To elaborate on this subject, we refer to the convex reformulation of the risk parity problem proposed by Bai et al. [5],

$$\min_{\boldsymbol{y}} \quad \frac{1}{2}\boldsymbol{y}^\top \boldsymbol{\Sigma} \boldsymbol{y} - \kappa \sum_{i=1}^{n} \ln(\beta_i y_i) \tag{2.18a}$$

$$\text{s.t.} \quad \beta_i y_i > 0, \quad i = 1, \ldots, n, \tag{2.18b}$$

where $\beta_i \in \{-1, 1\}\ \forall\ i$ and $\kappa > 0$ is an arbitrary positive constant. At optimality, we find that

$$[\boldsymbol{\Sigma} \boldsymbol{y}]_i = \frac{\kappa}{y_i}\ \ \forall\ i \quad \Rightarrow \quad y_i [\boldsymbol{\Sigma} \boldsymbol{y}]_i = \kappa\ \ \forall\ i,$$

retrieving a solution where all individual risk contributions are equal to a constant, and, therefore, to each other. The variable $\boldsymbol{y} \in \mathbb{R}^n$ is a placeholder for the vector of asset weights $\boldsymbol{x}$, and stems from the lack of a budget constraint in (2.18). Discarding the budget constraint allows the decision variable $\boldsymbol{y}$ to match the risk contributions to κ. However, there is no guarantee that the sum of weights will be equal to one. Instead, the optimal risk parity solution can be recovered as follows, $x_i^* = y_i / \sum_{i=1}^{n} y_i$. Since $\boldsymbol{\Sigma}$ is PSD and the sum of logarithms is strictly concave, the objective function (2.18a) is strictly convex. In turn, this means any optimal solution will be a unique global solution.

This is particularly relevant to us because it means that up to 2^{n-1} risk parity solutions may exist. These risk parity solutions correspond to every possible combination of $\beta_i y_i > 0$ for each $\beta_i \in \{-1, 1\}$, with each solution corresponding to every possible long–short combination between the assets.

From an optimization perspective, all of these optimal solutions meet the risk parity condition of equalizing the asset risk contributions. However, in practice, these solutions may lead to very different portfolios. These portfolios will have different levels of risk and return. In turn,

this highlights two fundamental deficiencies of the nominal risk parity framework. First, risk parity seeks only to equalize individual risk contributions, disregarding any impact this may have on the overall portfolio risk. Other things equal, an investor will always prefer to hold a portfolio with the lowest possible risk. The second deficiency is that any short positions we wish to hold must be predetermined before optimization takes place. Indeed, should we wish to find the most desirable risk parity solution, we will first need to find the subset of 2^{n-1} optimal solutions before selecting the most desirable solution based on a secondary level of criteria (e.g., lowest portfolio variance, highest rate of return).

2.2.2 Long-only convex risk parity problems

The general consensus in risk parity portfolio optimization is to restrict ourselves to the positive orthant in $\mathbb{R}^n$ [5, 70]. In other words, we can guarantee the uniqueness of the risk parity solution if we only allow long-only portfolios.

Restricting the feasible set to $\mathbb{R}^n_+$ allows us to formulate the risk parity problem as a convex optimization problem. For example, consider the optimization problem in (2.18). If we let $\beta_i = 1 \; \forall \; i$, then we have the following convex risk parity problem,

$$\boldsymbol{y}^* = \underset{\boldsymbol{y} \in \mathbb{R}^n_+}{\operatorname{argmin}} \quad \frac{1}{2}\boldsymbol{y}^\top \boldsymbol{\Sigma} \boldsymbol{y} - \kappa \sum_{i=1}^{n} \ln(y_i), \tag{2.19a}$$

$$\boldsymbol{x}^* = \frac{\boldsymbol{y}^*}{\sum_{i=1}^{n} y_i^*}. \tag{2.19b}$$

As we previously noted, the covariance matrix $\boldsymbol{\Sigma}$ is PSD and the sum of logarithms is strictly concave, meaning this version of the risk parity problem is strictly convex.

Alternatively, the risk parity problem can also be reformulated as a second-order cone program (SOCP). As shown by Mausser and Romanko [72], the SOCP version of the risk parity

problem is

$$\min_{\boldsymbol{x},\boldsymbol{z},u,v} \quad u - v \tag{2.20a}$$

$$\text{s.t.} \qquad \mathbf{1}^\top \boldsymbol{x} = 1, \tag{2.20b}$$

$$[\boldsymbol{\Sigma}\boldsymbol{x}]_i = z_i, \qquad i = 1, \ldots, n, \tag{2.20c}$$

$$\left\| \begin{bmatrix} 2v \\ x_i - z_i \end{bmatrix} \right\|_2 \leq x_i + z_i, \quad i = 1, \ldots, n, \tag{2.20d}$$

$$\left\| \boldsymbol{\Sigma}^{1/2} \boldsymbol{x} \right\|_2 \leq \sqrt{n}\, u, \tag{2.20e}$$

$$\boldsymbol{x}, \boldsymbol{z} \geq \mathbf{0}, \tag{2.20f}$$

$$u, v \geq 0, \tag{2.20g}$$

where $u,\ v \in \mathbb{R}_+$, and $\boldsymbol{z} \in \mathbb{R}^n_+$ are auxiliary variables. Since $x_i, z_i \geq 0$, constraint (2.20d) is equivalent to the hyperbolic constraint $x_i z_i \geq v^2$. Moreover, the objective function in (2.20a) is equivalent to

$$\sqrt{\frac{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}}{n}} - \sqrt{\min_{1 \leq i \leq n} \{x_i [\boldsymbol{\Sigma}\boldsymbol{x}]_i\}},$$

where the square roots are simply a construct of the SOCP reformulation. By design, the objective function is zero at optimality and otherwise positive. Thus, optimality is attained when the smallest risk contribution is equal to the average risk contribution of the portfolio. As we will see in Chapter 3, the SOCP formulation will allow us to design a highly tractable robust counterpart to the nominal risk parity problem.

2.3 Measures of financial performance and risk concentration

Here we define a few additional measures of financial performance and risk concentration. In particular, we discuss *ex post* measures of financial performance, i.e., measures based on out-of-sample observations instead of forecasts. Conversely, we measure risk concentration in an *ex ante* basis, i.e., these are based on forecasts rather than realizations. These measures will be useful in subsequent chapters to evaluate whether our proposed frameworks are, in fact, improving the nominal risk parity problem.

We begin by discussing the measures of return and risk. We previously defined the ex ante measures of portfolio return and risk. Specifically, we defined the portfolio expected return $\hat{\mu}_\pi$

in (2.5) and variance $\hat{\sigma}_\pi^2$ in (2.6).

However, the ex post measures must be computed using the time series of realized portfolio values over discrete time, which we refer to as the portfolio wealth evolution. To compute the portfolio wealth evolution, we must be conscious that both the total portfolio wealth and the asset weights change over time as the value of the underlying assets evolves.

We can calculate the realized portfolio returns per time step using the portfolio wealth evolution. In turn, we can use the time series of realized portfolio returns to calculate the *excess* returns by subtracting the risk-free rate over each time step. For example, if we use U.S. stocks for our experiments, then the risk-free rate can be approximated using U.S. Treasury bonds or bills.

Next, we can use the time series of realized portfolio excess returns to calculate the corresponding geometric mean and variance. Furthermore, we can use the variance to derive the portfolio *volatility*, which is simply the standard deviation of the portfolio wealth evolution. Finally, the geometric mean and volatility can be annualized by scaling them relative to the number of discrete time steps per year, yielding the portfolio's annualized excess return and annualized volatility.

The Sharpe ratio is a very popular measure of risk-adjusted return [94] often used to compare the financial performance between different portfolios. For consistency, our out-of-sample numerical experiments will use the following definition of the Sharpe ratio: the ratio between a portfolio's annualized excess return and its annualized volatility.

Our out-of-sample numerical experiments will rely on repeatedly rebalancing and re-optimizing our portfolios periodically over an investment horizon. For example if our out-of-sample horizon spans the 2000–2016 time period (17 years), and we rebalance our portfolios every six months, then we will re-optimize our portfolio a total of 34 times over the entire investment horizon.

We measure the average turnover rate of our optimal portfolio period-over-period to assess its *stability*, i.e., how much do our asset weights change every time we rebalance our portfolio. Stability is a desirable property because it reduces the transaction costs associated with asset management. These transaction costs include everything from market friction and brokerage fees to management overhead. We define the turnover rate as the absolute change in asset weights at the instant where rebalancing takes place summed over all assets in the portfolio. Following our previous example, the average turnover rate would correspond to the average over the 34 rebalancing periods.

Finally, we discuss some ex ante measures of risk concentration. Our proposed advancements to risk parity may force us into portfolios that do not satisfy the nominal definition of risk parity. Thus, it will be important to measure the risk concentration of our portfolios. We use the following three measures of ex ante risk concentration:

- Coefficient of Variation (CV): the CV is calculated by dividing the standard deviation of the asset risk contributions by their average

$$\text{CV} = \frac{\text{SD}\big(\boldsymbol{x} \circ (\hat{\boldsymbol{\Sigma}}\boldsymbol{x})\big)}{\frac{1}{n}\boldsymbol{x}^\top \hat{\boldsymbol{\Sigma}}\boldsymbol{x}}, \tag{2.21}$$

  where the operator 'SD$(\cdot)$' computes the standard deviation of the input vector. If we have perfect risk-based diversification, then the portfolio CV is equal to zero.

- Highest Risk Contribution (HRC): as its name suggests, the HRC is defined as the ratio of the highest individual risk contribution over the total portfolio variance,

$$\text{HRC} = \max_{i \in \{1,\ldots,n\}} \frac{x_i(\hat{\boldsymbol{\Sigma}}\boldsymbol{x})_i}{\boldsymbol{x}^\top \hat{\boldsymbol{\Sigma}}\boldsymbol{x}}. \tag{2.22}$$

  The value of the HRC ranges between $1/n$ for the risk parity portfolio and 1 for a fully concentrated portfolio.

- Herfindahl index: the H-index is defined as follows

$$\text{H-index} = \sum_{i=1}^{n} \left(\frac{x_i(\hat{\boldsymbol{\Sigma}}\boldsymbol{x})_i}{\boldsymbol{x}^\top \hat{\boldsymbol{\Sigma}}\boldsymbol{x}} \right)^2. \tag{2.23}$$

  The value of the H-index also ranges between $1/n$ for the risk parity portfolio and 1 for a fully concentrated portfolio.

We reiterate that these are ex ante measures of risk concentration, and will allow us to measure how diverse (or, conversely, how concentrated) is the risk of our portfolios. For most cases, we will defer to the CV as our sole risk concentration measure. The only exception is Chapter 3, which will use all three measures.

Chapter 3

A robust framework for risk parity

Regardless of the statistical procedure used to estimate the asset covariance matrix, any estimate derived from data will always have some degree of estimation error. It follows that the uncertainty surrounding our estimate can usually be statistically quantified, and this information can be exploited during the subsequent optimization step. Our proposed robust risk parity framework accepts the estimation error as a secondary input parameter. This robust framework is specifically tailored towards risk parity. From a risk parity perspective, our objective is to equalize the asset risk contributions. Thus, we are not only concerned with the overall portfolio risk, but we are also concerned with our measure of individual risk contributions. As such, our contribution is to design an optimization problem that uses the estimation error of the risk measure to introduce robustness around the assets' marginal risk contributions.

The robust risk parity problem presented in this chapter is able to accommodate any generic estimation procedure of the covariance matrix, provided this allows us to specify both the nominal estimate of the covariance and its corresponding uncertainty set. After we estimate the covariance matrix and quantify its estimation error, we then develop a robust optimization framework based on the methodology described in [13]. The nominal risk parity problem is formulated as a SOCP. Thus, as outlined in [13], we are able to formulate a robust optimization framework that retains the same level of complexity as the original problem (i.e., the robust counterpart is also a SOCP). Computational results show this robust problem is able to improve both portfolio returns and risk-adjusted returns relative to the nominal problem.

Unlike traditional MVO portfolios, risk parity is not explicitly concerned with minimizing portfolio risk. Instead, risk parity aims to attain perfect risk diversification by equalizing the

asset risk contributions. Thus, mitigating the effect of misspecification on the marginal risk contributions is paramount when constructing risk parity portfolios. We begin by introducing a SOCP version of the nominal risk parity problem as in [72]. We proceed to introduce robustness by targeting the two constraints pertinent to the risk measure: the overall portfolio risk and the marginal risk contributions. The rationale for selecting a SOCP formulation is that it allows for a tractable introduction of robustness while still maintaining the same level of complexity as the original problem (i.e., the robust problem is also a SOCP).

Robust optimization allows us to frame the problem deterministically by taking the most extreme estimates of our uncertain parameters within some distributional bounds when solving our optimization problem [9, 10, 12]. The robust instance of most MVO problems is attained when the worst-case estimate of the risk measure is assumed [32, 38, 52, 69, 96]. In the case of portfolios where short positions are not allowed, a robust 'worst-case' estimate of the variance is attained simply by assuming the upper bound estimate of the covariance matrix. This shields a portfolio whose objective is to minimize risk from estimation error in the risk measure. This same approach can be directly applied to risk parity, as proposed in [60], where a robust problem is formulated by taking the worst-case estimate of the covariance matrix subject to a risk diversification constraint. In addition, [26] propose a robust risk parity problem that implements the uncertainty structure proposed in [49]. This method relies on deriving a worst-case estimate of the covariance matrix arising from the errors in the regression coefficients.

Introducing robustness by taking the 'worst-case' instance of the portfolio variance is an intuitive approach when we only seek to minimize our financial risk. However, this is not tailored towards an optimization problem where the objective is to fully diversify our risk, regardless of the overall portfolio risk. Therefore, our proposed robust formulation deviates away from traditional robust methods based solely on worst-case estimates of the covariance matrix. The aforementioned techniques may be unable to fully capture the intricacies behind a risk parity portfolio, where overall portfolio risk is not a concern, but where we seek to correctly estimate the risk contribution per asset in order to allocate wealth accordingly. Thus, we propose a model that relaxes the 'worst-case' variance assumption by focusing not only on the overall risk of the portfolio, but also on the more relevant marginal risk contributions.

3.1 Covariance uncertainty structure

In this section we discuss the uncertainty structure that we will use to construct a robust risk parity problem. The only estimated parameter involved in the construction of these portfolios is the measure of risk. Thus, in our case, we are only concerned with uncertainty in the asset covariance matrix.

We can design the uncertainty set in a straightforward fashion by placing a set of box constraints on each element of the covariance matrix. Let $\mathbf{\Sigma} \in \mathbb{S}_+^n$ be the true (but latent) asset covariance matrix. If we assume that the true covariance matrix lies somewhere in-between upper and lower bounds estimated from our data, then we can define the uncertainty set of box constraints as

$$\mathcal{U}_{\mathbf{\Sigma}} = \left\{ \mathbf{\Sigma} \in \mathbb{S}_+^n : \ \underline{\Sigma}_{ij} \leq \Sigma_{ij} \leq \overline{\Sigma}_{ij} \right\}, \tag{3.1}$$

where both $\underline{\Sigma}_{ij}$ and $\overline{\Sigma}_{ij}$ are defined by the user depending on their choice of parameter estimation technique. Examples of uncertainty bounds on the covariance matrix can be found in [32, 69].

For the case of long-only minimum variance portfolios, it has been established that a robust formulation can be achieved simply by taking the upper bound of the covariance matrix, i.e., we assume the worst-case estimate of the covariance matrix [96]. Thus, having $\Sigma_{ij} = \overline{\Sigma}_{ij} \ \forall \ i, j$ as our covariance matrix may introduce robustness if we seek to minimize the portfolio variance, but it may not necessarily align with the risk parity objective. The objective of risk parity is to equalize the risk contribution per asset, and not necessarily to minimize portfolio risk. Therefore, assuming the worst-case instance of the covariance matrix may not be relevant for our purpose.

The set of box constraints on the covariance matrix in (3.1) can be expressed as a perturbation on the nominal estimate as follows

$$\begin{aligned} \mathcal{U}_{\mathbf{\Sigma}} = \big\{ \mathbf{\Sigma} \mid \exists \mathbf{\Delta} \in \mathbb{R}^{n \times n} : \ & \mathbf{\Sigma} = \hat{\mathbf{\Sigma}} + \mathbf{\Delta} \circ \mathbf{\Sigma}^{\Delta}, \\ & \underline{\Sigma}_{ij} \leq (\hat{\Sigma}_{ij} + \Delta_{ij} \Sigma_{ij}^{\Delta}) \leq \overline{\Sigma}_{ij}, \ -1 \leq \Delta_{ij} \leq 1 \big\}, \end{aligned} \tag{3.2}$$

where $\hat{\mathbf{\Sigma}} \in \mathbb{S}_+^n$ is the nominal covariance matrix estimated from data, $\mathbf{\Delta} \in \mathbb{R}^{n \times n}$ are independent and identically distributed random variables with mean zero and serve to define the independent perturbations on each element Σ_{ij}, and $\mathbf{\Sigma}^{\Delta} \in \mathbb{R}^{n \times n}$ is a constant that appropriately scales the perturbation on the nominal estimate.

In the case of where the covariance matrix is bounded by fixed upper and lower limits, a simple and tractable way to size the perturbation is by proceeding as in Tütüncü and Koenig [96] and let it equal the difference between the nominal and the worst-case variance, i.e., $\mathbf{\Sigma}^{\Delta} = \overline{\mathbf{\Sigma}} - \hat{\mathbf{\Sigma}}$.

We seek to create a robust portfolio that will reduce our exposure to assets with a higher degree of error in their estimated risk contribution per asset. The risk contribution of asset i, r_i, is defined as the product of the marginal risk contribution multiplied by the asset weight, i.e., $r_i \triangleq x_i[\mathbf{\Sigma}\boldsymbol{x}]_{\boldsymbol{i}}$. Since the estimation error is intrinsic to the covariance matrix, we will introduce robustness into our risk parity model by targeting the marginal risk contributions. This procedure is explained in Section 3.2. For now, we proceed to show how to estimate the size of the covariance perturbation, $\mathbf{\Sigma}^{\Delta}$, using a factor model of asset returns.

3.1.1 Derivation of the uncertainty structure from factor models

The covariance perturbation from (3.2), $\mathbf{\Sigma}^{\Delta}$, can be defined using any desirable set of bounds on the covariance matrix. As an example, this section demonstrates how to derive $\mathbf{\Sigma}^{\Delta}$ from the estimation error of the coefficients from a factor model of asset returns. Specifically, we will size the perturbation using the standard error from the estimated factor loadings.

The factor model and its estimation procedure were presented in detail in Section 2.1.2. In particular, the calculation of the standard error from the estimated factor loadings was presented in (2.12).

In turn, we can use the standard errors to construct the uncertainty set around the covariance matrix. Let $\boldsymbol{V}$ be the true (but unknown) matrix of factor loadings, and let it belong to the uncertainty set

$$\begin{aligned} \mathcal{U}_V = \big\{ \boldsymbol{V} \mid \exists \mathbf{\Delta} \in \mathbb{R}^{m \times n} :\ & \boldsymbol{V} = \hat{\boldsymbol{V}} + \mathbf{\Delta} \circ \boldsymbol{V}^{\Delta}, \\ & -\operatorname{SE}(\hat{V}_{ij}) \leq V_{ij}^{\Delta} \leq \operatorname{SE}(\hat{V}_{ij}),\ -1 \leq \Delta_{ij} \leq 1 \big\} \end{aligned} \tag{3.3}$$

where $\hat{\boldsymbol{V}} \in \mathbb{R}^{m \times n}$ is the least squares estimate of $\boldsymbol{V}$ and $\operatorname{SE}(\hat{V}_{ij})$ denotes the standard error of the estimated factor loading of factor i corresponding to asset j. Estimating a worst-case covariance matrix from a factor model with uncertain parameters is a difficult process. A closed-form solution is not possible due to the nature of the factor covariance matrix $\boldsymbol{F}$, where having either positive or negative covariance between the factors may influence the size and

direction of the perturbation.

Instead, we can find the set of factor loadings that define the upper bound of the covariance matrix by formulating a simple mathematical program, where we seek to maximize the sum of all elements in the covariance matrix under the constraints given by $\mathcal{U}_V$, i.e.,

$$\boldsymbol{V}^* = \underset{\boldsymbol{V} \in \mathcal{U}_V}{\operatorname{argmax}} \quad \mathbf{1}^\top (\boldsymbol{V}^\top \boldsymbol{F} \boldsymbol{V} + \boldsymbol{D}) \mathbf{1},$$

$$\overline{\boldsymbol{\Sigma}} = \boldsymbol{V}^{*\top} \boldsymbol{F} \boldsymbol{V}^* + \boldsymbol{D}.$$

We can then use the nominal and worst-case estimates of the covariance matrix to recover the corresponding perturbation $\boldsymbol{\Sigma}^\Delta$,

$$\boldsymbol{\Sigma}^\Delta = \overline{\boldsymbol{\Sigma}} - \hat{\boldsymbol{\Sigma}}.$$

Finally, we use the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$ and the element-wise perturbation $\boldsymbol{\Sigma}^\Delta$ to define the uncertainty set in (3.2).

The remainder of this chapter assumes that we use a factor model to define the perturbation. However, we note that the robust risk parity framework in this chapter accepts any alternative method to estimate $\boldsymbol{\Sigma}^\Delta$.

3.2 Robust risk parity

This section presents a robust optimization problem specifically tailored towards risk parity. We will use the SOCP formulation of the risk parity problem shown in (2.20). The reason we use the SOCP formulation is because this allows us to introduce robustness to the marginal risk contributions per asset. Moreover, using the SOCP formulation will allow us to maintain the same level of complexity of the original problem, i.e., the robust counterpart is also a SOCP.

The nominal SOCP in (2.20) defined the portfolio risk in constraint (2.20e) and the marginal risk contributions in constraint (2.20c). In particular, constraint (2.20e) pertains to the minimization of the average portfolio risk, which is fundamentally equivalent to minimizing the total portfolio risk. Given that our risk parity portfolio disallows short sales, we can introduce robustness in a similar fashion to Tütüncü and Koenig [96] and take the worst-case instance of the covariance matrix as the input parameter. In other words, the original constraint (2.20e)

becomes

$$\left\|(\boldsymbol{\Sigma})^{1/2}\boldsymbol{x}\right\|_2 \leq \sqrt{n}\ p \quad \Rightarrow \quad \left\|(\hat{\boldsymbol{\Sigma}}+\boldsymbol{\Sigma}^{\Delta})^{1/2}\boldsymbol{x}\right\|_2 \leq \sqrt{n}\ p.$$

Constraint (2.20c) assigns the marginal risk contribution per asset to an auxiliary variable $\boldsymbol{z} \in \mathbb{R}^n_+$, i.e., $z_i = [\boldsymbol{\Sigma}\boldsymbol{x}]_i\ \forall\ i$. First, we should realize that this constraint can be relaxed to $z_i \leq [\boldsymbol{\Sigma}\boldsymbol{x}]_i$, as it will become tight at optimality. This relaxation allows us to proceed in a similar fashion to Bertsimas and Sim [13] and introduce robustness in the form of an error term that penalizes the marginal risk contributions. Thus, the original constraint (2.20c) becomes

$$z_i \leq [\boldsymbol{\Sigma}\boldsymbol{x}]_i \quad \Rightarrow \quad \Omega\zeta \leq [\hat{\boldsymbol{\Sigma}}\boldsymbol{x}]_i - z_i,$$

where $\Omega \in \mathbb{R}_+$ is a penalty parameter and $\zeta \in \mathbb{R}_+$ is an auxiliary variable. In turn, ζ allows us to model the error of the marginal risk contributions as a second-order cone constraint (SOCC) as follows

$$\sqrt{n}\ \zeta \geq \left\|\boldsymbol{\Sigma}^{\Delta}\boldsymbol{x}\right\|_2.$$

By design, ζ imposes the same penalty term on our n marginal risk contributions. The penalty is modelled in this fashion because the marginal risk contributions are fundamentally linked by the vector of asset weights, $\boldsymbol{x}$. Thus, all marginal risk contributions should be equally penalized.

After addressing the constraints pertinent to $\boldsymbol{\Sigma} \in \mathcal{U}_{\boldsymbol{\Sigma}}$, we can modify our original SOCP from (2.20) to define the robust risk parity problem,

$$\min_{\boldsymbol{x},\boldsymbol{z},u,v,\zeta} \quad u - v \tag{3.4a}$$

$$\text{s.t.} \qquad \mathbf{1}^\top\boldsymbol{x} = 1, \tag{3.4b}$$

$$\left\|\boldsymbol{\Sigma}^{\Delta}\boldsymbol{x}\right\|_2 \leq \sqrt{n}\ \zeta, \tag{3.4c}$$

$$\Omega\zeta \leq [\hat{\boldsymbol{\Sigma}}\boldsymbol{x}]_i - z_i, \quad i = 1,\ldots,n, \tag{3.4d}$$

$$\left\|\begin{bmatrix} 2v \\ x_i - z_i \end{bmatrix}\right\|_2 \leq x_i + z_i, \qquad i = 1,\ldots,n, \tag{3.4e}$$

$$\left\|(\hat{\boldsymbol{\Sigma}}+\boldsymbol{\Sigma}^{\Delta})^{1/2}\boldsymbol{x}\right\|_2 \leq \sqrt{n}\ u, \tag{3.4f}$$

$$\boldsymbol{x}, \boldsymbol{z} \geq \mathbf{0}, \tag{3.4g}$$

$$u, v, \zeta \geq 0. \tag{3.4h}$$

Conceptually, the robust problem attempts to reduce the size of the error term in constraint (3.4d) as it tightens. This not only diminishes our exposure to assets with larger estimation error in their marginal risk contributions, but also yields a set of error-adjusted marginal risk contributions in the form of the variable $\boldsymbol{z}$. The subsequent attempt to equalize risk contributions is based on $\boldsymbol{z}$, implicitly reducing our exposure to assets with larger marginal risk contributions and further shielding the portfolio against riskier assets. Additionally, constraint (3.4f) ensures we are still bracing the portfolio against the worst-case instance of the covariance matrix.

Introducing robustness in this fashion is desirable for two reasons. First, introducing individual error terms for the marginal risk contribution of each asset would not be able to capture the intricacies of the risk contributions, where the marginal risk contribution of asset i is not only dependent on itself, but also on its interaction with asset j as measured by their covariance $\sigma_{ij}\ \forall\ j \neq i$. Thus, formulating this as a SOCC ensures these dependencies are captured. Second, this formulation allows us to model the error term as a single SOCC, thereby avoiding a convoluted set of constraints that would otherwise increase the computational complexity of the problem.

The penalty parameter Ω scales the sensitivity of the marginal risk contributions to the error term. Therefore, a larger value of Ω corresponds to a more conservative outlook. The parameter Ω may be sized to provide a probabilistic guarantee on the results, as shown by Bertsimas and Sim [13]. However, we prefer to proceed with a data-driven approach to determine an appropriate value of Ω, which is based on the relative size between our nominal covariance matrix $\hat{\boldsymbol{\Sigma}}$ and its corresponding perturbation $\boldsymbol{\Sigma}^{\Delta}$. We have that

$$\Omega \triangleq \omega\ \frac{\|\boldsymbol{\Sigma}^{\Delta}\|_F}{\|\hat{\boldsymbol{\Sigma}}\|_F},$$

where $\omega \in \mathbb{R}_+$ serves to scale Ω based on the ratio of the nominal covariance matrix to its perturbation. Intuitively, setting $\omega = 0$ will revert the robust risk parity problem to its nominal counterpart. On the other hand, an excessively large value of ω would be prohibitively conservative, rendering the robust SOCP infeasible. Thus, to guarantee feasibility, we must have that

$$0 \leq \omega < \sqrt{n} \cdot \frac{[\hat{\boldsymbol{\Sigma}}\boldsymbol{x}]_i}{\|\boldsymbol{\Sigma}^{\Delta}\boldsymbol{x}\|_2} \cdot \frac{\|\hat{\boldsymbol{\Sigma}}\|_F}{\|\boldsymbol{\Sigma}^{\Delta}\|_F}, \quad i = 1, ..., n.$$

We can use ω to determine the level of robustness according to the risk appetite of the user.

In other words, ω is a risk aversion parameter. As we proceed onto the numerical experiment section, we will test robust portfolios with different values of ω.

3.3 Numerical experiments

This section presents two experiments designed to test the out-of-sample performance of the robust risk parity problem. The first experiment tests multiple portfolios each with $n = 25$ assets and studies six robust portfolios with varying degrees of robustness, testing different values of the parameter ω. The second experiment tests the impact of robustness for portfolios of different sizes, testing different number of assets n.

The general methodology is the same for both experiments, both described below. The nominal estimate of the covariance matrix, $\hat{\mathbf{\Sigma}}$ and its maximum absolute perturbation, $\mathbf{\Sigma}^{\Delta}$, are built using a factor model as shown in Section 2.1.2. Specifically, we use the Fama–French three-factor model [39] as the basis model for both experiments. The Fama–French model was purposely chosen to show how a well-known multi-factor model would behave under our proposed robust framework. The historical factor returns were obtained from Kenneth R. French's data library [42].

For both experiments, the assets used for portfolio construction are selected from a universe of 250 diverse stocks that are regularly traded in major U.S. exchanges. A list of these stocks is shown in Table 3.1, with representative stocks from each of the eleven *Global Industry Classification Standard* (GICS) sectors. The historical stock prices were obtained from Quandl.com [84]. The experiment is performed using weekly historical data from 01-Jan-1995 to 31-Dec-2016, with the first five-year period used to perform the initial factor model calibration and subsequent parameter estimation (i.e., the out-of-sample period begins on 01-Jan-2000). Thereafter, the portfolios are rebalanced every six months, re-estimating the parameters using data from the preceding five-year period. The length of this window introduces a survivorship bias to our experiments as we only consider stocks with sufficient historical data. However, we expect this bias to have a similar effect on all portfolios. Thus, it is the relative performance between the portfolios that is of main interest.

Both experiments consist of 1,000 independent trials, where a basket of n assets is randomly drawn from the universe of 250 assets at the beginning of each trial. This basket of n assets is held constant throughout the duration of the trial. Using the rolling window approach de-

Table 3.1: List of assets

GICS Sector	**Company Tickers**								
Energy	XOM	HAL	OXY	MRO	SLB	CVX	HES	APA	COG
	COP	DVN	EOG	AE	APC	BP	CKH	EGN	
Materials	MOS	NEM	IP	IFF	NUE	PPG	APD	AVY	BLL
	ECL	FMC	SEE	SHW	VMC	AP	BMS	CCK	CRS
	EMN	FUL	GLT	MLM	OLN	SON	TG		
Industrials	BA	CAT	GE	DE	HON	LMT	GD	CSX	CMI
	DOV	ETN	EFX	FDX	ABM	AIR	ALG	AME	AOS
	BGG	CSL	MMM	HNI	NPK	RHI	SPA	SSD	
Cons. Disc.	MCD	F	GPS	HRB	NKE	AZO	HOG	HAS	HD
	JWN	TGT	AAN	CBRL	BKS	ETH	GT	LOW	NWL
	PZZA	RCL	WWW						
Cons. Staples	WMT	KO	KR	HSY	CL	CLX	WBA	COST	MO
	CPB	GIS	MKC	PEP	PG	ALCO	FARM	FLO	CAG
	CASY	LANC	ODC	TSN	UVV	STZ	WMK	TR	
Healthcare	BMY	PFE	JNJ	BAX	CVS	CI	DHR	HUM	ABT
	AGN	AMGN	BDX	BSX	LLY	BIO	CAH	HAE	HRC
	LH	MCK	MRK	OMI	PKI	STE	SYK	TMO	
Financials	JPM	BAC	AON	AFL	WFC	AXP	BK	BEN	MS
	AFG	ALL	BANF	BXS	CMA	EV	KEY	LM	MSL
	PNC	STT	WTM	TMP					
IT	HPQ	IBM	MSFT	INTC	ADSK	ADP	CSCO	GLW	ORCL
	PAYX	TXN	PLT	XRX	ROG	TSS			
Comm. Serv.	DIS	VZ	S	EA	CTL	T	IPG	OMC	NYT
	VOD	CBB	RDI	SSP	MDP				
Utilities	CNP	DTE	DUK	ED	AEP	D	ETR	ATO	ES
	EXC	NEE	PEG	SO	AVA	AWR	BKH	CMS	CPK
	GXP	NFG	NJR	OGE	PNM	PNW	PPL	SWX	UGI
	XEL	SJW							
Real Estate	REG	HCP	UDR	DRE	AIV	WY	KIM	MAA	AVB
	EQR	HST	PSA	SPG	VNO	ADC	ALX	BDN	BFS
	CLI	CPT	CTO	CUZ	EGP	ELS	GTY	JOE	LXP
	MAC	WRI							

scribed in the previous paragraph, we estimate our parameters and optimize our portfolios using three different optimization models to construct our optimal portfolios. The three optimization models are described below.

- Nominal Model: The nominal risk parity SOCP from (2.20) with the nominal covariance estimate, $\hat{\mathbf{\Sigma}}$, as the input.
- Worst-Case Model: The optimization model is the nominal SOCP from (2.20), except we use the 'worst-case' instance of the portfolio variance, $\overline{\mathbf{\Sigma}}$, as the input.
- Robust Model: The robust SOCP from (3.4), which takes both $\hat{\mathbf{\Sigma}}$ and $\mathbf{\Sigma}^{\Delta}$ as inputs, as well as the user-defined risk aversion parameter ω. A larger value of ω yields robust portfolios with increasingly conservative outlooks.

Once a trial is complete, we repeat the process by randomly drawing a new set of n assets and creating the corresponding optimal portfolios using the aforementioned rolling window approach. This process is repeated for each of the 1,000 trials.

All experiments were conducted on an Apple MacBook Pro computer (2.8 GHz Intel Core i7, 16 GB 2133 MHz DDR3 RAM) running macOS 'Catalina'. The computer script was written in the Julia programming language (version 1.4.0) using the modelling language 'JuMP' [33] with MOSEK (version 9.0.1) as the optimization solver.

3.3.1 Experiment with varying degrees of robustness

Our first experiment analyzes portfolios with $n = 25$ assets. There are eight portfolios in this experiment: one nominal portfolio, one 'worst-case' portfolio, and six robust portfolios with $\omega =$ 0.5, 1.0, 1.5, 2.0, 2.5, 3.0. The experiment consists of 1,000 trials. Each trial begins by randomly drawing 25 assets from those listed in Table 3.1. We then construct eight portfolios using these 25 assets. The necessary parameters are estimated using the Fama–French model and the eight portfolios are built using the corresponding optimization models. The trial is conducted using a rolling window where parameters are re-estimated and portfolios are re-optimized every six months over the 17-year investment horizon. We then repeat this process for every subsequent trial, randomly drawing a new set of 25 constituent assets at the beginning of every trial.

The portfolio wealth evolution is presented in Figure 3.1. For clarity, this figure shows the absolute and relative wealth of only three of our eight portfolios. The top plot in Figure 3.1

shows the average wealth evolution of the nominal, 'worst-case', and robust ($\omega = 2.0$) portfolios. The average wealth is calculated by taking the average value of the 1,000 trials at each discrete point in time. The error bars display the standard deviation corresponding to the distribution of the 1,000 trials. The bottom plot shows the wealth of the 'worst-case' and robust portfolios relative to the nominal, i.e., $(W_i^t/W_{\text{Nom}}^t - 1) \times 100$ for the wealth W_i^t of portfolio i at time t.

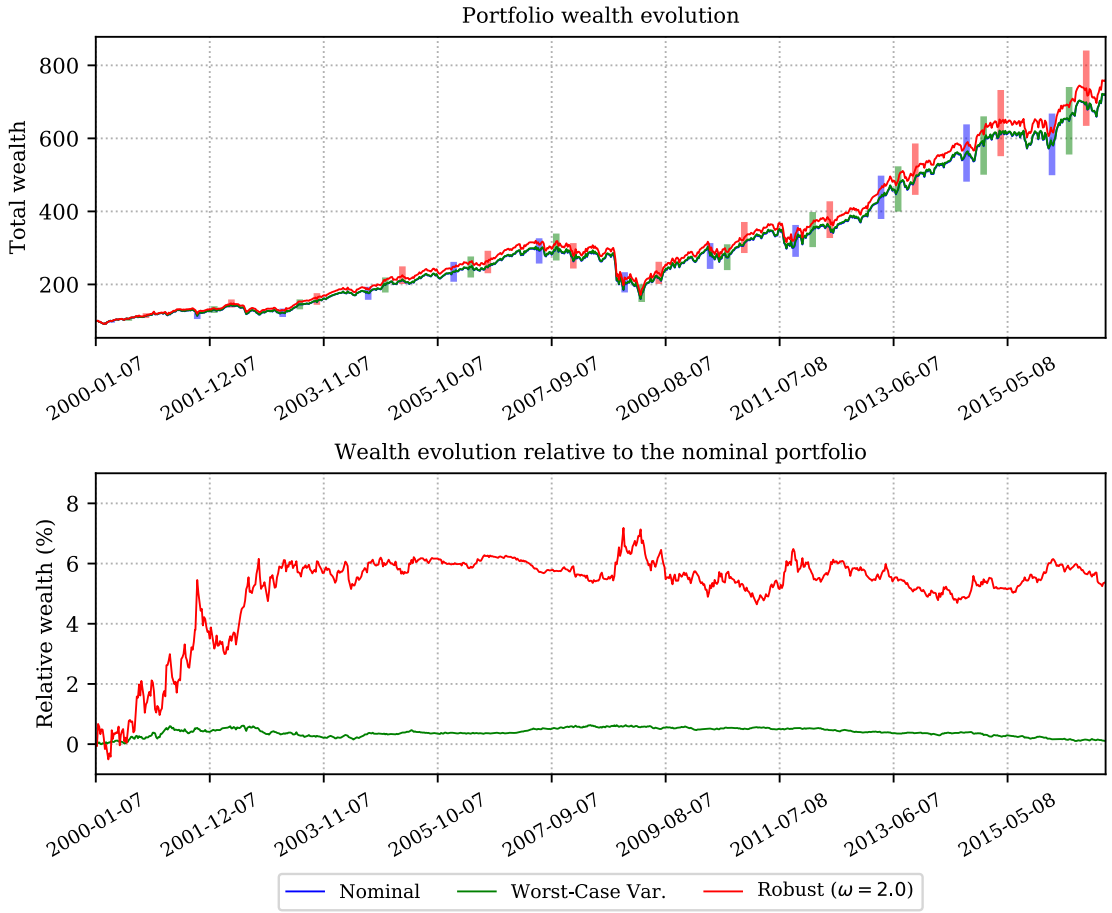

Figure 3.1: Wealth evolution of 1,000 individual trials with 25 assets per portfolio over the period 2000–2016

Notes: The top plot shows the average wealth of portfolios over the 1,000 trials with the standard deviation shown as error bars. The bottom plot shows the average wealth relative to the nominal portfolio.

The results show that our robust portfolio with $\omega = 2.0$ behaves as expected, outperforming the nominal during bear market periods. This is exemplified in the bottom plot of Figure 3.1 by the relative gains made by the robust portfolio during the recession of the early 2000s, as well as the spike observed during the recent financial crisis of 2008. Moreover, the robust portfolio appears to be able to maintain its relative advantage throughout bull market periods as

well. On the other hand, the 'worst-case' variance portfolio closely mimics the nominal, having a marginally better performance but not displaying a meaningful behaviour during different market cycles.

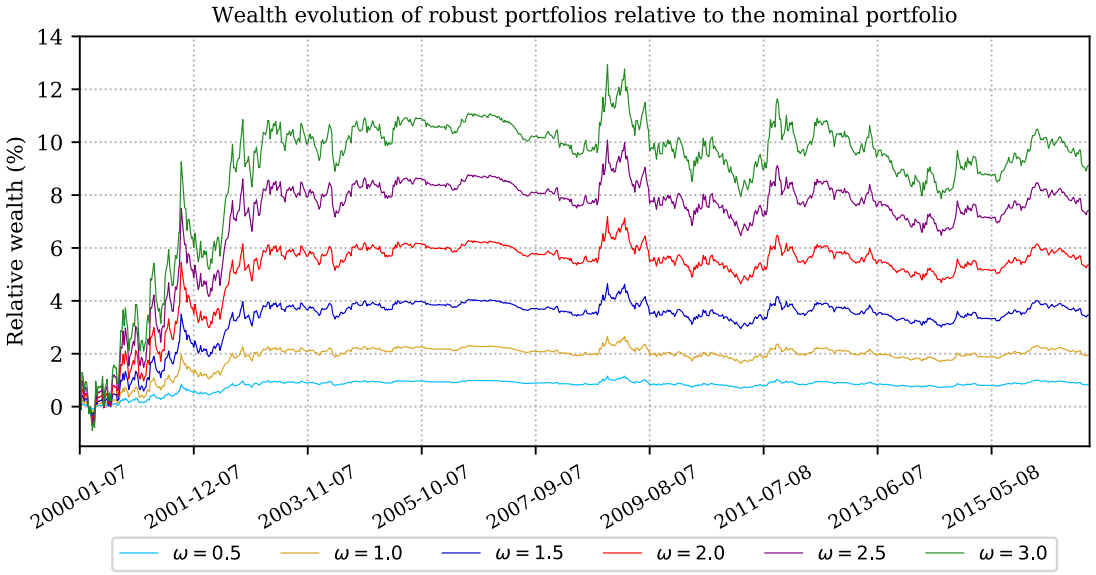

Figure 3.2: Average wealth evolution of robust portfolios relative to the nominal portfolio for several values of ω

Next, we discuss the performance of the six robust portfolios for varying values of ω. The purpose is to analyze the out-of-sample effect of different levels of robustness. Figure 3.2 shows the average wealth evolution of the six robust portfolios relative to the nominal. The plot of the robust portfolio with $\omega = 2.0$ is the same as in Figure 3.1. Figure 3.2 suggests that taking an increasingly conservative stance can significantly improve the relative performance of the robust portfolios. For example, the robust portfolio with $\omega = 3.0$ is able to attain a peak advantage of 12.94% over the nominal during the financial crisis of 2008. It is worthwhile to note that the upward trend seen during bearish periods does not necessarily imply that the robust portfolios had a positive return, but rather that their losses were not as pronounced as those of the nominal portfolio. Moreover, the volatility observed on a relative wealth plot, such as Figure 3.2, is the result of the relative difference in wealth evolution between the nominal portfolio and its robust counterpart. In other words, the relative wealth plots may appear volatile even when the robust portfolios exhibit low volatility if this coincides with a period of high volatility of the underlying nominal portfolio.

A summary of the financial performance of all eight portfolios is presented in Table 3.2.

These results are calculated as follows. The portfolio excess returns are calculated by finding the weekly return of each portfolio from each individual trial and subtracting the corresponding weekly risk-free rate. We then use the time series of observed weekly excess returns to calculate the annualized excess return, volatility and ex-post Sharpe ratio for each portfolio per trial. Finally, we calculated the average and standard deviation over the 1,000 trials. For clarity, the row in Table 3.2 labelled 'Ann. Volatility' corresponds to the average annualized volatility experienced by the portfolios over the 1,000 trials; the row labelled 'σ_{Vol}' is the standard deviation of the observed volatilities over the 1,000 trials. Finally, the turnover rate measures the average period-over-period absolute change in the asset weights per trial, with the average over the 1,000 trials labelled as the 'Turnover Rate' and its corresponding standard deviation as 'σ_{TR}'. The turnover rate serves to exemplify potential transaction costs.

Table 3.2: Summary of financial performance of 1,000 trials with $n = 25$

	Nom.	WC	Robust					
			$\omega = 0.5$	$\omega = 1.0$	$\omega = 1.5$	$\omega = 2.0$	$\omega = 2.5$	$\omega = 3.0$
Ann. Ex. Return (%)	9.76	9.77	9.81	9.88	9.96	10.07	10.18	10.27
σ_{Ret} (%)	0.84	0.83	0.84	0.84	0.84	0.85	0.85	0.86
Ann. Volatility (%)	16.19	16.20	16.09	15.98	15.86	15.73	15.60	15.49
σ_{Vol} (%)	0.94	0.95	0.94	0.94	0.94	0.94	0.95	0.97
Sharpe Ratio (%)	60.49	60.51	61.17	62.01	63.04	64.24	65.48	66.57
σ_{Sharpe} (%)	6.45	6.40	6.48	6.52	6.60	6.71	6.77	6.93
Turnover Rate (%)	5.13	4.97	5.58	6.20	7.04	8.22	9.88	11.94
σ_{TR} (%)	3.17	3.03	3.42	3.81	4.50	5.71	7.68	9.83

Notes: Nom, nominal. WC, worst-case. Ann, annualized. Ex, excess.

The results in Table 3.2 show a clear trend when comparing the robust portfolios against the nominal. As the level of robustness is increased, the realized excess return increases while the volatility decreases. In turn, this translates into an increasing ex-post Sharpe ratio as the level of robustness increases. The robust risk parity problem attempts to reduce our exposure to assets with large ex-ante errors in their estimated risk contributions. Thus, our results suggest that a reduced exposure to such assets helps to mitigate portfolio losses. In general, mitigating portfolio losses yields an improved long-run rate of return while maintaining a lower volatility.

In contrast to the robust risk parity problem, the worst-case portfolio is based on the nominal risk parity problem with the only difference being that we use the worst-case estimate of the

covariance matrix as the input. Thus, it is not surprising to see that the worst-case portfolio has a similar behaviour to the nominal itself. This can be observed in the relative wealth evolution in the bottom plot of Figure 3.1, as well as summarized in the ex-post Sharpe ratio in Table 3.2.

Finally, the turnover rate of our robust portfolios increased as the level of robustness increased. This suggests that penalizing our positions on assets with noisier marginal risk contributions resulted in greater period-over-period changes to our optimal weights. It is worthwhile to note that nominal risk parity portfolios are generally very stable in their period-over-period wealth reallocation when compared to a broader range of optimal portfolios, such as MVO [22]. Thus, although the robust portfolios exhibit a higher turnover rate when compared against the nominal risk parity portfolio, their turnover rate may still be lower when compared against other asset allocation strategies.

An analysis of the Sharpe ratio indicates that the observed differences between the nominal and robust portfolios is statistically significant. A summary of this analysis is shown in Table 3.3. The first row of this table indicates the total number of trials for which the Sharpe ratio of a given portfolio was greater than the nominal portfolio. For example, from the first row we can see that the Sharpe ratio of the robust portfolio with $\omega = 2.0$ was greater than its nominal counterpart during 998 out of 1,000 trials. Similarly, by looking at Table 3.3, we can see that the robust portfolios outperformed the nominal in almost every trial. On the other hand, the worst-case portfolio had a superior performance in only 542 times out of 1,000. The second row of this table presents the t-statistic for a paired sample t-test. We performed this analysis by calculating the pair-wise difference in the ex-post Sharpe ratio between the nominal portfolio and all other competing portfolios over each of the 1,000 trials. Thus, for this number of trials, we have 999 degrees of freedom. The t-statistics in Table 3.3 imply that we have more than 99.9% confidence that the Sharpe ratio of the robust portfolios is greater than the nominal. However, we can only say the same of the worst-case portfolio with less than 90% confidence.

Next, we discuss the difference in risk concentration between the portfolios. The results in Table 3.4 serve to quantify our trade-off between risk parity and robustness. This 'cost' of robustness can be measured by our deviation away from perfect risk diversification. For our experiment, the portfolio risk concentration is measured relative to $\hat{\mathbf{\Sigma}}$. Since each individual portfolio is re-optimized and rebalanced 34 times per trial, the values in Table 3.4 correspond to the average and standard deviation of 34,000 observations for each portfolio. As defined

Table 3.3: Sharpe ratio analysis of 1,000 trials with $n = 25$

	WC	Robust					
		$\omega = 0.5$	$\omega = 1.0$	$\omega = 1.5$	$\omega = 2.0$	$\omega = 2.5$	$\omega = 3.0$
# of trials greater than nominal	542	1000	1000	999	998	997	988
t-statistic	1.247	73.68	70.51	68.07	68.61	72.20	69.77

Notes: WC, worst-case.

in Section 2.3, we report the following three risk concentration measures: CV, HRC, and the H-index. These three measures are calculated as shown in equations (2.21), (2.22) and (2.23), respectively.

Table 3.4: Summary of risk concentration measures of 1,000 trials with $n = 25$

	Nom.	WC	Robust					
			$\omega = 0.5$	$\omega = 1.0$	$\omega = 1.5$	$\omega = 2.0$	$\omega = 2.5$	$\omega = 3.0$
CV	0.000	0.054	0.035	0.080	0.139	0.219	0.322	0.448
σ_{CV}	0.000	0.029	0.023	0.060	0.121	0.208	0.319	0.437
HRC	0.040	0.044	0.044	0.049	0.056	0.066	0.080	0.097
σ_{HRC}	0.000	0.002	0.003	0.008	0.016	0.029	0.047	0.068
H-index	0.040	0.040	0.040	0.040	0.041	0.044	0.048	0.055
σ_H	0.000	0.000	0.000	0.001	0.003	0.010	0.023	0.040

Notes: Nom, nominal. WC, worst-case. CV, coefficient of variation. HRC, highest risk contribution. H-index, Herfindahl index.

Table 3.4 shows that, as robustness increases, we deviate further away from perfect risk diversification. This holds for all three measures of risk concentration. This is not surprising, as it becomes increasingly difficult to attain equalized asset risk contributions while simultaneously limiting our exposure to assets with noisier marginal risk contributions. Nevertheless, we are able to maintain a sufficient degree of diversity even when we assess the risk contributions relative to the nominal estimate of the covariance matrix.[1] This is particularly true of the robust portfolios with low ω values, where the average HRC and H-index values are similar to the nominal portfolio. However, risk concentration increases as we become increasingly conservative. Overall, the results suggest that the proposed robust framework is still able to

[1]When we use the nominal covariance matrix, $\hat{\boldsymbol{\Sigma}}$, as the risk measure, only one perfect risk-diverse portfolio exists: the nominal risk parity portfolio. Any other portfolio $\boldsymbol{x} \in \mathbb{R}^n$ that differs from the nominal will have a CV greater than zero, as well as H-index and HRC values greater than $1/n$. Thus, when risk contributions are measured using $\hat{\boldsymbol{\Sigma}}$, none of the competing portfolios can attain perfect risk diversification.

attain a risk-diverse portfolio, particularly for low values of ω.

3.3.2 Experiment with different portfolio sizes

Our second experiment compares portfolios of different sizes with n =15, 50, 75, 100. As before, we performed 1,000 trials for each value of n. For each trial, we construct four portfolios: the nominal portfolio, the 'worst-case' portfolio, and two robust portfolios with $\omega = 1.0$, 2.0.

For the sake of brevity, we summarize the experimental results in tables. Table 3.5 presents the results corresponding to the 1,000 trials for each value of n. These results are calculated in the same fashion as our previous experiment. Evaluating the average portfolio performance over the course of the entire investment horizon shows that the robust portfolios outperform the nominal in all three categories: excess return, volatility and Sharpe ratio. As before, the turnover rate is larger for the robust portfolios. In particular, the rebalancing of the robust portfolios is more drastic as the number of assets increases. This is not surprising since, as n increases, we have greater choice for our asset allocation strategy, as well as having more noise from the increase in estimated parameters.

Moreover, while the robust portfolios are consistently able to financially outperform the nominal, the performance of the worst-case portfolio worsens as the portfolio size increases. In particular, the worst-case portfolio fails to surpass the nominal once $n \geq 50$. Finally, we note that the performance of the nominal portfolio improves as the number of assets increases. This improvement with size is not steady. The portfolios benefit from greater diversification between more assets, but this benefit diminishes as n increases.

A statistical analysis of the Sharpe ratio is shown in Table 3.6. Similar to the previous experiment, we find that the Sharpe ratio of the robust portfolios exceeds that of the nominal portfolio in almost every single trial. In fact, for $n \geq 50$, the Sharpe ratio of the robust portfolio with $\omega = 1.0$ exceeds the nominal in every single trial. Moreover, the t-statistic of the difference in the Sharpe ratio between robust and nominal portfolios suggests that the robust portfolios have a better risk-adjusted performance with more than 99.9% confidence for over all values of n. However, we note that the Sharpe ratio of the robust portfolios with $\omega = 2.0$ begins to deteriorate when $n \geq 50$. Finally, in line with other results, the Sharpe ratio of the worst-case portfolio worsens as size increases. In fact, when $n \geq 50$, we can see that the t-statistic indicates with more than 99.9% confidence that the Sharpe ratio is lower for the worst-case portfolio when compared against the nominal.

Table 3.5: Summary of financial performance of 1,000 trials with $n = 15,\ 50,\ 75,\ 100$

	$n = 15$				$n = 50$			
	Nom.	WC	Robust		Nom.	WC	Robust	
			$\omega = 1.0$	$\omega = 2.0$			$\omega = 1.0$	$\omega = 2.0$
Ann. Ex. Return (%)	9.59	9.61	9.65	9.73	9.92	9.89	10.17	10.58
σ_{Ret} (%)	1.08	1.08	1.08	1.08	0.61	0.59	0.62	0.68
Ann. Volatility (%)	16.86	16.85	16.71	16.54	15.68	15.73	15.35	15.09
σ_{Vol} (%)	1.27	1.28	1.27	1.27	0.65	0.65	0.66	0.69
Sharpe Ratio (%)	57.27	57.44	58.10	59.19	63.41	63.03	66.43	70.32
σ_{Sharpe} (%)	8.19	8.20	8.25	8.36	4.97	4.84	5.28	5.79
Turnover Rate (%)	4.61	4.56	5.24	6.21	5.80	5.46	8.01	14.53
σ_{TR} (%)	3.02	2.96	3.39	4.10	3.42	3.17	5.18	14.62
	$n = 75$				$n = 100$			
	Nom.	WC	Robust		Nom.	WC	Robust	
			$\omega = 1.0$	$\omega = 2.0$			$\omega = 1.0$	$\omega = 2.0$
Ann. Ex. Return (%)	10.01	9.96	10.40	10.34	10.07	10.00	10.57	10.15
σ_{Ret} (%)	0.43	0.41	0.46	0.72	0.39	0.36	0.43	0.68
Ann. Volatility (%)	15.42	15.50	15.00	14.95	15.32	15.42	14.82	14.92
σ_{Vol} (%)	0.5	0.5	0.52	0.57	0.40	0.40	0.42	0.53
Sharpe Ratio (%)	65.02	64.34	69.42	69.31	65.79	64.92	71.35	68.15
σ_{Sharpe} (%)	3.78	3.60	4.24	5.79	3.27	3.02	3.87	5.72
Turnover Rate (%)	6.18	5.72	9.65	20.32	6.41	5.86	11.28	24.44
σ_{TR} (%)	3.58	3.26	7.10	21.64	3.68	3.31	9.53	25.64

Notes: Nom, nominal. WC, worst-case. Ann, annualized. Ex, excess.

Table 3.6: Sharpe ratio analysis of 1,000 trials with $n = 15,\ 50,\ 75,\ 100$

	$n = 15$			$n = 50$		
	WC	Robust		WC	Robust	
		$\omega = 1.0$	$\omega = 2.0$		$\omega = 1.0$	$\omega = 2.0$
# of trials greater than nominal	681	983	983	229	1000	988
t-statistic	13.50	50.85	48.62	-24.29	92.61	82.42
	$n = 75$			$n = 100$		
	WC	Robust		WC	Robust	
		$\omega = 1.0$	$\omega = 2.0$		$\omega = 1.0$	$\omega = 2.0$
# of trials greater than nominal	87	1000	839	26	1000	697
t-statistic	-38.35	107.20	30.91	-46.68	123.92	16.66

Notes: WC, worst-case.

The risk concentration measures are presented in Table 3.7. The results show that, as n increases, the risk diversification of the portfolio deteriorates (with the exception of the nominal portfolio, which serves as our benchmark). With that said, the worst-case portfolios deviate the least from perfect risk diversifaction. This falls in line with our previous findings, where the cost of robustness manifests by deviating away from risk parity and increasing the turnover rate. In turn, this reflects our aversion to assets with noisier marginal risk contributions, limiting our capacity to attain perfect risk diversification.

Both experiments suggest that our proposed robust framework for risk parity portfolios can lead to higher out-of-sample risk-adjusted returns while maintaining a sufficiently risk-diverse portfolio. The trade-off between performance and risk diversification is expected of any investment model that differs from the nominal model. Moreover, our results suggest that simply taking the worst-case estimate of the covariance matrix does not deliver the robust behaviour expected during bear market periods. In particular, we observe that our proposed robust risk parity portfolios make most of their relative gains during periods of financial distress.

3.4 Conclusion

This chapter introduced a robust framework for risk parity portfolio optimization. Unlike MVO, risk parity does not seek to minimize risk. Conventional techniques in portfolio optimization, where robustness is introduced by assuming the 'worst-case' estimate of the covariance matrix,

Table 3.7: Summary of risk concentration measures of 1,000 trials with $n = 15,\ 50,\ 75,\ 100$

	$n = 15$				$n = 50$			
	Nom.	WC	Robust		Nom.	WC	Robust	
			$\omega = 1.0$	$\omega = 2.0$			$\omega = 1.0$	$\omega = 2.0$
CV	0.000	0.045	0.054	0.131	0.000	0.065	0.142	0.528
σ_{CV}	0.000	0.028	0.035	0.100	0.000	0.032	0.151	0.782
HRC	0.067	0.071	0.075	0.088	0.020	0.022	0.031	0.068
σ_{HRC}	0.000	0.003	0.006	0.019	0.000	0.001	0.014	0.097
H	0.067	0.067	0.067	0.068	0.020	0.020	0.021	0.037
σ_{H}	0.000	0.000	0.000	0.004	0.000	0.000	0.002	0.064
	$n = 75$				$n = 100$			
	Nom.	WC	Robust		Nom.	WC	Robust	
			$\omega = 1.0$	$\omega = 2.0$			$\omega = 1.0$	$\omega = 2.0$
CV	0.000	0.071	0.209	0.996	0.000	0.074	0.281	1.370
σ_{CV}	0.000	0.033	0.265	1.646	0.000	0.033	0.398	2.221
HRC	0.013	0.015	0.027	0.103	0.010	0.011	0.026	0.121
σ_{HRC}	0.000	0.001	0.019	0.183	0.000	0.001	0.025	0.216
H-index	0.013	0.013	0.015	0.062	0.010	0.010	0.012	0.077
σ_{H}	0.000	0.000	0.004	0.137	0.000	0.000	0.007	0.169

Notes: Nom, nominal. WC, worst-case. CV, coefficient of variation. HRC, highest risk contribution. H-index, Herfindahl index.

are not applicable when our objective is to equalize the asset risk contributions. Instead, due to the nature of our objective, we are most concerned with correctly estimating the asset risk contributions.

Our proposed framework was developed specifically to address uncertainty in risk parity. Thus, by design, this robust problem seeks to shield the portfolio from overexposure to assets with a high degree of estimation error in their marginal risk contributions. Intuitively, we define the uncertainty set on the asset covariance matrix. We then formulate the risk parity optimization problem such that it explicitly measures the assets' marginal risk contributions, allowing us to introduce robustness that limits our exposure to assets with noisier marginal risk contributions. As shown by the experimental results, our proposed framework is able to attain both a higher rate of return and a higher risk-adjusted rate of return, both of which can be attributed to a reduced exposure to riskier assets, which tend to drive portfolio losses.

Moreover, the proposed robust formulation relies on the insertion of a single new constraint upon the nominal risk parity SOCP. Thus, by design, the robust counterpart maintains the same level of complexity as the nominal problem. In other words, our robust SOCP is equally as tractable and efficient as the original risk parity problem.

Our framework accepts any uncertainty set derived from any acceptable estimation method that yields a set of box constraints on the covariance matrix. As an example, we showed how to derive this uncertainty set from the standard error of regression coefficients from a factor model of asset returns. Specifically, our numerical experiments used the Fama–French three-factor model.

The experimental results show that robust risk parity portfolios are able to outperform their nominal counterpart throughout a long investment horizon. As would be expected, the robust portfolios were able to realize a significant advantage during periods of market distress. More surprisingly, the robust portfolios were able to maintain their relative advantage during bullish periods. Aside from an improved rate of return, robust portfolios are also able to attain a higher risk-adjusted rate of return. However, the trade-off for this enhanced financial performance manifests in two manners: (i) by deviating away from perfect risk diversification, and (ii) by increasing our period-over-period turnover rate. This is the cost of robustness, and stems from our aversion to assets with noisier marginal risk contributions, in turn limiting our capacity to attain perfect risk diversification.

Chapter 4

Distributionally robust risk parity

As we discussed in Chapter 3, estimation errors may have a profound impact on a portfolio's ex post financial performance. The sensitivity of portfolio optimization to errors in estimated parameters has been widely explored in the literature [14, 20, 25, 73], leading to what is sometimes referred to as 'error maximization' given the poor out-of-sample performance of these (ex ante) optimal portfolios.

Accounting for uncertainty during optimization has become paramount in any decision-making problem where parameters are non-deterministic. If we assume we have complete knowledge of the underlying uncertainty (or a high degree of confidence in our estimate of the distribution), we can choose an appropriate probability distribution to represent the uncertainty of our parameters, leading to a collection of problems known as stochastic programs [15, 91]. On the other hand, when we have no distributional knowledge, we can ignore any distributional estimates and instead solve the worst-case instance of the problem and focus simply on retaining feasibility. This is the basis of robust optimization, which motivated our robust risk parity framework in Chapter 3.

Instead, this chapter is based on a class of problems that sits somewhere in-between stochastic programming and robust optimization. Such problems attempt to use distributional information during optimization, but accept that the underlying probability distribution is unknown. We assume this distribution lies within an ambiguity set of probability distributions.

Similar to robust optimization, we then take a worst-case approach, but with the distinction that we do this at the distributional level. Such a robust formulation for stochastic programs was proposed by Scarf [88]. Since then, this class of problems has often been referred to as

minimax problems or, more recently, as DRO problems [32].

The minimax problem has its roots in game theory [80]. In the context of this chapter, we seek to minimize our cost function with respect to our decision variable, while the secondary player, i.e., 'nature', is adversarial and seeks to maximize our cost with respect to our uncertain parameters. Thus, our true goal is to minimize our cost within the decision space against the most adversarial instance of the underlying distribution of the uncertain parameters. Minimax problems have been widely studied in literature in both theory and applications [19, 34, 90, 92, 100]. We note that minimax problems are sometimes referred to as saddle-point problems [61, 87] due to the 'saddle' shape of the cost function when viewed in the higher-dimensional space created by the decision variable and the uncertain parameters. In particular, we focus on the subset of well-behaved convex–concave saddle-point problems.

Specifically, this chapter introduces a distributionally robust counterpart of the convex risk parity problem in (2.19) proposed by Bai et al. [5]. As shown by Calafiore [21], distributional robustness can be introduced into the asset allocation problem by targeting the scenarios from which we derive our estimated parameters. When parameters are estimated from data, it is typically assumed that each scenario in the dataset is equally likely (i.e., we implicitly assume a uniform discrete probability distribution to describe the probability of each scenario). Instead, Calafiore [21] breaks this assumption and allows the scenarios to have different probabilities. In turn, this discrete probability distribution can be modelled as a set of decision variables, allowing us to design a maximization problem to find the most adversarial discrete probability distribution such that we attain the worst-case instance of the estimated parameters. Given that only the portfolio risk measure is pertinent for risk parity, this chapter focuses solely on the derivation of the asset covariance matrix from data.

Addressing distributional robustness through a discrete probability distribution aligns naturally with a data-driven parameter estimation process. This assumes that market efficiency holds and that raw market data suffices to accurately represent the set of possible future returns. More importantly, this avoids making any assumptions about the underlying probability distribution of the asset returns, as well as avoiding assigning a structured process (such as a factor model) to model the returns. Thus, we are not required to impose a structure on the raw market data, which fully exempts us from the risk of model misspecification. This follows a similar rationale to another popular scenario-based portfolio risk measure known as historical value-at-risk, which assumes that raw market data suffices to represent the set of possible future

outcomes. With that in mind, the application of a discrete probability distribution avoids the biases that could arise from assuming that the asset returns have a specific structure, and provides us with the flexibility to derive a robust estimate of the asset covariance matrix implied by the raw market data themselves.

The result is a distributionally robust risk parity (DRRP) problem with a discrete probability ambiguity set on the portfolio risk measure. We treat the discrete probability distribution as an adversarial player to formulate a minimax problem. Specifically, this minimax problem seeks to equalize the asset risk contributions against the worst-case instance of the portfolio variance.

The distributional ambiguity is modelled as a convex set. This convex set is defined by constraints corresponding to the axioms of probability and, in particular, by a constraint that bounds the statistical distance between a nominal (i.e., assumed) probability distribution and its adversarial counterpart. The nominal distribution can be defined as any reasonable discrete probability distribution, but this amounts to a uniform distribution when we assume that all scenarios are equally likely. Thus, the adversarial distribution is allowed to deviate from the 'equally likely' nominal distribution by a maximum permissible limit defined by our choice of statistical distance measure and our confidence level.

Given the conditions of our problem, we are limited to statistical distance measures for discrete probability distributions. The statistical distance measure used in [21] was the Kullback–Leibler (KL) divergence. However, the KL divergence is not a proper distance metric,[1] making it difficult for an investor to appropriately define this distance based on a given confidence level. Thus, this chapter focuses on statistical distance measures that satisfy the following two conditions: the measure must be a proper distance metric with finite bounds, and we must be able to formulate it as a computationally-tractable convex function. Therefore, the distributional ambiguity set is predominantly defined by our choice of distance measure and confidence level, which in turn defines our distributional robustness. For the purpose of this thesis, we limit ourselves to the following three statistical distance measures: the Jensen–Shannon (JS) divergence, Hellinger distance, and total variation (TV) distance. However, we note that our framework extends naturally to any finite statistical distance measure that can be represented as a convex function.

This chapter models the nominal risk parity problem as a convex minimization problem. In

[1] A metric or 'distance function' must be non-negative and satisfy the following axioms: symmetry, identity of indiscernibles, and the triangle inequality.

turn, the corresponding DRRP problem is a convex–concave minimax problem, where we seek to maximize our objective by finding the most adversarial instance of a discrete probability distribution. Our asset allocation variable is pragmatically constrained by the set of admissible portfolios, while the adversarial distribution is fundamentally constrained both by the axioms of probability and by the measure of statistical distance from the nominal distribution. Our modelling framework gives the user the flexibility to choose their preferred measure of statistical distance, provided this can be modelled as a convex function. The result is a constrained minimax problem, which we can solve with a projected gradient method [11]. Specifically, we present a projected gradient descent–ascent (PGDA) algorithm that alternates between the descent and ascent steps to reach the saddle point. The standard approaches to solve constrained minimax problems are projection-type methods [78, 98].

We proceed to introduce a novel algorithm to solve the DRRP minimax problem that is grounded in projected gradient descent and sequential convex programming. The standard PGDA algorithm requires that we take alternating descent and ascent steps as we move towards the saddle point of our problem. Such an approach typically requires double the number of design parameters when compared to algorithms that move in a single direction. Moreover, these design parameters must be defined by the user a priori (e.g., initial point, step sizes). Finally, iterating in two directions increases the possibility of numerical divergence.

Instead, we exploit the existence of a unique optimal risk parity portfolio for any given discrete probability distribution. Thus, our proposed algorithm operates iteratively through gradient ascent in the probability space while solving a risk parity minimization problem in the asset weight space after every iteration. We can interpret our proposed algorithm as an implementation of sequential convex programming (SCP), where we ascend in the probability space towards the most adversarial instance of the portfolio risk measure after every iteration using a projected gradient ascent (PGA) method. Thus, we refer to our proposed algorithm as SCP–PGA.

Compared to the PGDA algorithm, each iteration of the SCP–PGA algorithm is computationally more expensive. However, the exactness of each step translates to significantly fewer iterations until convergence. Additionally, we will see that the structure of the problem, combined with modern optimization software packages, allows for a computationally tractable and scalable implementation. Finally, from the user's perspective, this algorithm is conceptually easier to understand and implement: we iteratively ascend in the probability space while main-

taining the risk parity condition in the asset weight space.

As we will see in Section 4.3, our numerical experiments show that our SCP–PGA algorithm is computationally efficient and scales well for problems with a large number of assets and scenarios. Moreover, the in-sample experiments show that the DRRP problem behaves as expected, while the out-of-sample experiments demonstrate good ex post performance. Specifically, the DRRP portfolio is able to attain a higher risk-adjusted rate of return when compared to the nominal risk parity portfolio.

In summary, the contributions from this chapter are the following. First, we introduce the DRRP problem, which seeks risk parity with respect to the most adversarial estimate of the portfolio risk measure through a purely data-driven process (i.e, the probabilistic ambiguity is implied by the data themselves). Second, we explicitly define how to construct this ambiguity set using different statistical distance metrics and we show how to use an investor's confidence level to size the ambiguity set. Finally, we propose the SCP–PGA algorithm to solve the resulting DRRP minimax problem. We note that the flexible structure of the SCP–PGA algorithm means that it may be applied to solve other asset allocation problems, as well as other constrained convex–concave minimax problems from different disciplines.

4.1 Preliminaries

4.1.1 Estimation of parameters

We begin by restating and reinterpreting the measures of financial risk and reward that we previously discussed in Section 2.1.1. We aptly reframe the parameters specifically to serve the development of our DRRP problem. Moreover, we highlight the difference in notation that will be used in this chapter. For the purpose of clarity, many of the parameters and solutions in this chapter are defined using a functional notation, i.e., we define these output parameters in terms of some other input parameters. In turn, this will help to better understand the relationship between the inputs and outputs, and how changes in the former affect the latter.

As we saw in Section 2.1.1, we assume we can describe the random asset returns through the first two moments of their joint distribution, i.e., $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. If we have a dataset $\hat{\boldsymbol{\xi}} \in \mathbb{R}^{n \times T}$ consisting of T scenarios (where $\hat{\boldsymbol{\xi}}^t \in \mathbb{R}^n$ is the t^{th} scenario of the dataset $\hat{\boldsymbol{\xi}}$), we can derive the estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ as shown in (2.3–2.4). This process naturally assumes that each scenario in $\hat{\boldsymbol{\xi}}$ is equally likely.

Instead, assume there exists some probability p_t associated with each scenario t. In vector notation, this is the probability mass function (PMF) $\boldsymbol{p} \in \mathcal{P}$, where

$$\mathcal{P} \triangleq \left\{ \boldsymbol{p} \in \mathbb{R}_+^T : \mathbf{1}^\top \boldsymbol{p} = 1 \right\} \tag{4.1}$$

is the simplex defined by the axioms of probability. We note that the true distribution $f_{\boldsymbol{\xi}}$ is latent and typically assumed to be continuous. However, if we use a scenario-based estimation method we implicitly approximate such a distribution through a discrete counterpart. Thus, we shift our focus towards discrete probability distributions not because we believe the distribution of the random vector $\boldsymbol{\xi}$ is discrete, but rather because this aligns well with the parameter estimation process.

If we have knowledge of $\boldsymbol{p}$, then we can statistically derive the estimated parameters corresponding to the first two moments of the joint probability distribution of asset returns. Below we reintroduce the first two moments of the asset returns distribution, except we state them as functions of the probability distribution $\boldsymbol{p}$,

$$\hat{\boldsymbol{\mu}}(\boldsymbol{p}) \triangleq \mathbb{E}[\boldsymbol{\xi}] = \sum_{t=1}^{T} p_t \cdot \hat{\boldsymbol{\xi}}^t, \tag{4.2}$$

$$\hat{\boldsymbol{\Sigma}}(\boldsymbol{p}) \triangleq \mathbb{E}\big[\big(\boldsymbol{\xi} - \hat{\boldsymbol{\mu}}(\boldsymbol{p})\big)^2\big] = \sum_{t=1}^{T} p_t \cdot \big(\hat{\boldsymbol{\xi}}^t - \hat{\boldsymbol{\mu}}(\boldsymbol{p})\big)\big(\hat{\boldsymbol{\xi}}^t - \hat{\boldsymbol{\mu}}(\boldsymbol{p})\big)^\top, \tag{4.3}$$

where $\hat{\boldsymbol{\mu}}(\boldsymbol{p}) \in \mathbb{R}^n$ and $\hat{\boldsymbol{\Sigma}}(\boldsymbol{p}) \in \mathbb{S}_+^n$ are the data-driven estimates of the latent parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively. These two estimates are shown as functions of some discrete probability distribution $\boldsymbol{p}$. Thus, $\hat{\boldsymbol{\mu}}(\boldsymbol{p})$ and $\hat{\boldsymbol{\Sigma}}(\boldsymbol{p})$ are generalizations of the nominal estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ previously shown in (2.3–2.4). In other words, if we assume each scenario is equally likely, then (4.2) and (4.3) are simply the standard sample arithmetic mean and sample covariance matrix[2] from (2.3–2.4). Finally, we note that $\hat{\boldsymbol{\Sigma}}(\boldsymbol{p})$ in (4.3) is the result of the weighted sum of T rank-1 symmetric matrices, meaning $\hat{\boldsymbol{\Sigma}}(\boldsymbol{p})$ is guaranteed to be a PSD matrix, and is very likely to be positive definite if $T >> n$.

We will constrain our portfolios $\boldsymbol{x}$ to the set of admissible portfolios $\mathcal{X}$, which, in our case,

[2]We note that $\hat{\boldsymbol{\Sigma}}(\boldsymbol{p})$, where $p_t = 1/T$ for $t = 1, \ldots, T$ is the 'equally likely' nominal distribution, yields the standard scenario-based estimate of the covariance matrix. However, if we wish to recover the standard *unbiased* estimate of the covariance matrix shown in (2.4), we should multiply $\hat{\boldsymbol{\Sigma}}(\boldsymbol{p})$ by $T/(T-1)$. For the purpose of this chapter, we will see that this distinction has no impact in the formulation of our distributionally robust optimization problem.

disallows short sales and imposes a unit budget constraint. Short sales are prohibited to address the non-convexity of risk parity discussed in Section 2.2.1. Doing this will allow us to use a convex formulation of the nominal risk parity problem to develop the corresponding DRRP problem. It follows that the set of admissible portfolios is the following simplex

$$\mathcal{X} \triangleq \big\{\boldsymbol{x} \in \mathbb{R}^n_+ : \mathbf{1}^\top \boldsymbol{x} = 1\big\}. \tag{4.4}$$

The equality constraint in $\mathcal{X}$ ensures that the entirety of our available budget is invested in the assets.

In section 2.1.1, we saw that the portfolio random return is $\pi = \boldsymbol{\xi}^\top \boldsymbol{x}$. Thus, if we have the asset return data $\hat{\boldsymbol{\xi}}$ and a portfolio $\boldsymbol{x}$, the corresponding observed portfolio returns are $\hat{\boldsymbol{\pi}}(\boldsymbol{x}) = \hat{\boldsymbol{\xi}}^\top \boldsymbol{x} \in \mathbb{R}^T$. Accordingly, the estimated portfolio expected return and variance are functions of the probability distribution $\boldsymbol{p}$ and the portfolio $\boldsymbol{x}$. We previously saw how to calculate the estimated portfolio expected return and variance in (2.5–2.6) if we assume that all scenarios are equally likely. However, we can generalize and express both parameters as functions of $\boldsymbol{p}$ and $\boldsymbol{x}$. This functional notation will aid in our development of the DRRP problem throughout this chapter.

It follows that the estimated portfolio expected return and variance are

$$\begin{aligned}
\hat{\mu}_\pi(\boldsymbol{x}, \boldsymbol{p}) &\triangleq \boldsymbol{x}^\top \hat{\boldsymbol{\mu}}(\boldsymbol{p}) && \text{(4.5a)}\\
&= \boldsymbol{p}^\top \hat{\boldsymbol{\pi}} && \text{(4.5b)}\\
\hat{\sigma}^2_\pi(\boldsymbol{x}, \boldsymbol{p}) &\triangleq \boldsymbol{x}^\top \hat{\boldsymbol{\Sigma}}(\boldsymbol{p}) \boldsymbol{x} && \text{(4.6a)}\\
&= \mathbb{E}\Big[\big(\pi - \mathbb{E}[\pi]\big)^2\Big] = \mathbb{E}[\pi^2] - \big(\mathbb{E}[\pi]\big)^2 \\
&= \boldsymbol{p}^\top \hat{\boldsymbol{\pi}}^{\circ 2}(\boldsymbol{x}) - \boldsymbol{p}^\top \hat{\boldsymbol{\Gamma}}(\boldsymbol{x}) \boldsymbol{p}, && \text{(4.6b)}
\end{aligned}$$

where $\hat{\boldsymbol{\pi}}^{\circ 2}(\boldsymbol{x}) \in \mathbb{R}^T_+$ denotes the element-wise square of the vector of portfolio return scenarios, and $\hat{\boldsymbol{\Gamma}}(\boldsymbol{x}) \triangleq \hat{\boldsymbol{\pi}}(\boldsymbol{x}) \hat{\boldsymbol{\pi}}(\boldsymbol{x})^\top$. By definition, we have that $\hat{\boldsymbol{\Gamma}}(\boldsymbol{y}) \in \mathbb{S}^T_+$ for any vector $\boldsymbol{y} \in \mathbb{R}^n$, meaning the portfolio variance in (4.6b) is concave over $\boldsymbol{p} \in \mathcal{P}$. Moreover, since $\hat{\boldsymbol{\Sigma}}(\boldsymbol{p}) \in \mathbb{S}^n_+$ for any probability distribution $\boldsymbol{p} \in \mathcal{P}$, the portfolio variance in (4.6a) is convex over $\boldsymbol{x} \in \mathcal{X}$.

The convexity over $\boldsymbol{x} \in \mathcal{X}$ and concavity over $\boldsymbol{p} \in \mathcal{P}$ of the portfolio variance $\hat{\sigma}^2_\pi(\boldsymbol{x}, \boldsymbol{p})$ will allow us to formulate a convex–concave DRRP minimax problem. Working with the portfolio variance aligns well with the convex risk parity problem introduced by Bai et al. [5] and shown

in (2.19). For the development of the DRRP problem, we restate this convex risk parity problem as a function of $\boldsymbol{p} \in \mathcal{P}$ through the following system of equations,

$$f_{\mathrm{RP}}(\boldsymbol{y}, \boldsymbol{p}) \triangleq \frac{1}{2}\boldsymbol{y}^\top \hat{\boldsymbol{\Sigma}}(\boldsymbol{p})\boldsymbol{y} - \kappa \sum_{i=1}^{n} \ln(y_i), \tag{4.7a}$$

$$\boldsymbol{y}^{\mathrm{RP}}(\boldsymbol{p}) \triangleq \underset{\boldsymbol{y} \in \mathbb{R}^n_+}{\operatorname{argmin}} \; f_{\mathrm{RP}}(\boldsymbol{y}, \boldsymbol{p}), \tag{4.7b}$$

$$\boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p}) \triangleq \Pi_{\mathcal{X}}\big(\boldsymbol{y}^{\mathrm{RP}}(\boldsymbol{p})\big), \tag{4.7c}$$

where $\boldsymbol{y} \in \mathbb{R}^n_+$ is a placeholder for the vector of asset weights, $f_{\mathrm{RP}} : \mathbb{R}^n_+ \times \mathcal{P} \to \mathbb{R}$ is our risk parity objective function, and $\Pi_{\mathcal{X}}\big(\boldsymbol{y}^{\mathrm{RP}}(\boldsymbol{p})\big)$ is the projection of the resulting vector $\boldsymbol{y}^{\mathrm{RP}}(\boldsymbol{p})$ onto the set of admissible portfolios $\mathcal{X}$. The logarithmic barrier in (4.7a) ensures that, at optimality, we converge to an optimal solution $\boldsymbol{y}^{\mathrm{RP}}(\boldsymbol{p}) \in \mathbb{R}^n_+$. Therefore, the projection onto the set $\mathcal{X}$ only needs to enforce the budget equality constraint. Finally, the risk parity portfolio for an arbitrary instance of $\boldsymbol{p} \in \mathcal{P}$ is the vector $\boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p}) \in \mathbb{R}^n_+$.

For some arbitrary vector $\boldsymbol{z} \in \mathbb{R}^n_+$, the projection onto the set $\mathcal{X}$ is

$$\Pi_{\mathcal{X}}(\boldsymbol{z}) \triangleq \frac{\boldsymbol{z}}{\sum_{i=1}^{n} z_i}. \tag{4.8}$$

We conclude this subsection by highlighting that we can use the optimization problem and projection in (4.7) to find the optimal risk parity portfolio for any estimate of the covariance matrix $\hat{\boldsymbol{\Sigma}}(\boldsymbol{p})$ with respect to an arbitrary instance of $\boldsymbol{p} \in \mathcal{P}$.

4.1.2 Distributionally robust optimization

Here we present a brief overview of DRO and its applications to the asset allocation problem. Our decision variable is the vector of asset weights, $\boldsymbol{x} \in \mathcal{X}$. Additionally, we have the vector of asset random returns $\boldsymbol{\xi} \in \mathbb{R}^n$. Let us momentarily depart from the notion of risk parity, and assume our objective is defined by some generic cost function $f(\boldsymbol{x}, \boldsymbol{\xi})$. We can define following the optimization problem to minimize the expected value of our cost function

$$\min_{\boldsymbol{x} \in \mathcal{X}} \quad \mathbb{E}_{\boldsymbol{\xi}}\big[f(\boldsymbol{x}, \boldsymbol{\xi})\big].$$

Given that this optimization problem involves a decision variable $\boldsymbol{x}$ and a random variable $\boldsymbol{\xi}$, this generic problem is a stochastic program. Note that the function $f(\boldsymbol{x}, \boldsymbol{\xi})$ may embed both measures of financial risk and reward. However, formulating our problem in this fashion assumes we have perfect knowledge of the probability density function $f_{\boldsymbol{\xi}}$, which is not true in general. Instead, we can introduce distributional robustness to protect ourselves against plausible adversarial forms of $f_{\boldsymbol{\xi}}$.

We can do this by formulating a minimax stochastic program [88, 92]. By design, the minimax problem seeks to minimize our objective with respect to our decision variable $\boldsymbol{x}$ while simultaneously defining the most adversarial form of the objective with respect to $f_{\boldsymbol{\xi}}$,

$$\min_{\boldsymbol{x} \in \mathcal{X}} \max_{f_{\boldsymbol{\xi}} \in \mathcal{U}_f} \quad \mathbb{E}_{\boldsymbol{\xi}}\big[f(\boldsymbol{x}, \boldsymbol{\xi})\big] \tag{4.9}$$

where, in this example, $\mathcal{U}_f$ is a set of plausible probability measures. This type of minimax problem is sometimes defined as DRO [24, 82, 32].

Unlike the generic problem in (4.9), we restrict ourselves to discrete probability distributions to align with traditional data-driven estimation procedures. Thus, for the purpose of this chapter, we assume that the PMF $\boldsymbol{p} \in \mathcal{P}$ is associated with the scenarios that describe the finite set of possible outcomes of our asset returns $\boldsymbol{\xi}$. In turn, our minimax problem becomes

$$\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{p} \in \mathcal{U}_{\boldsymbol{p}}} \quad \mathbb{E}_{\boldsymbol{\xi}}\big[f(\boldsymbol{x}, \boldsymbol{\xi})\big] \tag{4.10}$$

where $\mathcal{U}_{\boldsymbol{p}}$ is the ambiguity set of plausible discrete probability distributions. In addition to complying with the axioms of probability, the set $\mathcal{U}_{\boldsymbol{p}}$ includes the distributional ambiguity of $\boldsymbol{p}$. In a similar fashion to Calafiore [21], we define this ambiguity relative to some nominal distribution $\boldsymbol{q} \in \mathcal{P}$. For example, we typically assume that every scenario in a dataset is equally likely. In other words, we have that $q_t = 1/T$ for $t = 1, \ldots, T$. In this case, if we use $\boldsymbol{q}$ as an input in (4.2) and (4.3), then we recover the typical estimated expected returns $\hat{\boldsymbol{\mu}}(\boldsymbol{q})$ and covariance matrix[3] $\hat{\boldsymbol{\Sigma}}(\boldsymbol{q})$. As we will see next, we can use a measure of statistical distance to limit the ambiguity of $\boldsymbol{p}$.

[3]We note that if $q_t = 1/T$, then $\hat{\boldsymbol{\Sigma}}(\boldsymbol{q})$ is the *biased* estimate of the covariance matrix.

4.1.3 Statistical distance measures

Statistical distances can be used to quantify the similarity between two probability distributions. We limit our choice of statistical distance measures to a subset of convex functions that operate on discrete distributions. As we move forward into Section 4.2, this will allow us to retain computational tractability. Specifically, we will discuss a non-exhaustive list of five different statistical distance measures that are applicable to our framework.

The first distance measure we discuss is the KL divergence [63], sometimes referred to as 'relative entropy'. The KL divergence is not a formal distance metric since it is not symmetric and does not respect the triangle inequality. Nevertheless, the KL divergence has become an increasingly popular tool to measure the difference between two probability distributions. For two discrete probability distributions $\boldsymbol{p},\ \boldsymbol{q} \in \mathcal{P}$, the KL divergence is defined as

$$G_{\mathrm{KL}}(\boldsymbol{p}, \boldsymbol{q}) \triangleq \sum_{t=1}^{T} p_t \ln\left(\frac{p_t}{q_t}\right). \tag{4.11}$$

The asymmetry of the KL divergence becomes apparent if we reverse the order of the arguments $\boldsymbol{p}$ and $\boldsymbol{q}$ (i.e., $G_{\mathrm{KL}}(\boldsymbol{p}, \boldsymbol{q}) \neq G_{\mathrm{KL}}(\boldsymbol{q}, \boldsymbol{p})$). By definition, the KL divergence is always non-negative. However, the upper bound of the KL divergence is not properly defined, making it difficult to define an appropriate permissible distance between $\boldsymbol{p}$ and $\boldsymbol{q}$. In spite of these issues, the KL divergence is a useful tool to constrain the divergence between our nominal distribution $\boldsymbol{q}$ and our adversarial distribution $\boldsymbol{p}$. An example of its application to the asset allocation problem is shown in [21].

The KL divergence can be made symmetric by averaging its forward and reverse forms. We can define the symmetric KL (SKL) divergence as

$$G_{\mathrm{SKL}}(\boldsymbol{p}, \boldsymbol{q}) \triangleq \frac{1}{2} G_{\mathrm{KL}}(\boldsymbol{p}, \boldsymbol{q}) + \frac{1}{2} G_{\mathrm{KL}}(\boldsymbol{q}, \boldsymbol{p}). \tag{4.12}$$

Although the SKL divergence is symmetric, it still does not respect the triangle inequality and cannot be defined as a true metric. Moreover, it lacks a properly defined upper bound.

A measure closely related to the KL divergence is the JS divergence, which was introduced by Lin [67]. Unlike the KL divergence, the JS divergence is symmetric and has finite bounds.

The JS divergence is defined as

$$G_{\mathrm{JS}}(\boldsymbol{p},\boldsymbol{q}) \triangleq \frac{1}{2}G_{\mathrm{KL}}(\boldsymbol{p},\boldsymbol{m}) + \frac{1}{2}G_{\mathrm{KL}}(\boldsymbol{q},\boldsymbol{m}), \tag{4.13}$$

where $\boldsymbol{m} = \frac{1}{2}(\boldsymbol{p}+\boldsymbol{q}) \in \mathcal{P}$. Given that our definition of the KL divergence in (4.11) uses the natural logarithm, our definition of the JS divergence has the useful property of being bounded between zero and ln(2), i.e.,

$$0 \leq G_{\mathrm{JS}}(\boldsymbol{p},\boldsymbol{q}) \leq \ln(2).$$

Additionally, we can take the square root of the JS divergence to derive a formal distance metric known as the JS distance [37, 43] (i.e., the JS distance is $\sqrt{G_{\mathrm{JS}}(\boldsymbol{p},\boldsymbol{q})}$). This distance measure is bounded between zero and $\sqrt{\ln(2)}$. As we will see in Section 4.2, the finite bounds on the JS distance will allow us to properly define the permissible distance d between the nominal distribution $\boldsymbol{q}$ and its adversarial counterpart $\boldsymbol{p}$.

Next, we discuss the Hellinger distance,

$$G_{\mathrm{H}}(\boldsymbol{p},\boldsymbol{q}) \triangleq \frac{1}{\sqrt{2}}\sqrt{\sum_{t=1}^{T}\left(\sqrt{p_t}-\sqrt{q_t}\right)^2}. \tag{4.14}$$

The Hellinger distance is a formal distance metric and, by its definition, is bounded between zero and one. Similarly to the JS distance, this will allow us to define a permissible distance d between $\boldsymbol{q}$ and $\boldsymbol{p}$.

The last distance measure we discuss is the TV distance,

$$G_{\mathrm{TV}}(\boldsymbol{p},\boldsymbol{q}) \triangleq \frac{1}{2}\|\boldsymbol{p}-\boldsymbol{q}\|_1, \tag{4.15}$$

which is also a formal distance metric. The TV distance is bounded between zero and one. As with the JS and Hellinger distances, its formal definition as a bounded metric will allow us to define a permissible distance d between $\boldsymbol{q}$ and $\boldsymbol{p}$.

4.2 Distributionally robust risk parity

This section presents our two main contributions: a data-driven DRRP problem and the SCP–PGA algorithm to solve it. The 'data-driven' aspect arises from using a discrete probability

distribution to assign different weights to our scenarios. Thus, we can emphasize the scenarios that are more likely to adversely impact our estimate of the covariance matrix.

Consider the minimax problem from (4.10) and recall the set of plausible discrete probability distributions $\mathcal{U}_{\boldsymbol{p}}$. Our immediate goal is twofold: to design an appropriate ambiguity set $\mathcal{U}_{\boldsymbol{p}}$ for our adversarial probability distribution $\boldsymbol{p}$, and to recast the generic minimax problem from (4.10) into our DRRP problem. We address these two issues in the following two subsections, before proceeding into the algorithmic development. Finally, we will conclude this section by discussing a variant of the risk parity problem where an investor can incorporate estimated expected returns into the optimization problem.

4.2.1 Probability distribution ambiguity set

Our adversarial probability distribution $\boldsymbol{p}$ belongs to the ambiguity set $\mathcal{U}_{\boldsymbol{p}}$, which we will now formally define. A probability distribution must adhere to the simplex $\mathcal{P}$ defined by the axioms of probability. The ambiguity set $\mathcal{U}_{\boldsymbol{p}}$ is defined as the set of probability distributions that lie within a maximum permissible distance d from a predetermined nominal distribution $\boldsymbol{q}$. Thus, the ambiguity set is

$$\mathcal{U}_{\boldsymbol{p}}(\boldsymbol{q}, d) \triangleq \big\{\boldsymbol{p} \in \mathcal{P} : \psi(\boldsymbol{p}, \boldsymbol{q}) \leq d\big\} \tag{4.16}$$

where $\psi(\boldsymbol{p}, \boldsymbol{q})$ is some convex function that describes any of the statistical distance measures defined in (4.11–4.15), while $d \in \mathbb{R}_+$ is a user-defined bound on the maximum distance between $\boldsymbol{p}$ and $\boldsymbol{q}$. By definition, $\mathcal{U}_{\boldsymbol{p}} \subseteq \mathcal{P}$.

For the purpose of computational tractability, our five statistical distance measures are restated or, where appropriate, reformulated as follows,

$$\psi_{\mathrm{KL}}(\boldsymbol{p}, \boldsymbol{q}) \triangleq \sum_{t=1}^{T} p_t \ln\left(\frac{p_t}{q_t}\right), \tag{4.17a}$$

$$\psi_{\mathrm{SKL}}(\boldsymbol{p}, \boldsymbol{q}) \triangleq \frac{1}{2}\sum_{t=1}^{T} (p_t - q_t) \ln\left(\frac{p_t}{q_t}\right), \tag{4.17b}$$

$$\psi_{\mathrm{JS}}(\boldsymbol{p}, \boldsymbol{q}) \triangleq \frac{1}{2}\sum_{t=1}^{T} p_t \ln(p_t) + q_t \ln(q_t) - (p_t + q_t) \ln\left(\frac{p_t + q_t}{2}\right), \tag{4.17c}$$

$$\psi_{\mathrm{H}}(\boldsymbol{p}, \boldsymbol{q}) \triangleq \frac{1}{2}\sum_{t=1}^{T} p_t - 2\sqrt{p_t}\sqrt{q_t} + q_t, \tag{4.17d}$$

$$\psi_{\mathrm{TV}}(\boldsymbol{p}, \boldsymbol{q}) \triangleq \frac{1}{2}\|\boldsymbol{p} - \boldsymbol{q}\|_1. \tag{4.17e}$$

We make the distinction in notation between $\psi(\boldsymbol{p}, \boldsymbol{q})$ in (4.17) and $G(\boldsymbol{p}, \boldsymbol{q})$ in (4.11–4.15) because, in some cases, the distance measure $G(\boldsymbol{p}, \boldsymbol{q})$ was not stated in a computationally-tractable form. In particular, we have simplified the expressions of the SKL and JS divergences, and we have squared the Hellinger distance. The distance measures in (4.17) are convex functions over $\boldsymbol{p} \in \mathcal{P}$ for any $\boldsymbol{q} \in \mathcal{P}$. Since $\mathcal{P}$ is the probability simplex, this means that the ambiguity set $\mathcal{U}_{\boldsymbol{p}}(\boldsymbol{q}, d)$ is also convex. Moreover, with the exception of (4.17d) and (4.17e), these definitions of distance can be directly implemented and solved by any modern non-linear optimization software package. The functions in (4.17d) and (4.17e) can be computationally implemented by introducing auxiliary variables during optimization, and doing this does not fundamentally alter the problem. An example of how to computationally implement them is shown in Appendix A.

In Section 4.1.2, we defined the nominal probability distribution as $q_t = 1/T$ for $t = 1, \ldots, T$, which is merely a probabilistic representation of the assumption that each scenario in our dataset is equally likely. Our modelling framework provides sufficient flexibility for the user to prescribe their own choice of $\boldsymbol{q} \in \mathcal{P}$. However, we formally reiterate our definition of the nominal probability distribution as a discrete uniform distribution, i.e.,

$$\boldsymbol{q} \triangleq \begin{bmatrix} 1/T \\ \vdots \\ 1/T \end{bmatrix} \in \mathbb{R}^T. \tag{4.18}$$

This falls in line with our goal to define the most adversarial distribution $\boldsymbol{p}$ relative to the distribution implied by the data.

To finalize the definition of $\mathcal{U}_{\boldsymbol{p}}$, we must determine the value of the maximum permissible distance d based on the investor's confidence level. Given that the KL divergence in (4.17a) and SKL divergence in (4.17b) are not formal statistical distance metrics, we will not discuss how to appropriately determine d for these two measures. However, our modelling framework will retain sufficient flexibility in case the user is able to specify an appropriate maximum permissible distance d for the KL or SKL divergences.

The remainder of this subsection describes how to define an appropriate value for d for the JS, Hellinger and TV measures. As discussed in Section 4.1.3, these three distance measures have theoretical lower and upper bounds. In particular, the upper bounds are only attainable if the nominal distribution $\boldsymbol{q}$ differs the most from our adversarial distribution $\boldsymbol{p}$. For a discrete

probability distribution, this happens when both the nominal and adversarial distributions assign a probability $q_i = p_j = 1$ for scenarios $i \neq j$, with all other scenarios having a probability of zero. In practice, the upper bounds presented in Section 4.1.3 are unattainable under the assumption that $\boldsymbol{q}$ is a discrete uniform distribution. Consider the following example of an extreme probability distribution

$$\boldsymbol{w} \triangleq \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^T,$$

which assigns all of its weight to a single scenario. The distribution $\boldsymbol{w}$ is the most we can differ from the uniform distribution $\boldsymbol{q}$. Thus, in practice, the true upper bound is defined as $K(T) \triangleq \psi(\boldsymbol{w}, \boldsymbol{q}) \in \mathbb{R}_+$, where the argument T corresponds to the dimension of the fixed distributions $\boldsymbol{w}$ and $\boldsymbol{q}$. We define the practical upper bounds of our three distance measures as

$$K_{\mathrm{JS}}(T) = \psi_{\mathrm{JS}}(\boldsymbol{w}, \boldsymbol{q}), \tag{4.19a}$$

$$K_{\mathrm{H}}(T) = \psi_{\mathrm{H}}(\boldsymbol{w}, \boldsymbol{q}), \tag{4.19b}$$

$$K_{\mathrm{TV}}(T) = \psi_{\mathrm{TV}}(\boldsymbol{w}, \boldsymbol{q}). \tag{4.19c}$$

For example, if our data consist of ten scenarios, $T = 10$, then the upper bounds of the JS divergence in (4.17c), the squared Hellinger distance in (4.17d), and the TV distance in (4.17e) are

$$\begin{aligned} K_{\mathrm{JS}}(10) &= \frac{1}{2}\Big((0.1)\ln(0.1) - (1.1)\ln(0.55) + (9)(0.1)\ln(2)\Big) \approx 0.5256, \\ K_{\mathrm{H}}(10) &= \frac{1}{2}\Big(1 - 2\sqrt{0.1} + (10)(0.1)\Big) \approx 0.6838, \\ K_{\mathrm{TV}}(10) &= \frac{1}{2}\Big(0.9 + (9)(0.1)\Big) = 0.9. \end{aligned}$$

Furthermore, as T increases, the upper bounds approach their theoretical values (i.e., as $T \to \infty$, we have $K_{\mathrm{JS}} \to \ln(2)$, $K_{\mathrm{H}} \to 1$ and $K_{\mathrm{TV}} \to 1$). The purpose of this exercise is to avoid defining d in terms of a theoretical upper bound. Instead, we seek a more pragmatic definition based on the number of scenarios in our dataset.

In Section 4.1.3 we saw that the square root of the JS divergence is a true distance metric. We can exploit this property to define an appropriate maximum permissible distance d_{JS} between our nominal and adversarial distributions based on the upper bound K_{JS} and the user-defined

confidence level $0 \leq \delta \leq 1$. In turn, we can use this to constrain the statistical distance between $\boldsymbol{p}$ and $\boldsymbol{q}$ (i.e., $\psi_{\mathrm{JS}}(\boldsymbol{p}, \boldsymbol{q}) \leq d_{\mathrm{JS}}$). Recall that we must square the investor's confidence level since the JS divergence is the square of the JS distance. Thus, for a given confidence level δ and number of scenarios T, the maximum permissible distance is

$$d_{\mathrm{JS}}(\delta, T) \triangleq \delta^2 K_{\mathrm{JS}}(T). \tag{4.20}$$

Similarly, since $\psi_{\mathrm{H}}(\boldsymbol{p}, \boldsymbol{q})$ in (4.17d) is defined as the square of the Hellinger distance, the appropriate maximum permissible distance is

$$d_{\mathrm{H}}(\delta, T) \triangleq \delta^2 K_{\mathrm{H}}(T). \tag{4.21}$$

Finally, given that the TV distance in (4.17e) is already a true metric of statistical distance, we can define the maximum permissible distance as

$$d_{\mathrm{TV}}(\delta, T) \triangleq \delta K_{\mathrm{TV}}(T). \tag{4.22}$$

To properly define the ambiguity set $\mathcal{U}_{\boldsymbol{p}}$ in (4.16), the user must choose their preferred distance measure and define $\psi(\boldsymbol{p}, \boldsymbol{q})$ as one of the options in (4.17), while defining d accordingly. In particular, if the selected distance measure is either the JS divergence, Hellinger distance or TV distance, then d can be defined from the options shown in (4.20–4.22).

4.2.2 Minimax problem

For a given dataset $\hat{\boldsymbol{\xi}}$, our problem is defined by the investor's choice of statistical distance measure (e.g., JS or TV) and confidence level δ. Given this information, we aim to construct an optimal DRRP portfolio $\boldsymbol{x}^*$. The nominal risk parity problem in (4.7) is strictly convex for any given estimate of the covariance matrix $\hat{\boldsymbol{\Sigma}}(\boldsymbol{p})$. Therefore, as discussed in Section 2.2, there exists a unique risk parity portfolio $\boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p})$ for every instance of $\boldsymbol{p} \in \mathcal{P}$.

The distinction between $\boldsymbol{x}^*$ and $\boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p})$ is the following. The latter is the optimal risk parity portfolio for a given instance of $\boldsymbol{p} \in \mathcal{P}$, as shown in (4.7). On the other hand, we use $\boldsymbol{x}^*$ to denote the portfolio resulting from the most adversarial instance of $\boldsymbol{p} \in \mathcal{U}_{\boldsymbol{p}}$ such that it maximizes our risk parity objective function. Thus, our optimal DRRP portfolio $\boldsymbol{x}^*$ can be formulated as a minimax problem where we seek an optimal portfolio against an optimally

adversarial discrete probability distribution.

The risk parity problem in (4.7) requires that we first optimize an unconstrained problem and then project it onto the set of admissible portfolios. However, for simplicity, let us ignore the projection step and treat the unconstrained auxiliary variable $\boldsymbol{y}$ as a proxy[4] for our asset weights $\boldsymbol{x}$. Thus, for now, let the variables of our minimax problem be $\boldsymbol{y}$ and $\boldsymbol{p}$.

Recall our original definition of the portfolio variance, which was expressed in two equivalent forms in (4.6a) and (4.6b). Moreover, recall our original definition of the risk parity objective function $f_{\text{RP}}(\boldsymbol{y}, \boldsymbol{p})$ in (4.7a). Using both expressions of the portfolio variance, we can restate our risk parity objective function in two equivalent forms

$$f_{\text{RP}}(\boldsymbol{y}, \boldsymbol{p}) \triangleq \frac{1}{2}\boldsymbol{y}^\top \hat{\boldsymbol{\Sigma}}(\boldsymbol{p})\boldsymbol{y} - \kappa \sum_{i=1}^{n} \ln(y_i) \tag{4.23a}$$

$$\triangleq \frac{1}{2}\left(\boldsymbol{p}^\top \hat{\boldsymbol{\pi}}^2(\boldsymbol{y}) - \boldsymbol{p}^\top \hat{\boldsymbol{\Gamma}}(\boldsymbol{y})\boldsymbol{p}\right) - \kappa \sum_{i=1}^{n} \ln(y_i), \tag{4.23b}$$

where (4.23a) is exactly the same as (4.7a) and is restated for clarity, while (4.23b) presents $f_{\text{RP}}(\boldsymbol{y}, \boldsymbol{p})$ explicitly in terms of $\boldsymbol{p}$. Formulating the objective function in these two equivalent forms allows us to observe how the function acts upon both the decision variable $\boldsymbol{y}$ and the adversarial probability $\boldsymbol{p}$.

The corresponding DRRP problem, stated as a minimax problem, is the following

$$\min_{\boldsymbol{y} \in \mathbb{R}^n_+} \max_{\boldsymbol{p} \in \mathcal{U}_{\boldsymbol{p}}} \quad f_{\text{RP}}(\boldsymbol{y}, \boldsymbol{p}). \tag{4.24}$$

As we saw in Section 4.1.1, both $\hat{\boldsymbol{\Sigma}}(\boldsymbol{p})$ and $\hat{\boldsymbol{\Gamma}}(\boldsymbol{y})$ are PSD for any $\boldsymbol{p} \in \mathcal{P}$ and $\boldsymbol{y} \in \mathbb{R}^n_+$, respectively. Therefore, the function $f_{\text{RP}}(\cdot, \boldsymbol{p}) : \mathbb{R}^n_+ \to \mathbb{R}$ is strictly convex for every $\boldsymbol{p} \in \mathcal{P}$, while $f_{\text{RP}}(\boldsymbol{y}, \cdot) : \mathcal{P} \to \mathbb{R}$ is concave for every $\boldsymbol{y} \in \mathbb{R}^n_+$. Moreover, the sets $\mathcal{X}$ and $\mathcal{U}_{\boldsymbol{p}}$ are convex. This means we have a convex–concave minimax problem, which means that any local optimum is a global optimum. In turn, this leads to the following observation

$$f_{\text{RP}}(\boldsymbol{y}^*, \boldsymbol{p}) \leq f_{\text{RP}}(\boldsymbol{y}^*, \boldsymbol{p}^*) \leq f_{\text{RP}}(\boldsymbol{y}, \boldsymbol{p}^*) \;\; \forall\, \boldsymbol{y} \in \mathbb{R}^n_+,\; \boldsymbol{p} \in \mathcal{U}_{\boldsymbol{p}}, \tag{4.25}$$

where $(\boldsymbol{y}^*, \boldsymbol{p}^*)$ is the optimal solution (i.e., the saddle point) of $f_{\text{RP}}(\boldsymbol{y}, \boldsymbol{p})$.

[4] Projecting the auxiliary variable $\boldsymbol{y} \in \mathbb{R}^n_+$ onto the set of of admissible portfolios such that we find a portfolio $\boldsymbol{x}$ is a trivial step, as shown in (4.8).

The maximization step in (4.24) is also meaningful in a financial context. Consider the definition of $f_{\mathrm{RP}}(\boldsymbol{y}, \boldsymbol{p})$ in (4.23b) where, without loss of generality, we have defined the portfolio variance using the unnormalized proxy variable $\boldsymbol{y}$. The maximization step in (4.24) pertains solely to the portfolio variance given that the logarithmic barrier term only acts on the variable $\boldsymbol{y}$. Thus, intuitively, the maximization step aims to find the most adversarial probability distribution $\boldsymbol{p}$ such that we attain the worst-case instance of the portfolio variance. This leads to the following conclusion: the minimax problem in (4.24) seeks the optimal risk parity portfolio with respect to the worst-case portfolio variance.

4.2.3 Projected gradient descent–ascent

We can exploit the convex–concave structure of our minimax problem to attain global optimality through a gradient-based algorithm. In particular, we discuss a PGDA algorithm that sequentially alternates between descending in $\boldsymbol{y}$ and ascending in $\boldsymbol{p}$ until convergence.

To retain feasibility after each iteration, we project each step in $\boldsymbol{y}$ and $\boldsymbol{p}$ onto the sets $\mathbb{R}^n_+$ and $\mathcal{U}_{\boldsymbol{p}}$, respectively. In particular, the non-linearity of the statistical distance measure means that the projection onto the ambiguity set $\mathcal{U}_{\boldsymbol{p}}$ is non-trivial and cannot be solved in closed form. Instead, the projection must be solved as a constrained optimization problem. A Euclidean projection ensures the problem is strictly convex, guaranteeing the uniqueness of our solution. We define the projection of some arbitrary vector $\boldsymbol{u} \in \mathbb{R}^T$ onto the set $\mathcal{U}_{\boldsymbol{p}}$ as follows,

$$\Pi_{\mathcal{U}_{\boldsymbol{p}}}(\boldsymbol{u}) \triangleq \underset{\boldsymbol{p}}{\operatorname{argmin}} \quad \|\boldsymbol{u} - \boldsymbol{p}\|_2^2 \tag{4.26a}$$

$$\text{s.t.} \qquad \mathbf{1}^T \boldsymbol{p} = 1, \tag{4.26b}$$

$$\psi(\boldsymbol{p}, \boldsymbol{q}) \leq d, \tag{4.26c}$$

$$\boldsymbol{p} \geq 0, \tag{4.26d}$$

where the constraints (4.26b–4.26d) arise from the ambiguity set $\mathcal{U}_{\boldsymbol{p}}(\boldsymbol{q}, d)$ in (4.16). In particular, constraint (4.26c) is shown with respect to a generic distance measure $\psi(\boldsymbol{p}, \boldsymbol{q})$, which can be defined by the user as any of the measures in (4.17a–4.17e) with an appropriate maximum permissible distance d. Thus, for some point $\boldsymbol{u}$, the projection $\Pi_{\mathcal{U}_{\boldsymbol{p}}}(\boldsymbol{u})$ finds the closest solution within the ambiguity set $\mathcal{U}_{\boldsymbol{p}}(\boldsymbol{q}, d)$.

Likewise, we retain feasibility in the descent step by projecting each iteration onto the set

$\mathbb{R}^n_+$. This projection is trivial and, for some arbitrary point $\boldsymbol{z} \in \mathbb{R}^n$, can be computed as follows

$$\Pi_{\mathbb{R}^n_+}(\boldsymbol{z}) \triangleq \begin{cases} z_i & \text{if } z_i > 0 \\ 0^+ & \text{otherwise} \end{cases} \quad \text{for } i = 1, \ldots, n, \tag{4.27}$$

where we inspect every element of $\boldsymbol{z}$ and set any non-positive element to an arbitrarily small positive value.[5]

We proceed to discuss the PGDA algorithm. Like an unconstrained gradient descent–ascent algorithm, we take steps to descend in $\boldsymbol{y}$ and ascend in $\boldsymbol{p}$ in the direction of the respective gradients of $f_{\text{RP}}(\boldsymbol{y}, \boldsymbol{p})$. The gradients of $f_{\text{RP}}(\boldsymbol{y}, \boldsymbol{p})$ are

$$\nabla_{\boldsymbol{y}} f_{\text{RP}}(\boldsymbol{y}, \boldsymbol{p}) = \hat{\boldsymbol{\Sigma}}(\boldsymbol{p})\boldsymbol{y} - \kappa \boldsymbol{y}^{\circ -1}, \tag{4.28}$$

$$\nabla_{\boldsymbol{p}} f_{\text{RP}}(\boldsymbol{y}, \boldsymbol{p}) = \frac{1}{2}\hat{\boldsymbol{\pi}}^2(\boldsymbol{y}) - \hat{\boldsymbol{\Gamma}}(\boldsymbol{y})\boldsymbol{p} \tag{4.29}$$

where $\boldsymbol{y}^{\circ -1} = [1/y_1 \cdots 1/y_n]^\top$.

Given that the feasible sets $\mathbb{R}^n_+$ and $\mathcal{U}_{\boldsymbol{p}}$ are convex, we can design the search directions in both $\boldsymbol{y}$ and $\boldsymbol{p}$ such that we retain feasibility after each iteration. Assume we have some feasible solutions $\boldsymbol{y}^k \in \mathbb{R}^n_+$ and $\boldsymbol{p}^k \in \mathcal{U}_{\boldsymbol{p}}$. The search directions at iteration k are

$$\begin{aligned} \boldsymbol{g}^k &\triangleq \Pi_{\mathbb{R}^n_+}\Big(\boldsymbol{y}^k - \gamma_k^{\boldsymbol{y}} \nabla_{\boldsymbol{y}} f_{\text{RP}}(\boldsymbol{y}^k, \boldsymbol{p}^k)\Big) - \boldsymbol{y}^k, \\ \boldsymbol{h}^k &\triangleq \Pi_{\mathcal{U}_{\boldsymbol{p}}}\Big(\boldsymbol{p}^k + \gamma_k^{\boldsymbol{p}} \nabla_{\boldsymbol{p}} f_{\text{RP}}(\boldsymbol{y}^{k+1}, \boldsymbol{p}^k)\Big) - \boldsymbol{p}^k, \end{aligned}$$

where $\gamma_k^{\boldsymbol{y}}$ and $\gamma_k^{\boldsymbol{p}}$ are the step sizes in each direction. To ensure that our next iteration remains within the feasible set, we define the search parameters $\eta_{\boldsymbol{y}}, \eta_{\boldsymbol{p}} \in [0, 1]$. Thus, our next iterations in each direction are

$$\begin{aligned} \boldsymbol{y}^{k+1} &= \boldsymbol{y}^k + \eta_{\boldsymbol{y}} \boldsymbol{g}^k, \\ \boldsymbol{p}^{k+1} &= \boldsymbol{p}^k + \eta_{\boldsymbol{p}} \boldsymbol{h}^k. \end{aligned}$$

The points $\boldsymbol{y}^{k+1}$ and $\boldsymbol{p}^{k+1}$ are the result of linear combinations between two feasible points

[5] We must replace any non-positive value with a strictly positive value, 0^+. This stems from the derivative of the logarithm barrier in our objective function, where $(d/dz_i)\ln(z_i) = 1/z_i$. Thus, if z_i is not strictly positive, this will lead to a numerical error when we implement the algorithm.

in each set, respectively. Since the sets are convex, the points $\boldsymbol{y}^{k+1}$ and $\boldsymbol{p}^{k+1}$ are feasible by definition.

We defer to the Barzilai–Borwein method [6] to define the step sizes $\gamma_k^{\boldsymbol{y}}$ and $\gamma_k^{\boldsymbol{p}}$. Specifically, we use the following definition of the Barzilai–Borwein method. For any iteration $k \geq 1$ where $k = 0, 1, \ldots$, we have

$$\gamma_k^{\boldsymbol{y}} = \frac{\left|(\boldsymbol{y}^k - \boldsymbol{y}^{k-1})^\top \left(\nabla_{\boldsymbol{y}} f_{\mathrm{RP}}(\boldsymbol{y}^k, \boldsymbol{p}^k) - \nabla_{\boldsymbol{y}} f_{\mathrm{RP}}(\boldsymbol{y}^{k-1}, \boldsymbol{p}^{k-1})\right)\right|}{\left\|\nabla_{\boldsymbol{y}} f_{\mathrm{RP}}(\boldsymbol{y}^k, \boldsymbol{p}^k) - \nabla_{\boldsymbol{y}} f_{\mathrm{RP}}(\boldsymbol{y}^{k-1}, \boldsymbol{p}^{k-1})\right\|_2^2}, \tag{4.30}$$

$$\gamma_k^{\boldsymbol{p}} = \frac{\left|(\boldsymbol{p}^k - \boldsymbol{p}^{k-1})^\top \left(\nabla_{\boldsymbol{p}} f_{\mathrm{RP}}(\boldsymbol{y}^k, \boldsymbol{p}^k) - \nabla_{\boldsymbol{p}} f_{\mathrm{RP}}(\boldsymbol{y}^{k-1}, \boldsymbol{p}^{k-1})\right)\right|}{\left\|\nabla_{\boldsymbol{p}} f_{\mathrm{RP}}(\boldsymbol{y}^k, \boldsymbol{p}^k) - \nabla_{\boldsymbol{p}} f_{\mathrm{RP}}(\boldsymbol{y}^{k-1}, \boldsymbol{p}^{k-1})\right\|_2^2}. \tag{4.31}$$

The Barzilai–Borwein step size is sometimes referred to as the 'spectral step size'. In the case of projected gradient descent, this class of algorithms is sometimes referred to as 'spectral projected gradient descent' [16].

Next, we discuss how to determine the search parameters $\eta_{\boldsymbol{y}}, \eta_{\boldsymbol{p}} \in (0, 1]$. Specifically, we favour the non-monotone Grippo–Lampariello–Lucidi (GLL) line search proposed in [51]. The GLL line search method has been shown to work well with spectral projected gradient descent and ensures global convergence on closed convex sets [16, 31].

A brief overview of this line search method follows. Consider the descent step in $\boldsymbol{y}$. For a given integer $m \geq 1$, we are searching for $\eta_{\boldsymbol{y}} \in (0, 1]$ such that

$$f_{\mathrm{RP}}(\boldsymbol{y}^k + \eta_{\boldsymbol{y}} \boldsymbol{g}^k, \boldsymbol{p}^k) \leq \max_{j \in \mathcal{J}} f_{\mathrm{RP}}(\boldsymbol{y}^{k-j}, \boldsymbol{p}^{k-j}) + \beta \eta_{\boldsymbol{y}} (\boldsymbol{g}^k)^\top \nabla_{\boldsymbol{y}} f_{\mathrm{RP}}(\boldsymbol{y}^k, \boldsymbol{p}^k) \tag{4.32}$$

where $\mathcal{J} \triangleq \big\{j \in \mathbb{Z} : 0 \leq j \leq \min\{k, m-1\}\big\}$ and $\beta \in (0, 1)$ is some predefined constant. Intuitively, a more aggressive descent step corresponds to a larger value of $\eta_{\boldsymbol{y}}$. This typically means we initially set $\eta_{\boldsymbol{y}} = 1$ and shrink it appropriately by some fixed factor $\tau \in (0, 1)$, resulting in an inexact but fast method to determine an appropriate value for $\eta_{\boldsymbol{y}}$.

The GLL method stems from an Armijo-type line search, but it allows us to be greedier with our step sizes. For example, if we set $m = 1$, then we revert back to a traditional Armijo-type line search method and the condition in (4.32) causes our objective function to decrease monotonically. Thus, by considering multiple previous iterations of the objective value we allow

for a non-monotonic decrease.

For the purpose of the PGDA algorithm, the ascent direction follows the same logic. However, we do not discuss it in detail for the sake of brevity. Instead, the complete PGDA algorithm is presented in Algorithm 1, which shows how to calculate the steps in the descent and ascent directions.

Although the global convergence of spectral projected gradient descent with a GLL line search has been established [16, 31], we avoid making any claims that this is true for our PGDA algorithm. However, we note that for appropriate step sizes, the convergence of convex–concave constrained minimax problems has been previously established [78]. From our perspective, the PGDA algorithm serves as a stepping stone towards the development of the SCP–PGA algorithm in Section 4.2.4 below.

4.2.4 Sequential convex programming with projected gradient ascent

The PGDA algorithm served two purposes. First, it provided a straightforward approach to solve a convex–concave minimax problem. More importantly, it showed the steps required to navigate such a problem and highlighted some structural weaknesses. In particular, the PGDA algorithm requires that we determine two appropriate step sizes, $\gamma_k^{\boldsymbol{y}}$ and $\gamma_k^{\boldsymbol{p}}$, during each iteration. Moreover, the set $\mathbb{R}_+^n$ in which $\boldsymbol{y}$ exists is not compact. Therefore, the PGDA algorithm necessitates careful initialization and, in general, may be prone to diverge.

The PGDA has three structural weaknesses. First, the design of the risk parity problem in (4.7) means that the descent step in the PGDA algorithm must ignore the budget equality constraint. In other words, the algorithm operates in the unbounded set $\mathbb{R}_+^n$ instead of the compact set $\mathcal{X}$. Only after convergence of the PGDA algorithm do we project our solution onto $\mathcal{X}$.

Second, the PGDA algorithm is twice as susceptible to the problem of vanishing gradients. As we approach a saddle point, the gradient information in both directions starts to vanish, slowing the convergence of the algorithm to an optimal saddle point.

The third and final weakness is the burden placed on the user to define an initial step size in both $\boldsymbol{y}$ and $\boldsymbol{p}$ directions, as well as the initial guess $\boldsymbol{y}^0$. Since the proxy variable $\boldsymbol{y}$ does not have an upper bound, an improperly sized $\boldsymbol{y}^0$ may slow down convergence.

These three weaknesses can be remediated by redesigning the algorithm to operate directly on the set of admissible portfolios $\mathcal{X}$. The strict convexity of the nominal risk parity problem in

Algorithm 1: PGDA for DRRP portfolio optimization

Input: Data $\hat{\boldsymbol{\xi}} \in \mathbb{R}^{n\times T}$; Confidence level $\delta \in (0,1)$; Distance measure {JS, Hellinger, TV}; Nominal distribution $\boldsymbol{q} \in \mathcal{P}$; Risk parity constant $\kappa > 0$; Initial step sizes $\gamma_0^{\boldsymbol{y}}, \gamma_0^{\boldsymbol{p}} > 0$; Initial proxy portfolio $\boldsymbol{y}^0$; Convergence tolerance ε_0; Search control parameters $\beta, \tau \in (0,1)$; GLL parameter $m \geq 1$

Output: Optimal DRRP portfolio $\boldsymbol{x}^*$

Find the distance limit $d(\delta, T)$ as shown in either of (4.20–4.22)
Initialize the adversarial distribution: $\boldsymbol{p}^0 = \boldsymbol{q}$
Initialize the convergence measure: $\varepsilon = 1$
Initialize the counter: $k = 0$
while $\varepsilon > \varepsilon_0$ **do**
  if $k \geq 1$ **then**
    Update $\gamma_k^{\boldsymbol{y}}$ as shown in (4.30)
    Update $\gamma_k^{\boldsymbol{p}}$ as shown in (4.31)
  $\boldsymbol{g}^k = \Pi_{\mathbb{R}^n_+}\Big(\boldsymbol{y}^k - \gamma_k^{\boldsymbol{y}} \nabla_{\boldsymbol{y}} f_{\mathrm{RP}}(\boldsymbol{x}^k, \boldsymbol{p}^k)\Big) - \boldsymbol{y}^k$
  $\eta_{\boldsymbol{y}} = 1$
  $\bar{\boldsymbol{y}} = \boldsymbol{y}^k + \eta_{\boldsymbol{y}} \boldsymbol{g}^k$
  while $f_{RP}(\bar{\boldsymbol{y}}, \boldsymbol{p}^k) > \max\limits_{j\in\mathcal{J}} f_{RP}(\boldsymbol{y}^{k-j}, \boldsymbol{p}^{k-j}) + \beta\eta_{\boldsymbol{y}} (\boldsymbol{g}^k)^\top \nabla_{\boldsymbol{y}} f_{RP}(\boldsymbol{x}^k, \boldsymbol{p}^k)$ **do**
    $\eta_{\boldsymbol{y}} = \eta_{\boldsymbol{y}} \tau$
    $\bar{\boldsymbol{y}} = \boldsymbol{y}^k + \eta_{\boldsymbol{y}} \boldsymbol{g}^k$
  $\boldsymbol{y}^{k+1} = \bar{\boldsymbol{y}}$
  $\boldsymbol{h}^k = \Pi_{\mathcal{U}_{\boldsymbol{p}}}\Big(\boldsymbol{p}^k + \gamma_k^{\boldsymbol{p}} \nabla_{\boldsymbol{p}} f_{\mathrm{RP}}(\boldsymbol{x}^k, \boldsymbol{p}^k)\Big) - \boldsymbol{p}^k$
  $\eta_{\boldsymbol{p}} = 1$
  $\bar{\boldsymbol{p}} = \boldsymbol{p}^k + \eta_{\boldsymbol{p}} \boldsymbol{h}^k$
  while $f_{RP}(\boldsymbol{y}^k, \bar{\boldsymbol{p}}) < \min\limits_{j\in\mathcal{J}} f_{RP}(\boldsymbol{y}^{k-j}, \boldsymbol{p}^{k-j}) + \beta\eta_{\boldsymbol{p}} (\boldsymbol{h}^k)^\top \nabla_{\boldsymbol{p}} f_{RP}(\boldsymbol{x}^k, \boldsymbol{p}^k)$ **do**
    $\eta_{\boldsymbol{p}} = \eta_{\boldsymbol{p}} \tau$
    $\bar{\boldsymbol{p}} = \boldsymbol{p}^k + \eta_{\boldsymbol{p}} \boldsymbol{h}^k$
  $\boldsymbol{p}^{k+1} = \bar{\boldsymbol{p}}$
  if $k \geq 1$ **then**
    $\varepsilon = \|\boldsymbol{p}^{k+1} - \boldsymbol{p}^k\|_2$
  $k = k + 1$
Find the optimal portfolio: $\boldsymbol{x}^* = \boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p}^k)$

Result: Optimal DRRP portfolio $\boldsymbol{x}^*$

(4.7) means that there exists a unique risk parity portfolio $\boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p})$ for every $\boldsymbol{p} \in \mathcal{U}_{\boldsymbol{p}}$. Assume we have an ascent algorithm and let $\boldsymbol{p}^k$ be the k^{th} iteration of our adversarial probability. Then, for every $k = 0, 1, \ldots$, there exists a corresponding risk parity portfolio $\boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p}^k)$. Thus, we can formulate an algorithm that ascends in $\boldsymbol{p} \in \mathcal{U}_{\boldsymbol{p}}$ while enforcing the risk parity condition in $\boldsymbol{x} \in \mathcal{X}$ after every iteration.

Conversely, we can interpret this algorithm as solving a sequence of convex problems. Specifically, we solve the risk parity problem $\boldsymbol{x}^k = \boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p}^k)$, where we update the covariance matrix $\hat{\boldsymbol{\Sigma}}(\boldsymbol{p}^k)$ after every iteration k. Thus, the resulting algorithm needs only to ascend in $\boldsymbol{p} \in \mathcal{U}_{\boldsymbol{p}}$, meaning it can be solved using PGA. In turn, this means that the user no longer needs to define any of the initial conditions and updates associated with the proxy variable $\boldsymbol{y}$. Given that the proposed PGA algorithm involves iteratively solving a sequence of convex problems, we refer to it as the SCP–PGA algorithm.

In (4.25) we stated that the following inequality holds for the saddle point $(\boldsymbol{y}^*, \boldsymbol{p}^*)$ given that $f_{\mathrm{RP}}(\boldsymbol{y}, \boldsymbol{p})$ is convex–concave,

$$f_{\mathrm{RP}}(\boldsymbol{y}^*, \boldsymbol{p}) \leq f_{\mathrm{RP}}(\boldsymbol{y}^*, \boldsymbol{p}^*) \leq f_{\mathrm{RP}}(\boldsymbol{y}, \boldsymbol{p}^*) \;\; \forall \; \boldsymbol{y} \in \mathbb{R}^n_+, \; \boldsymbol{p} \in \mathcal{U}_{\boldsymbol{p}}.$$

As per the risk parity problem in (4.7), we also have that the saddle point $(\boldsymbol{y}^*, \boldsymbol{p}^*)$ corresponds to $\boldsymbol{y}^* = \boldsymbol{y}^{\mathrm{RP}}(\boldsymbol{p}^*)$. Moreover, projecting $\boldsymbol{y}^*$ onto $\mathcal{X}$ yields the corresponding optimal DRRP portfolio $\boldsymbol{x}^*$, which leads to the following conclusion: $\boldsymbol{x}^* = \boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p}^*)$. We design the SCP–PGA algorithm to enforce the risk parity condition during every iteration k, i.e., we have $\boldsymbol{x}^k = \boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p}^k)$. Thus, at convergence, we reach the same conclusion as before: $\boldsymbol{x}^* = \boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p}^*)$. Consequently, the inequality above can be restated as follows

$$f_{\mathrm{RP}}\big(\boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p}), \boldsymbol{p}\big) \leq f_{\mathrm{RP}}\big(\boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p}^*), \boldsymbol{p}^*\big) \;\; \forall \; \boldsymbol{x} \in \mathcal{X}, \; \boldsymbol{p} \in \mathcal{U}_{\boldsymbol{p}},$$

indicating that we can use PGA to maximize $f_{\mathrm{RP}}\big(\boldsymbol{x}^{\mathrm{RP}}(\cdot), \cdot\big) : \mathcal{U}_{\boldsymbol{p}} \to \mathbb{R}$ while maintaining the risk parity condition during every iteration.

Since a closed-form solution to $\boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p})$ does not exist, we must proceed iteratively by solving $\boldsymbol{x}^k = \boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p}^k)$ at every iteration k. Although this increases the computational cost per iteration when compared against our previous PGDA algorithm, we note that convex optimization problems can be efficiently solved by modern optimization algorithms and software packages. Thus, as we will show numerically in Section 4.3, the actual computational cost is almost negligible

while also reducing the number of iterations required for convergence.

The SCP–PGA algorithm follows the same logic as the PGDA algorithm, except we are only concerned with the ascent step. Our maximization problem is quadratic concave over a compact convex set. Moreover, the gradient of the objective function is Lipschitz continuous since its Hessian is PSD for all $\boldsymbol{p}$. Therefore, using an appropriate line search can guarantee convergence. Specifically, the use of the GLL line search in our algorithm means that, by design, each iteration achieves a sufficient increase in $f_{\mathrm{RP}}\big(\boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p}), \boldsymbol{p}\big)$ such that we converge to the global maximum. We finish this subsection by presenting the complete SCP–PGA algorithm in Algorithm 2.

Algorithm 2: SCP–PGA for DRRP portfolio optimization

Input: Data $\hat{\boldsymbol{\xi}} \in \mathbb{R}^{n\times T}$; Confidence level $\delta \in (0,1)$; Distance measure {JS, Hellinger, TV}; Nominal distribution $\boldsymbol{q} \in \mathcal{P}$; Risk parity constant $\kappa > 0$; Initial step size $\gamma_0^{\boldsymbol{p}} > 0$; Convergence tolerance ε_0; Search control parameters $\beta, \tau \in (0,1)$; GLL parameter $m \geq 1$

Output: Optimal DRRP portfolio $\boldsymbol{x}^*$

Find the distance limit $d(\delta, T)$ as shown in either of (4.20–4.22)
Initialize the adversarial distribution $\boldsymbol{p}^0 = \boldsymbol{q}$
Initialize the convergence measure: $\varepsilon >> 1$
Initialize the counter: $k = 0$
while $\varepsilon > \varepsilon_0$ **do**
  if $k \geq 1$ **then**
    Update $\gamma_k^{\boldsymbol{p}}$ as shown in (4.31)
  $\boldsymbol{x}^k = \boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p}^k)$
  $\boldsymbol{h}^k = \Pi_{\mathcal{U}_{\boldsymbol{p}}}\Big(\boldsymbol{p}^k + \gamma_k^{\boldsymbol{p}} \nabla_{\boldsymbol{p}} f_{\mathrm{RP}}(\boldsymbol{x}^k, \boldsymbol{p}^k)\Big) - \boldsymbol{p}^k$
  $\eta_{\boldsymbol{p}} = 1$
  $\bar{\boldsymbol{p}} = \boldsymbol{p}^k + \eta_{\boldsymbol{p}} \boldsymbol{h}^k$
  while $f_{RP}(\boldsymbol{x}^k, \bar{\boldsymbol{p}}) < \min\limits_{j\in\mathcal{J}} f_{RP}(\boldsymbol{x}^{k-j}, \boldsymbol{p}^{k-j}) + \beta\eta_{\boldsymbol{p}} (\boldsymbol{h}^k)^\top \nabla_{\boldsymbol{p}} f_{RP}(\boldsymbol{x}^k, \boldsymbol{p}^k)$ **do**
    $\eta_{\boldsymbol{p}} = \tau\eta_{\boldsymbol{p}}$
    $\bar{\boldsymbol{p}} = \boldsymbol{p}^k + \eta_{\boldsymbol{p}} \boldsymbol{h}^k$
  $\boldsymbol{p}^{k+1} = \bar{\boldsymbol{p}}$
  if $k \geq 1$ **then**
    $\varepsilon = \|\boldsymbol{p}^{k+1} - \boldsymbol{p}^k\|_2$
  $k = k + 1$

Result: Optimal DRRP portfolio $\boldsymbol{x}^*$

4.3 Numerical Experiments

This section consists of three separate experiments. The first experiment serves to evaluate the numerical performance of the SCP–PGA algorithm (Algorithm 2) and is conducted using synthetic data to generate increasingly larger datasets. As a benchmark, this experiment also includes results from the PGDA algorithm (Algorithm 1). The second experiment assesses the in-sample performance of the DRRP portfolio in a financial context and uses historical data. The third experiment assesses the out-of-sample financial performance of the portfolio. To limit our scope, the experiments are conducted on DRRP portfolios based on the JS, Hellinger and TV distance measures.

The second and third experiments share the same data. The data consist of historical observations ranging from the start of 1998 until the end of 2016 for 30 industry portfolios. These industry portfolios serve as our financial assets and are akin to many popular exchange traded funds. The data were obtained from Kenneth R. French's data library [42]. Table 4.1 lists the 30 industry portfolios.

Table 4.1: List of assets

Food Products	Tobacco	Beer and Liquor	Recreation
Household Products	Apparel	Healthcare	Chemicals
Fabricated Products	Construction	Steel Works	Electrical Equip.
Aircraft, Ships, Rail Equip.	Mining	Coal	Oil and Gas
Communication	Services	Business Equip.	Paper
Restaurants and Hotels	Wholesale	Retail	Financials
Printing	Textiles	Automobiles	Utilities
Transportation	Other		

All experiments were conducted on an Apple MacBook Pro computer (2.8 GHz Intel Core i7, 16 GB 2133 MHz DDR3 RAM) running macOS 'Catalina'. The computer script was written in the Julia programming language (version 1.4.0) using the modelling language 'JuMP' [33] with IPOPT (version 3.12.6) as the optimization solver.

4.3.1 Numerical performance and tractability

The first part of the numerical performance experiment compares the SCP–PGA algorithm to the PGDA algorithm. The second part of the numerical performance experiment extends the

results of the SCP–PGA algorithm and compares the results against the nominal risk parity problem.

The comparison between the SCP–PGA and PGDA algorithms has the following experimental setup. We randomly generate synthetic datasets with $n = 50$, 200 assets and $T = 1{,}000$, 5,000 scenarios for a total of four different datasets. The largest dataset, with $n = 200$ and $T = 5,000$, simulates the conditions to create a portfolio with 200 constituents using approximately 20 years worth of daily scenarios. We construct optimal DRRP portfolios for three different confidence levels $\delta = 0.15$, 0.3, 0.45.

The remainder of the user-defined parameters are the following. The risk parity constant is set to $\kappa = 1$, while the convergence tolerance is set to $\varepsilon_0 = 10^{-6}$. As recommended in [16], we set $m = 10$. Moreover, we set the search parameters to $\beta = 10^{-6}$ and $\tau = 0.9$. The initial ascent step size is $\gamma_0^{\boldsymbol{p}} = 0.1$. In addition, we set the following values for the PGDA algorithm: $\gamma_0^{\boldsymbol{y}} = 30$, $y_i^0 = 5$ for $i = 1, \ldots, n$.

The SCP–PGA and PGDA algorithms are evaluated based on their runtime in seconds, the number of iterations until convergence, and the resulting portfolio variance. The two algorithms aim to construct risk parity portfolios with the worst-case estimate of the portfolio variance with respect to the probability ambiguity set $\mathcal{U}_{\boldsymbol{p}}$. Since, by design, both algorithms yield portfolios that satisfy the risk parity condition,[6] we evaluate the convergence quality of the two algorithms by comparing the corresponding portfolio variances. The numerical results are presented in Table 4.2.

The results in Table 4.2 show that the SCP–PGA algorithm is able to attain an equal or higher portfolio variance than the PGDA algorithm in every single instance, indicating that the SCP–PGA algorithm converges to a higher quality solution. This highlights the sensitivity of the PGDA algorithm to its initial conditions, where the algorithm appears to converge if the step sizes become ill-conditioned and not enough progress is made in the probability space (i.e., the PGDA algorithm terminates because $\varepsilon \leq \varepsilon_0$ after the step in the $\boldsymbol{p}$-direction becomes negligible). Moreover, for every instance where the variance of both algorithms is the same, the runtime of the SCP–PGA algorithm is significantly faster than the PGDA algorithm (e.g., see the results for $\delta = 0.15$, $n = 50$ and $T = 5,000$ with the JS divergence as the distance measure).

The second part of the numerical performance experiment focuses solely on the SCP–PGA

[6] The SCP–PGA algorithm finds a risk parity portfolio $\boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p}^k)$ during each iteration k, while the last line of the PGDA algorithm also enforces $\boldsymbol{x}^{\mathrm{RP}}(\boldsymbol{p}^*)$ after convergence in $\boldsymbol{p}$.

Table 4.2: Comparison of numerical performance between the PGDA algorithm (A.1) and the SCP–PGA algorithm (A.2)

	$n = 50$						$n = 200$					
	JS		Hellinger		TV		JS		Hellinger		TV	
	A.1	A.2	A.1	A.2	A.1	A.2	A.1	A.2	A.1	A.2	A.1	A.2
$\delta = 0.15$												
$\boldsymbol{T = 1,000}$												
Time (s)	26.3	1.49	3.49	2.40	57.9	8.32	5.96	5.94	18.2	6.11	7.84	10.5
Iterations	221	12	13	11	185	27	13	13	25	11	11	17
Var. ($\times 10^4$)	2.40	6.61	3.84	7.13	3.90	9.38	2.07	6.17	3.49	6.70	5.81	9.13
$\boldsymbol{T = 5,000}$												
Time (s)	335	9.15	72.3	17.9	191	33.9	79.6	35.3	4,172	33.2	322	104
Iterations	442	12	29	12	90	19	17	14	1,000	11	44	29
Var. ($\times 10^4$)	5.64	5.64	1.68	6.03	3.49	9.89	4.43	5.23	5.51	5.68	2.76	8.17
$\delta = 0.3$												
$\boldsymbol{T = 1,000}$												
Time (s)	7.04	2.98	9.42	7.44	7.69	10.9	436	7.99	143	14.2	19.9	14.9
Iterations	76	34	39	28	23	40	1000	16	272	25	31	22
Var. ($\times 10^4$)	10.2	10.8	11.7	11.9	5.62	14.4	1.16	10.5	0.97	11.7	3.28	14.2
$\boldsymbol{T = 5,000}$												
Time (s)	31.9	13.1	60.5	36.5	275	50.1	157	79.9	569	99.6	235	128
Iterations	22	24	37	23	127	32	59	33	132	28	48	37
Var. ($\times 10^4$)	5.80	10.8	11.6	11.6	1.28	16.2	3.97	9.06	3.67	9.98	0.71	12.9
$\delta = 0.45$												
$\boldsymbol{T = 1,000}$												
Time (s)	8.43	4.16	5.70	8.58	24.5	21.5	5.27	13.0	31.7	15.8	61.4	23.6
Iterations	87	45	22	38	86	72	11	27	51	27	108	35
Var. ($\times 10^4$)	16.1	16.1	0.51	17.8	4.87	19.3	10.2	16.1	10.3	17.9	9.50	19.1
$\boldsymbol{T = 5,000}$												
Time (s)	34.5	20.6	198	64.6	98.3	66.2	310	99.4	320	122	127	224
Iterations	51	37	78	38	47	40	141	41	42	38	14	62
Var. ($\times 10^4$)	17.9	17.9	1.81	19.3	22.0	22.5	14.3	14.3	1.70	15.7	4.87	17.6

Note: The maximum number of iterations is limited to 1,000, after which the algorithms terminate.

algorithm and extends our previous results to include synthetic datasets with $T = 100$ and $n = 500$. Thus, including the original values of T and n, we have a total of nine different datasets.

As before, the DRRP portfolios from the SCP–PGA algorithm are evaluated based on their runtime, the number of iterations until convergence, and the portfolio variance. We also include the runtime per iteration. Finally, the results also show the variance of the nominal risk parity portfolio for the same dataset. The nominal portfolio variance serves as a benchmark. When looking at the DRRP portfolio variance, we must keep in mind that this corresponds to the worst-case estimate of the variance as defined by the ambiguity set $\mathcal{U}_{\boldsymbol{p}}$. Therefore, we expect the DRRP portfolio variance to be larger than the nominal. The results are presented in Table 4.3.

The results in Table 4.3 indicate that, overall, the SCP–PGA algorithm converges within reasonable time, even for the largest dataset tested. We note that the largest dataset, with $n = 500$ and $T = 5,000$, exaggerates the number of scenarios T that we would normally consider for parameter estimation in a conventional environment.[7] Most financial data service providers tend to use anywhere from 10 days to five years when calculating risk metrics such as the portfolio variance or the CAPM 'beta' [93], and rely on daily, weekly or monthly scenarios for these calculations.

As shown by the different runtimes in Table 4.3, the SCP–PGA algorithm converged the fastest when we used the JS divergence to construct the ambiguity set (with a few exceptions where the Hellinger distance was faster). The runtime per iteration is relatively similar for all three distance measures. Thus, this suggests that having a faster convergence rate is mostly dependent on the number of iterations required until convergence.

If we inspect the portfolio variance, we can see that the JS divergence is more restrictive than either the Hellinger distance or the TV distance. In other words, for the same confidence level, the JS divergence provides a smaller feasible region in our variance maximization step, leading to portfolios with lower variances. Conversely, the TV distance is the most permissive, consistently having the highest variance for all trials. This suggests that, although the three distance measures have been scaled proportionally, their intrinsic differences suggest some are fundamentally more permissive than others from a portfolio variance perspective.

[7] This may exclude high frequency trading environments.

Table 4.3: Numerical performance of the SCP–PGA algorithm

	$n = 50$				$n = 200$				$n = 500$			
	Nom.	JS	H	TV	Nom.	JS	H	TV	Nom.	JS	H	TV
$\delta = 0.15$												
$\boldsymbol{T = 100}$												
Time (s)	-	0.24	0.34	0.98	-	1.26	0.98	2.24	-	5.57	5.19	8.57
Iterations	-	10	9	27	-	10	9	22	-	10	9	14
Time/Iter. (s)	-	0.02	0.04	0.04	-	0.13	0.11	0.1	-	0.56	0.57	0.61
Var. ($\times 10^4$)	3.17	4.90	5.19	5.64	2.98	4.96	5.31	5.96	3.29	5.91	6.37	7.44
$\boldsymbol{T = 1,000}$												
Time (s)	-	1.49	2.40	8.32	-	5.94	6.11	10.45	-	27.7	23.9	117
Iterations	-	12	11	27	-	13	11	17	-	12	11	42
Time/Iter. (s)	-	0.12	0.22	0.31	-	0.46	0.56	0.61	-	2.31	2.18	2.79
Var. ($\times 10^4$)	4.03	6.61	7.13	9.38	3.55	6.17	6.70	9.13	3.07	5.02	5.42	6.24
$\boldsymbol{T = 5,000}$												
Time (s)	-	9.15	17.9	33.9	-	35.3	33.2	104	-	147	132	277
Iterations	-	12	12	19	-	14	11	29	-	12	12	21
Time/Iter. (s)	-	0.76	1.50	1.78	-	2.52	3.02	3.6	-	12.2	11.0	13.2
Var. ($\times 10^4$)	3.11	5.64	6.03	9.89	3.11	5.23	5.68	8.17	3.39	5.91	6.37	9.76
$\delta = 0.3$												
$\boldsymbol{T = 100}$												
Time (s)	-	0.36	0.63	0.94	-	1.65	1.34	1.97	-	9.55	8.91	13.0
Iterations	-	17	917	28	-	17	13	18	-	16	15	22
Time/Iter. (s)	-	0.02	0.04	0.03	-	0.10	0.10	0.11	-	0.6	0.59	0.59
Var. ($\times 10^4$)	3.17	6.92	7.51	7.69	2.98	7.38	8.06	8.46	3.29	9.07	9.89	9.75
$\boldsymbol{T = 1,000}$												
Time (s)	-	2.98	7.44	10.9	-	7.99	14.2	14.9	-	64.9	62.3	65.1
Iterations	-	34	28	40	-	16	25	22	-	31	25	29
Time/Iter. (s)	-	0.09	0.27	0.27	-	0.5	0.57	0.67	-	2.09	2.49	2.24
Var. ($\times 10^4$)	4.03	10.8	11.9	14.4	3.55	10.5	11.7	14.2	3.07	7.57	8.52	9.25
$\boldsymbol{T = 5,000}$												
Time (s)	-	13.1	36.5	50.1	-	79.9	99.6	128	-	251	250	273
Iterations	-	24	23	32	-	33	28	37	-	29	23	30
Time/Iter. (s)	-	0.54	1.29	1.56	-	2.42	3.56	3.45	-	8.65	10.89	9.10
Var. ($\times 10^4$)	3.11	10.8	11.6	16.2	3.11	9.06	9.98	12.9	3.39	10.7	11.7	15.4
$\delta = 0.45$												
$\boldsymbol{T = 100}$												
Time (s)	-	0.66	0.93	1.32	-	2.16	1.87	2.25	-	12.7	11.1	16.3
Iterations	-	22	21	37	-	22	18	21	-	20	17	25
Time/Iter. (s)	-	0.03	0.04	0.04	-	0.1	0.1	0.11	-	0.64	0.65	0.65
Var. ($\times 10^4$)	3.17	9.04	9.84	9.61	2.98	9.69	10.5	10.5	3.29	11.1	12.0	10.7
$\boldsymbol{T = 1,000}$												
Time (s)	-	4.16	8.58	21.5	-	13.0	15.8	23.6	-	87.5	87.6	98.4
Iterations	-	45	38	72	-	27	27	35	-	44	36	44
Time/Iter. (s)	-	0.09	0.23	0.3	-	0.48	0.58	0.67	-	1.99	2.43	2.24
Var. ($\times 10^4$)	4.03	16.1	17.8	19.3	3.55	16.1	17.89	19.1	3.07	10.7	12.1	12.2
$\boldsymbol{T = 5,000}$												
Time (s)	-	20.6	64.6	66.2	-	99.4	122	224	-	288	501	510
Iterations	-	37	38	40	-	41	38	62	-	34	41	48
Time/Iter. (s)	-	0.56	1.70	1.66	-	2.43	3.22	3.61	-	8.47	12.2	10.6
Var. ($\times 10^4$)	3.11	17.87	19.3	22.5	3.11	14.3	15.7	17.6	3.39	16.9	18.5	20.7

Note: Nom, nominal.

4.3.2 In-sample experiment

To better understand how the distributionally robust framework works, we present a set of in-sample trials over varying levels of confidence. The DRRP portfolio is, by design, a risk parity portfolio under the worst-case estimate of the risk measure (i.e., the portfolio variance). This means that the portfolio risk is perfectly diversified among the constituent assets with respect to a given estimate of the covariance matrix. However, it is paramount to understand that only one risk parity portfolio exists for a specific instance of the covariance matrix. For example, if we have two estimates of the covariance matrix, $\boldsymbol{\Sigma}^a$ and $\boldsymbol{\Sigma}^b$, and we find the corresponding risk parity portfolios for each matrix, $\boldsymbol{x}^a$ and $\boldsymbol{x}^b$, then we have that $\boldsymbol{x}^a = \boldsymbol{x}^b$ if and only if $\boldsymbol{\Sigma}^a = c \cdot \boldsymbol{\Sigma}^b$ for any $c > 0$. It follows that if our covariance estimates differ, $\boldsymbol{\Sigma}^a \neq c \cdot \boldsymbol{\Sigma}^b \; \forall \; c > 0$, then $\boldsymbol{x}^a$ will not be a risk parity portfolio with respect to $\boldsymbol{\Sigma}^b$, and $\boldsymbol{x}^b$ will not be a risk parity portfolio with respect to $\boldsymbol{\Sigma}^a$.

With that said, our goal is to evaluate how the asset allocations and risk contributions differ between the robust portfolios and the nominal portfolio. We use the 30 industry portfolios listed in Table 4.1 as our assets ($n = 30$), and we use two years of weekly returns to estimate the covariance matrix, meaning we have 104 historical scenarios ($T = 104$). Specifically, the data corresponds to the time period from 01–Jan–2008 to 31–Dec–2009.

We begin by inspecting the asset weights and risk contributions for robust portfolios built with a confidence level $\delta = 0.3$. The asset weights are shown in Figure 4.1, and show that the DRRP portfolios exhibit a similar behaviour under all three statistical distance measures. Not only do the DRRP portfolios differ from the nominal portfolio, but the asset weights of all three DRRP portfolios also follow a similar pattern (i.e., the peaks and troughs in Figure 4.1 are similar for all portfolios). We also note that the wealth allocation of the DRRP portfolios is less pronounced than that of the nominal portfolio, with the DRRP portfolios exhibiting a more even distribution of wealth.

Figure 4.2 presents a similar analysis, except we compare the risk contribution per asset with respect to some estimate of the covariance matrix. For convenience, we restate that the risk contribution per asset is defined as $r_i \triangleq x_i[\hat{\boldsymbol{\Sigma}}\boldsymbol{x}]_i$ for some instance of the covariance matrix $\hat{\boldsymbol{\Sigma}}$. For a fair comparison, the top plot in Figure 4.2 shows the risk contribution per asset for all portfolios relative to the nominal estimate of the covariance matrix, $\boldsymbol{\Sigma}^{\text{nom}} \triangleq \boldsymbol{\Sigma}(\boldsymbol{q})$. The remaining three plots compare the risk contributions of the nominal portfolio with respect to

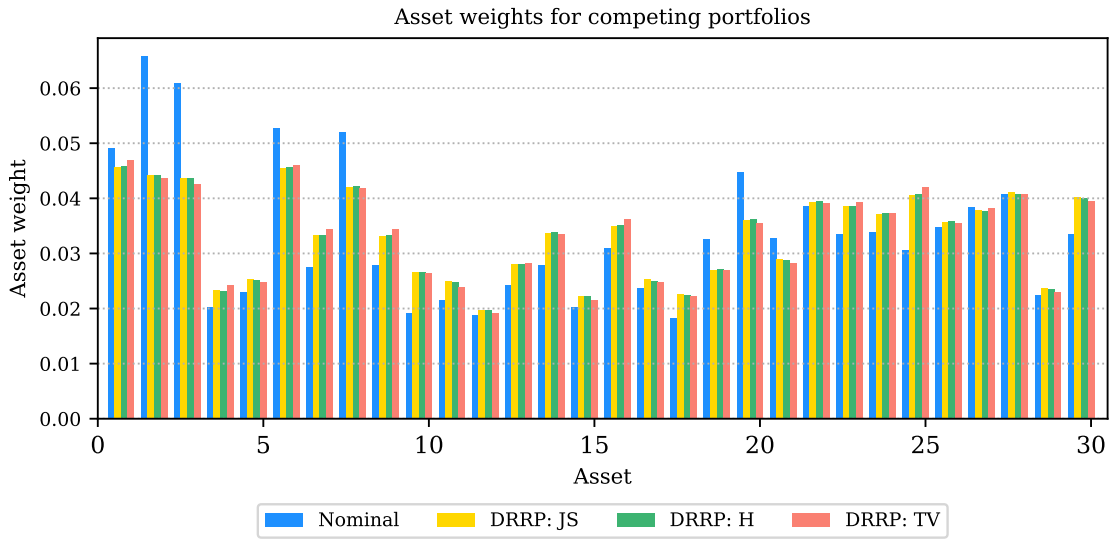

Figure 4.1: Asset weights of the nominal and DRRP portfolios with $\delta = 0.3$

$\boldsymbol{\Sigma}^{\text{JS}}$, $\boldsymbol{\Sigma}^{\text{H}}$ and $\boldsymbol{\Sigma}^{\text{TV}}$, respectively. These three matrices correspond to the estimated covariance matrix obtained after the convergence of Algorithm 2 for the respective distance measure.

The top plot in Figure 4.2 confirms the similarity between all three distributionally robust portfolios. For all risk contributions per asset, the three DRRP portfolios together are either lower or higher than the nominal portfolio (i.e., there is no asset where its risk contribution from a DRRP portfolio is higher than from the nominal portfolio while simultaneously lower from another DRRP portfolio). The DRRP portfolios choose the same assets to over- or under-contribute risk relative to the nominal portfolio, highlighting the structural similarity between the DRRP portfolios.

The three remaining plots in Figure 4.2 serve to show that all robust portfolios are true risk parity portfolios with respect to their corresponding estimate of the covariance matrix. As shown in the plots, the bars for the robust portfolios are of equal height.

The last component of the in-sample experiment replicates the same procedure, except we use varying levels of confidence δ. For brevity, these results are summarized in Table 4.4. The table shows the total variance of the four competing portfolios with respect to the nominal estimate of the covariance matrix, as well as a pairwise comparison of the total variance of the robust portfolios against the nominal using the corresponding worst-case estimates of the covariance matrix. In addition, we report the level of risk concentration through the CV of the risk contributions. The CV is calculated as shown in (2.21). In theory, an optimal risk parity portfolio should have a CV of zero.

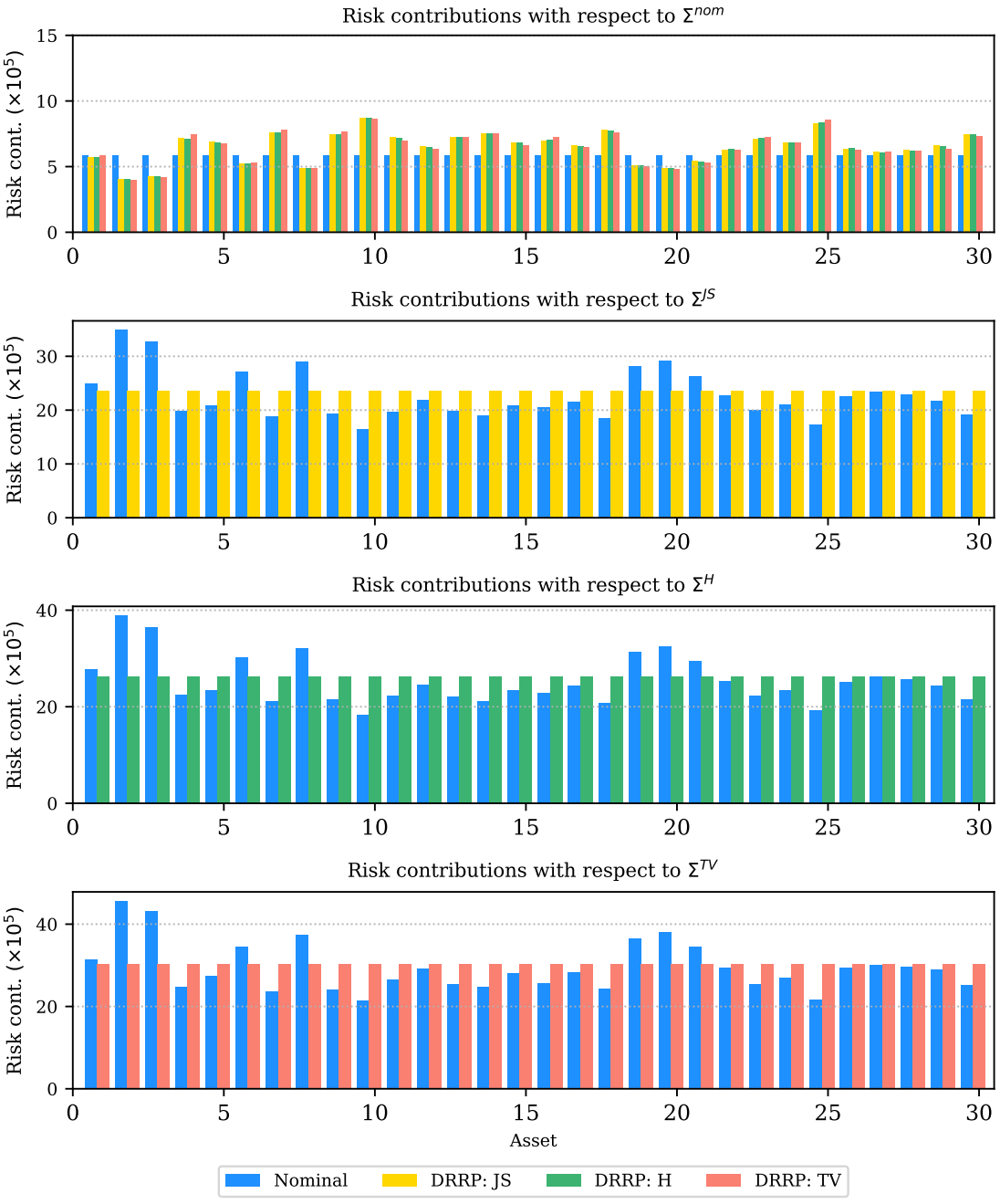

Figure 4.2: Risk contributions per asset of the nominal and DRRP portfolios with $\delta = 0.3$

Notes: The top plot shows the risk contributions with respect to the nominal estimate of the covariance matrix. The remaining plots show the risk contributions of the nominal portfolio and a single DRRP portfolio for the corresponding robust estimate of the covariance matrix.

Table 4.4: Portfolio variance and CV based on the nominal and worst-case estimates of the asset covariance matrix

	$\boldsymbol{\Sigma}^{\text{nom}}$				$\boldsymbol{\Sigma}^{\text{JS}}$		$\boldsymbol{\Sigma}^{\text{H}}$		$\boldsymbol{\Sigma}^{\text{TV}}$	
	$\boldsymbol{x}^{\text{nom}}$	$\boldsymbol{x}^{\text{JS}}$	$\boldsymbol{x}^{\text{H}}$	$\boldsymbol{x}^{\text{TV}}$	$\boldsymbol{x}^{\text{nom}}$	$\boldsymbol{x}^{\text{JS}}$	$\boldsymbol{x}^{\text{nom}}$	$\boldsymbol{x}^{\text{H}}$	$\boldsymbol{x}^{\text{nom}}$	$\boldsymbol{x}^{\text{TV}}$
$\boldsymbol{\delta = 0.1}$										
Var. ($\times 10^3$)	1.77	1.87	1.87	1.99	2.91	3.03	3.13	3.25	4.35	4.54
CV	7e-16	0.10	0.10	0.19	0.10	6e-16	0.11	2e-16	0.22	2e-16
$\boldsymbol{\delta = 0.2}$										
Var. ($\times 10^3$)	1.77	1.93	1.94	1.97	4.63	4.83	5.12	5.34	6.59	6.83
CV	7e-16	0.15	0.15	0.19	0.17	3e-16	0.17	3e-16	0.21	4e-16
$\boldsymbol{\delta = 0.3}$										
Var. ($\times 10^3$)	1.77	1.96	1.96	1.95	6.80	7.06	7.59	7.87	8.81	9.09
CV	7e-16	0.18	0.18	0.188	0.20	6e-16	0.20	4e-16	0.20	3e-16
$\boldsymbol{\delta = 0.4}$										
Var. ($\times 10^3$)	1.77	1.96	1.95	1.95	9.27	9.57	10.4	10.7	11.0	11.32
CV	7e-16	0.18	0.18	0.18	0.20	3e-16	0.20	3e-16	0.20	2e-16
$\boldsymbol{\delta = 0.5}$										
Var. ($\times 10^3$)	1.77	1.95	1.95	1.94	11.9	12.3	13.3	13.6	13.2	13.5
CV	7e-16	0.19	0.19	0.18	0.20	3e-16	0.20	3e-16	0.20	2e-16
$\boldsymbol{\delta = 0.6}$										
Var. ($\times 10^3$)	1.77	1.94	1.94	1.94	14.6	15.0	16.1	16.5	15.3	15.67
CV	7e-16	0.19	0.19	0.19	0.20	3e-16	0.20	6e-16	0.20	5e-16

Note: Nom, nominal. Var, variance. CV, coefficient of variation.

The results in Table 4.4 show that all portfolios have perfect risk diversification with respect to their corresponding estimates of the covariance matrix (i.e., the CV of the portfolios is approximately zero with respect to their corresponding instance of $\hat{\mathbf{\Sigma}}$). An interesting observation from Table 4.4 is that the nominal portfolio has the lowest total variance when compared against the robust portfolios for all instances of the covariance matrix. We note that this observation does not fundamentally conflict with our objective, as our robust portfolios aim to diversify risk, and not minimize it. Nevertheless, the results suggest that our robust portfolios incur more ex ante risk when compared to the nominal portfolio.

4.3.3 Out-of-sample experiment

The out-of-sample experiment allows us to evaluate the ex post financial performance of our DRRP portfolios. An overview of the experimental setup follows. Our portfolio constituents are the 30 assets listed in Table 4.1 ($n = 30$). The dataset consists of weekly historical returns from 01–Jan–1998 to 31–Dec–2016, with the data obtained from [42]. This is a rolling window experiment, where we use two years of weekly scenarios to calibrate our portfolios ($T = 104$) and we hold these portfolios for six month before rebalancing them. All estimated parameters and weights are recalibrated every time we rebalance our portfolios. To exemplify our approach, consider the first investment period. We use the data from 01–Jan-1998 to 31–Dec-1999 to calibrate our initial portfolios, and then we hold and observe the out-of-sample performance from 01–Jan–2000 to 30–Jun–2000. Afterwards, we roll the calibration window forward and recalibrate and rebalance our portfolios using the preceding two-year period (01–Jul–1998 to 30–Jun–2000). We then observe the out-of-sample performance from 01–Jul–2000 to 31–Dec–2000. We repeat these steps until the end of the investment horizon. Our out-of-sample experiment runs from 01–Jan–2000 until 31–Dec–2016, meaning we have a total of 34 six-month out-of-sample investment periods. We record the wealth evolution of the portfolios over the entire horizon. Finally, we note that this experiment is non-exhaustive since the portfolio performance is highly dependent on our choice of assets and historical time period. However, having a diverse basket of assets representative of major U.S. industries and a 17-year out-of-sample investment period should suffice for our analysis.

The first set of results, shown in Table 4.5 and Figure 4.3, correspond to portfolios with confidence level $\delta = 0.15, 0.3, 0.45$. The top plot in Figure 4.3 shows the total wealth evolution of the nominal portfolio. The remaining three plots show the relative wealth of the DRRP

portfolios. The 'relative wealth' is defined as a percentage, $(W_i^t/W_{\text{nom}}^t - 1) \times 100$, where W_i^t is the wealth of portfolio i at each weekly time step t, while W_{nom}^t is the nominal portfolio's wealth.

The robust portfolios exhibit a drop in their relative wealth over the bear market periods of 2000–2003 and 2008–2009. However, we note that the risk parity portfolios are not designed to minimize a portfolio's risk, but rather to be fully risk diverse. In turn, the results suggest that the DRRP portfolios are in a better position to take advantage of the subsequent bull market periods, where we can see sustained growth relative to the nominal. Moreover, the ex post portfolio performance aligns with our findings from the in-sample experiment, where we saw that the DRRP portfolios had a somewhat higher risk appetite given that they had a higher ex ante variance when compared to the nominal portfolio.

We summarize the ex post performance in Table 4.5, where we show the annualized excess return, annualized volatility, Sharpe ratio and average turnover rate over the entire investment horizon (2000–2016). We also provide the subset of results corresponding to a bear market period and its subsequent recovery (2007–2011). The results consistently show that the DRRP portfolios are able to attain a higher average excess return while maintaining a similar level of volatility, leading to higher Sharpe ratios. However, as the Sharpe ratio increases so does the average turnover rate, which serves as a proxy of transaction costs. Moreover, we note that the turnover rates of risk parity portfolios are typically very low when compared against other asset allocation strategies such as MVO (e.g., see [29]), and our results in Table 4.5 are no exception. Thus, the increased transaction costs incurred by the DRRP portfolios are somewhat negligible. Finally, we note that our observations are consistent over the 2007–2011 period, with the DRRP portfolios having a higher Sharpe ratio over this time period when compared to the nominal.

4.4 Conclusion

This chapter introduced a DRO problem specifically designed for risk parity portfolios. Distributional robustness is introduced through a discrete probability distribution that allows us to break away from the assumption that all scenarios in a data-driven parameter estimation process are equally likely. Instead, we can model the probability attached to each as a decision variable, which in turn allows us to formulate a minimax problem that seeks risk parity while simultaneously seeking the most adversarial instance of the discrete distribution such that the

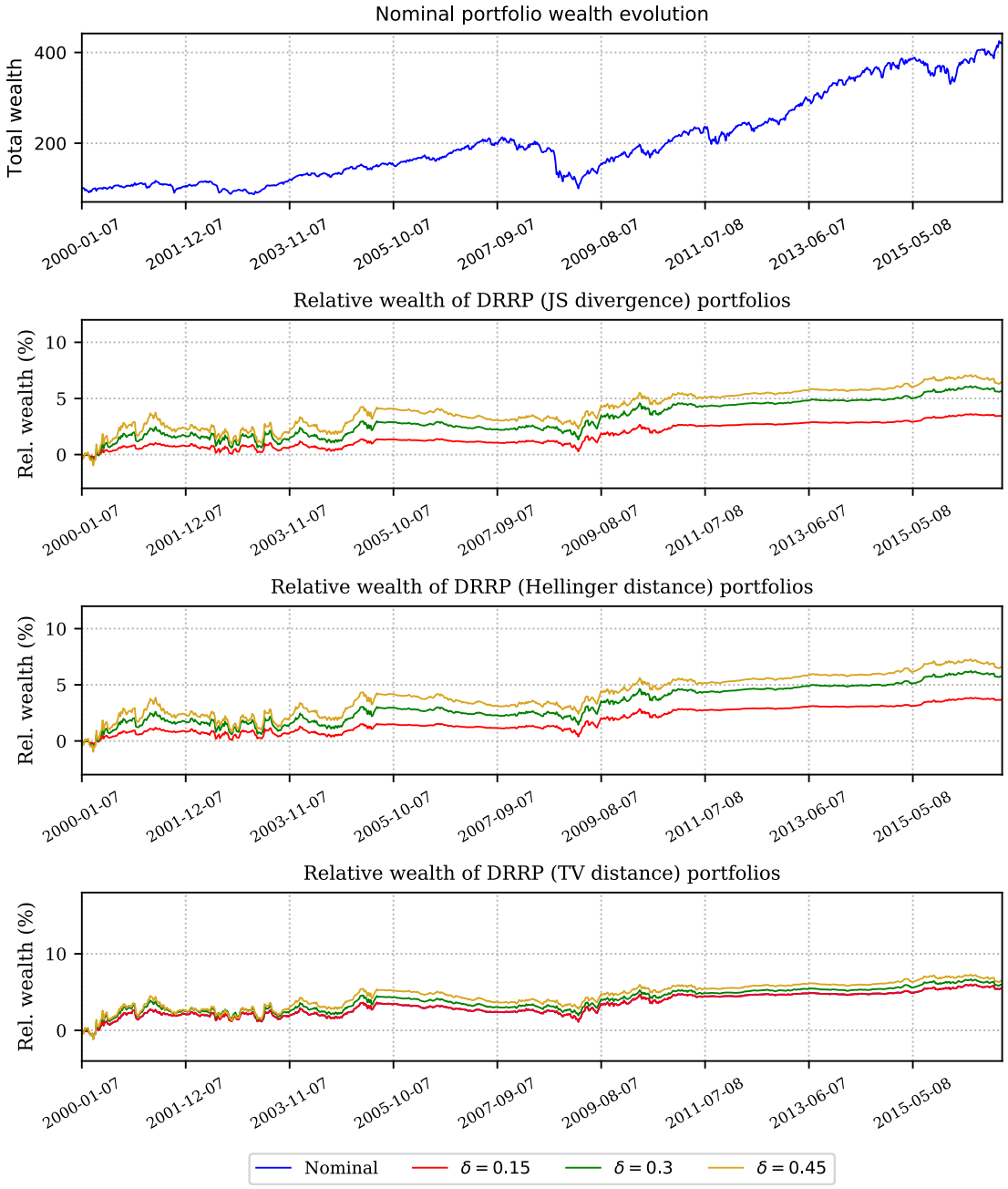

Figure 4.3: Wealth evolution for DRRP portfolios

Notes: The top plot shows the total wealth evolution of the nominal portfolio. The remaining three plots present the relative wealth evolution of the DRRP portfolios with respect to the nominal for varying confidence levels.

Table 4.5: Summary of financial performance of the risk parity portfolios over the periods 2000–2016 and 2007–2011

	Nom.	JS			Hellinger			TV		
$\delta =$		0.15	0.3	0.45	0.15	0.3	0.45	0.15	0.3	0.45
2000 – 2016										
Ann. Ex. Return (%)	6.64	6.84	6.96	7.01	6.85	6.97	7.02	6.95	6.98	7.01
Ann. Volatility (%)	17.0	17.2	17.2	17.2	17.2	17.2	17.2	17.2	17.2	17.2
Sharpe Ratio (%)	39.0	39.8	40.5	40.7	39.9	40.5	40.8	40.4	40.6	40.7
Avg. Turnover (%)	10.0	12.6	15.1	16.7	12.8	15.2	16.8	16.0	17.2	18.0
2007 – 2011										
Ann. Ex. Return (%)	2.39	2.67	2.77	2.73	2.69	2.77	2.73	2.71	2.66	2.62
Ann. Volatility (%)	23.3	23.7	23.8	23.8	23.7	23.8	23.8	23.9	23.8	23.8
Sharpe Ratio (%)	10.2	11.3	11.6	11.5	11.3	11.6	11.5	11.4	11.2	11.0
Avg. Turnover (%)	9.8	12.3	15.0	16.6	12.5	15.1	16.6	16.2	17.2	17.7

Notes: Nom, nominal. Ann, annualized. Ex, excess. Avg, average.

portfolio variance is maximized. Our modelling framework allows us to define the probability ambiguity set using any convex function to measure this statistical distance. We exemplify this by implementing three alternative statistical distances: JS, Hellinger, and TV.

The DRRP problem is a constrained convex–concave minimax problem over convex sets. We apply projected gradient methods to iterate over the DRRP problem in both descent and ascent directions. The projections ensure that we retain feasibility after each iteration. However, iteratively moving in both directions may lead to instability and slow convergence. Instead, we propose a novel algorithmic framework to solve our DRRP problem. The proposed SCP–PGA algorithm exploits the strict convexity of the risk parity problem, which guarantees that we have a unique risk parity portfolio for each instance of the adversarial probability distribution. Thus, we aim to iteratively ascend in the probability space through PGA while solving the corresponding convex risk parity problem after every iteration. The SCP–PGA algorithm dramatically improves computational runtime and, by design, retains the global convergence properties of general projected gradient methods. Our numerical results show that the SCP–PGA algorithm is computationally tractable and scalable. From a financial perspective, our experiments show that a DRRP portfolio is able to attain a higher risk-adjusted return when compared to the nominal portfolio.

Finally, we note that the DRRP problem can be adapted to solve other asset allocation

problems that may benefit from distributional robustness. Moreover, the general design of the SCP–PGA algorithm should allow it to solve other types of constrained convex–concave minimax problems, including problems in other disciplines. These topics are the subject of future research.

Chapter 5

A regime-switching factor model for risk parity

The purpose of this chapter is to improve the quality of the estimated input parameters used during optimization. To achieve this, we introduce a novel regime-switching factor model of asset returns that will allow us to incorporate an additional dimension of risk into the estimated parameters. Before delving into this topic, we first proceed by discussing the financial application of generic factor models, followed by regime-switching models.

Factor models attempt to explicitly explain the behaviour of a random variable either through a single factor, such as the capital asset pricing model [93, 68, 77], or through a combination of multiple factors, such as the Fama–French three-factor model [39]. Factor models have become popular in finance for the economic relevance of the factors and their ability to explain and quantify different sources of an asset's systematic risk. One important application of these models is to derive the estimated asset expected returns and covariance matrix, where the inherent properties of the factors are used to explain the asset returns and volatility.

Our intention is to maintain a traditional factor model structure that will allow us to naturally derive these two estimated parameters, while also introducing a regime-switching framework to capture the cyclical nature of financial markets. The original application of regime-switching in economics, proposed by Hamilton [55], was to describe the abrupt changes observed in business cycles. Due to their similarities, this concept was naturally extended to financial markets to describe a transition from a period of stability and growth to periods of financial distress, typically known as bullish or bearish market regimes. Existing literature suggests

a two-regime framework can effectively describe this cyclical behaviour [1, 2, 26, 29, 66, 81]. However, there is no consensus on the correct number of regimes. Guidolin and Timmermann [54] and Bae et al. [4] identify four separate market regimes to explain the joint distribution of asset returns.

This chapter maintains a purely data-driven approach. Since the overall signal-to-noise ratio of the financial market returns is low, our preference is to employ a two-regime framework. The changes in regime can be described through a Markov chain, a property that we exploit in the design of our proposed regime-switching factor model. A Markov chain that describes the transition between latent regimes is sometimes referred to as a hidden Markov model (HMM). Indeed, all the aforementioned literature use HMMs to model the cyclical behaviour of financial markets. An encompassing literature review on the use of regime-switching models in empirical finance is provided in [53]. In particular, this chapter follows the Markov process described by Hamilton [56] and Ang and Timmermann [3].

Preserving a linear structure allows us to naturally derive a set of regime-dependent asset expected returns and covariance matrix. Mathematically, this derivation has a similar complexity to the standard derivation from a traditional factor model. When used in optimization, the regime-dependency of the estimated parameters is implicitly incorporated into the resulting portfolio.

The result is a single-period regime-dependent portfolio that is tailored to the current market regime. 'Single-period' denotes the static nature of traditional portfolio optimization, where we find an optimal solution today to be used for all subsequent time periods, without consideration of any future transitions to other market regimes as time moves forward.

We could attempt to include future transitions stemming from our HMM conditions by using a dynamic programming approach for asset allocation in a multi-period setting [1, 81]. However, we favour a single-period approach to improve the computational tractability and scalability of our model, avoiding the 'curse of dimensionality' often associated with dynamic programming [8]. In other words, a single-period approach allows us to efficiently find optimal portfolios with a large number of constituent assets.

Furthermore, we can account for the transient nature of a regime-switching model by periodically re-estimating our parameters and rebalancing our portfolio, allowing the portfolio to align itself with the market over a long investment horizon. This, in turn, is compatible with traditional asset management practice, where institutions rely on rebalancing policies with fixed

calendar intervals, e.g., quarterly, semi-annually, or annually [76].

5.1 Regime-switching factor model

In this section we introduce a Markov regime-switching factor model of asset returns. As discussed in Section 2.1.2, the calibration of the standard factor model in (2.7) makes one important assumption: that recent historical data suffices to accurately estimate the future behaviour of the random asset returns $\boldsymbol{\xi}$.

As an example, imagine we wish to calibrate our factor model at the onset of the financial crisis of 2008. Using recent historical data to do this may provide a false sense of stability given the bullish market observed prior to the crisis. Thus, the standard factor model in (2.7) would fail to explain the abrupt change in behaviour of the asset returns.

Now, consider a market-dependent factor, such as the 'excess market return' factor from the Fama–French three-factor model[1] [39]. The aforementioned change in behaviour observed during the financial crisis of 2008 could be described as a transition into a different market regime. In turn, this transition can be described through a Markov process [3, 56]. Thus, we could assume that the factor itself is governed by a regime-dependent probability distribution that describes the observed factor returns.

We can extend our example and assume two market regimes exist: bullish and bearish. We can use a discrete-time Markov chain to describe the probability that our state variable s_t will be in regime j at time t given that our previous state was $s_{t-1} = i$, independent of any further preceding states, i.e.,

$$P\{s_t = j \mid s_{t-1} = i, s_{t-2} = k, \cdots\} = P\{s_t = j \mid s_{t-1} = i\}.$$

The state variable s_t is latent and corresponds to the market regime at time t. However, in practice, its probability distribution can be estimated from the observable market data through the Baum–Welch algorithm [7]. This algorithm is a special case of expectation maximization, an iterative technique used to find the maximum likelihood. In the case of our two-regime example,

[1] We note that all three factors in the Fama–French three-factor model are market-dependent, but we choose the 'excess market return' as the best example among the three factors to represent the market as a whole.

the transition matrix governing the probability of switching from one regime to another is

$$\boldsymbol{P} \triangleq \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}, \tag{5.1}$$

where p_{ij} is the probability that regime i at time t will be followed by regime j at time $t+1$. By the axioms of probability, each row must sum up to one, i.e., $\sum_{j=1}^{2} p_{ij} = 1$ for $i = 1, 2$.

Continuing with our two-regime example, the corresponding regime-switching model is structured as the superposition of two individual factor models

$$\boldsymbol{\xi} = \mathbb{1}_1 \cdot (\boldsymbol{\mu}^1 + [\boldsymbol{V}^1]^\top \boldsymbol{f}^1 + \boldsymbol{\epsilon}^1) + \mathbb{1}_2 \cdot (\boldsymbol{\mu}^2 + [\boldsymbol{V}^2]^\top \boldsymbol{f}^2 + \boldsymbol{\epsilon}^2), \tag{5.2}$$

where each factor model corresponds to a market regime and the superscripts indicate to which regime the parameters and variables belong. The indicator function is $\mathbb{1}_i = 1$ when the current regime is $s_t = i$ and $\mathbb{1}_i = 0$ if $s_t \neq i$. All the coefficients from the regression are assumed to be regime-dependent. Thus, we have that $\boldsymbol{\mu}^1, \boldsymbol{\mu}^2 \in \mathbb{R}^n$ are the regime-dependent asset expected returns, $\boldsymbol{V}^1, \boldsymbol{V}^2 \in \mathbb{R}^{m \times n}$ are the factor loadings under each regime, while $\boldsymbol{\epsilon}^1 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}^1) \in \mathbb{R}^n$ and $\boldsymbol{\epsilon}^2 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}^2) \in \mathbb{R}^n$ are the corresponding residuals. Accordingly, $\boldsymbol{D}^1, \boldsymbol{D}^2 \in \mathbb{S}_+^n$ denote the diagonal matrices of residual variance under each regime.

Moreover, the factors themselves are governed by this Markov chain. Therefore, the factor returns are described as $\boldsymbol{f}^1 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{F}^1) \in \mathbb{R}^m$ if we are under regime 1, or $\boldsymbol{f}^2 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{F}^2) \in \mathbb{R}^m$ if we are under regime 2. Accordingly, $\boldsymbol{F}^1, \boldsymbol{F}^2 \in \mathbb{S}_+^n$ denote the factor covariance matrices under each regime.

We will use (5.2) as our regime-switching factor model for the remainder of this chapter. The use of centred factors provides the added benefit of simplifying the subsequent derivation of the expectation, variance, and covariance of the random returns. These are derived as follows

$$\boldsymbol{\mu}^{s_i} \triangleq p_{i1}\ \boldsymbol{\mu}^1 + p_{i2}\ \boldsymbol{\mu}^2, \tag{5.3}$$

$$\begin{aligned} \boldsymbol{\Sigma}^{s_i} \triangleq\ & p_{i1}([\boldsymbol{V}^1]^\top \boldsymbol{F}^1 \boldsymbol{V}^1 + \boldsymbol{D}^1) + p_{i2}([\boldsymbol{V}^2]^\top \boldsymbol{F}^2 \boldsymbol{V}^2 + \boldsymbol{D}^2) \\ & + p_{i1}(1 - p_{i1})\boldsymbol{\mu}^1[\boldsymbol{\mu}^1]^\top + p_{i2}(1 - p_{i2})\boldsymbol{\mu}^2[\boldsymbol{\mu}^2]^\top - p_{i1}p_{i2}(\boldsymbol{\mu}^1[\boldsymbol{\mu}^2]^\top + \boldsymbol{\mu}^2[\boldsymbol{\mu}^1]^\top), \end{aligned} \tag{5.4}$$

where $\boldsymbol{\mu}^{s_i} \in \mathbb{R}^n$ and $\boldsymbol{\Sigma}^{s_i} \in \mathbb{R}^{n \times n}$ are the asset expected returns and covariance matrix corresponding to regime i, respectively. In addition, $p_{ij} = \mathbb{E}[\mathbb{1}_j \mid s_t = i]$, i.e. p_{ij} is the probability of

switching from regime i to regime j. These probabilities correspond to the two-state transition matrix in (5.1).

The benefit of this model is that the cyclical property of the market is captured implicitly through the regime-dependent covariance matrix in (5.4). In turn, we can construct a regime-dependent risk parity portfolio by using this covariance matrix during optimization. Since all the regime information is embedded in the covariance matrix, the optimization problem itself does not have any additional computational cost, i.e., the complexity of the regime-dependent risk parity optimization problem is the same as the nominal one.

The parameters in (5.3) and (5.4) lend themselves to single-period portfolio optimization. However, by design, these parameters are calibrated at time t to match the expectation of the current market regime, and do not take into consideration any future transitions from one regime to another. Thus, a limitation of this model is the lack of guarantees that our parameters will be properly aligned during any future time periods. Nevertheless, this limitation can be overcome through periodic re-estimation of the parameters, allowing them to re-align with the market.

5.1.1 Selection of market regimes

Here we discuss the number of regimes present in our model. Due to the signal-to-noise ratio of financial time series, determining the appropriate number of market regimes is a difficult task. Moreover, estimation procedures, such as the Baum–Welch algorithm [7], require that the user to define the following a priori: (i) the total number of regimes and (ii) the factor(s) used for estimation.

Our intention is not to suggest that a two-regime model is the most adequate for all scenarios, but rather it reflects our preference for parsimony. In direct reference to the argument presented by Ang and Bekaert [2], a two-regime specification is the most a data-driven approach can bear without significant computational problems in estimation, and it suffices to capture the main empirical differences in the asset behaviour. Other examples of two-regime models can be found in [1, 66, 81].

The number of factors used for estimation becomes an important consideration when we consider multi-factor models. As we will soon see, using all available factors in a model during the regime estimation process may not necessarily lead to a better-fitting model. Indeed, as we introduce more factors, we also introduce more noise during estimation. We note that, in the

case of multi-factor models, just because we feed a subset of the factors to the Baum–Welch algorithm does not mean we are discarding the remaining factors from the actual regression model.

For example, in the case of the Fama–French three-factor model, all three factors are market-dependent and are subject to the market's cyclical behaviour. However, we may have that the 'excess market return' factor alone conveys the latent market information more faithfully. Thus, we can feed this factor to the Baum–Welch algorithm to partition our time series into bull and bear periods. Once we identify which time periods correspond to each regime, we can use this information on all three factors to construct the regime-switching equivalent of the Fama–French model.

If desired, an investor with a predisposition towards a specific number of regimes and factors can modify the factor model in (5.2) to accommodate additional market regimes, and re-derive the regime-dependent parameters in (5.3) and (5.4) accordingly. Increasing the number of regimes has both potential benefits and drawbacks. In theory, increasing the number of regimes may allow us to detect additional nuances within the raw market data, providing a better-fitting model. However, this requires the estimation of additional parameters, which is computationally more expensive and also increases the potential for estimation error. The latter is particularly important due to the sensitivity of portfolio optimization to estimation errors. With this in mind, we note that there is existing literature advocating for a greater number of regimes [54, 4].

We proceed to justify our choice of number of regimes and factors through an empirical analysis. The analysis will rely on two criteria. The first is the Bayesian information criterion (BIC) Schwarz [89]. Briefly stated, the BIC is a measure of log-likelihood penalized by both the number of regimes and the number of factors used for estimation. The BIC is expressed as

$$\begin{aligned} a &\triangleq l \cdot b + l \cdot b^2 + l^2, \\ \text{BIC} &\triangleq -2\ln(L) + a\ln(T), \end{aligned} \tag{5.5}$$

where l is the number of regimes, b is the number of factors used for regime detection, L is the likelihood calculated from the Baum–Welch algorithm, and T is the total number of observations in the time series. A lower BIC value is preferable and indicates a better fitting model. The second criterion is more qualitative and relies on the inspection of the plot of smoothed probabilities from the Baum–Welch algorithm.

The analysis is performed on monthly observations of the three Fama–French factors [39] over the time period from Jan–1983 to Dec–2016. The data were obtained from Kenneth R. French's data library [42]. This is the same data we will also use in Section 5.3 for the numerical experiments. For now, we perform this empirical analysis to justify our choice of regimes and factors.

Table 5.1 presents the BIC of Markov models with varying numbers of regimes and factors. For reference, the 'Market' factor refers to the 'excess market return' factor, the 'Value' factor refers to the 'high-minus-low' factor, and the 'Size' factor refers to the 'small-minus-big' factor. The results shows that a Markov model with two regimes and a single factor is the most desirable since it has the lowest BIC.

Table 5.1: BIC values

	Mkt	Mkt and Value	Mkt and Size	All
2 regimes	34.9	81.4	81.4	152.6
3 regimes	77.0	147.5	147.5	254.8
4 regimes	131.1	225.6	225.6	369.0

Notes: A lower BIC value is preferable. Mkt, market.

In general, Table 5.1 strongly suggests that a single-factor model provides the best fit. Thus, Figure 5.1 shows several plots of the smoothed probabilities for multiple regimes corresponding solely to the Market factor. It is interesting to see that 'Regime 1' is relatively consistent between the three plots. However, this consistency is lost when we inspect the remaining regimes. In particular, the noise observed in the three- and four-regime models reinforces that a two-regime model provides the most stable fit. Again, citing our preference for parsimony, a two-regime model suffices to capture the main empirical differences between regimes. Thus, for the remainder of this chapter, we assume a two-regime environment.

5.2 Regime-dependency in risk parity

We can embed the regime-dependency into a risk parity portfolio by using the regime-dependent covariance matrix. For example, we can use the Baum–Welch algorithm to partition our asset and factor return data, $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\phi}}$, into two distinct datasets corresponding to each regime. Moreover, the Baum–Welch algorithm also yields the transition probability matrix $\boldsymbol{P}$.

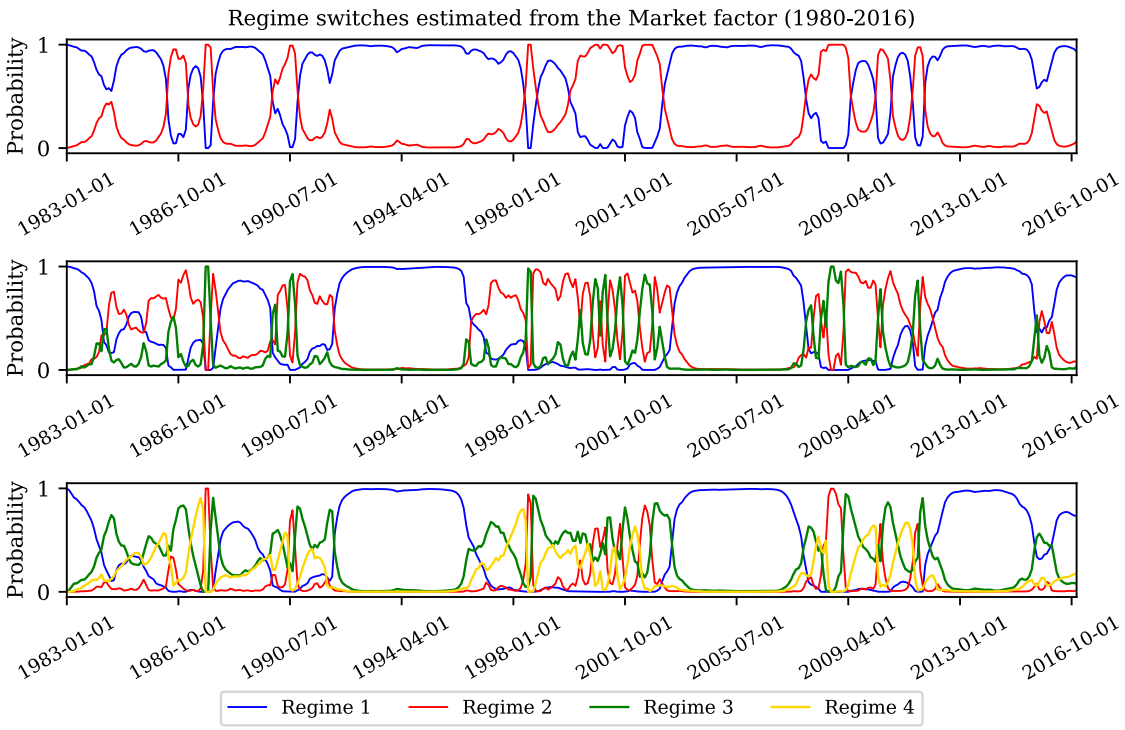

Figure 5.1: Estimated regime switches from two-regime (top), three-regime (center), and four-regime (bottom) Markov models

The factor model in (5.2) is composed of two distinct factor models, each corresponding to a different regime. Thus, we can take similar steps to those outlined in (2.8–2.10) from Section 2.1.2 to estimate the regime-specific parameters $\hat{\boldsymbol{\mu}}^i$, $\hat{\boldsymbol{V}}^i$, $\hat{\boldsymbol{F}}^i$ and $\hat{\boldsymbol{D}}^i$ for $i = 1, 2$. Finally, we can use the estimated parameters in conjunction with the transition probabilities to calculate $\hat{\boldsymbol{\Sigma}}^{s_i}$ as shown in (5.4).

Given that the regime-specific information is embedded within $\hat{\boldsymbol{\Sigma}}^{s_i}$, we do not need to modify the risk parity optimization problem itself. Thus, it suffices to use any of the long-only risk parity models from Section 2.2. For example, we can construct a regime-dependent risk parity portfolio by solving the convex problem in (2.19) using the covariance matrix $\hat{\boldsymbol{\Sigma}}^{s_i}$.

Mathematically, the regime-dependent risk parity optimization model has the same structure as its nominal counterpart, meaning it retains the same level of complexity. Simply put, the contribution from this chapter is to improve risk parity by focusing on improving the estimation of parameters rather than the optimization problem itself.

5.3 Numerical experiments

This section presents the results of a numerical experiment to test the out-of-sample performance of a regime-dependent risk parity portfolio. We chose the Fama–French three-factor model [39] as the basis to test the regime-switching framework due to its popularity as a trustworthy multiple factor model. However, we note that this framework is applicable to any model with market-dependent factors.

The experimental data is composed of monthly observations ranging from Jan–1983 until Dec–2016. The factor data were obtained from Kenneth R. French's data library [42]. The asset data consists of a basket of 52 diverse U.S. equities with a few representatives from each of the 11 *Global Industry Classification Standard* (GICS) sectors. In particular, these stocks were selected because of their data availability over the period from 1983 to 2016. Thus, there are a total of 52 stocks in our portfolio ($n = 52$), and their tickers are shown in Table 5.2. The historical stock prices were obtained from *Yahoo! Finance*.

The length of the time series required for the experiments introduces a survivorship bias to our results since we are only able to select stocks with sufficient historical data. However, we expect the bias to have a similar effect on all the portfolios tested, and it is the relative performance of the different risk parity portfolio strategies that are of main interest.

Table 5.2: List of assets

GICS Sector	**Company Tickers**					
Communication	DIS	NYT				
Consumer Disc.	F	FL	GT	MCD		
Consumer Staples	CL	KO	KR	PEP	PG	WMT
Energy	APA	CVX	HAL	MRO	XOM	
Financials	AXP	BK	C	WFC		
Healthcare	BMY	CHE	JNJ	LLY	MRK	PFE
Industrials	BA	CAT	GE	MMM	FDX	LMT
Information Tech.	HPQ	IBM	MSI	TXN	XRX	
Materials	DD	DOW	IP	LXU		
Real Estate	BRT	GTY	PEI	WY		
Utilities	AEP	CNP	DTE	ED	PCG	PEG

An overview of the experimental setup follows. The experiment involves four different portfolios: (*i*) a nominal risk parity portfolio, (*ii*) a robust risk parity portfolio built using the framework from Chapter 3, (*iii*) a regime-switching risk parity portfolio, and (*iv*) a robust

regime-switching risk parity portfolio that combines the robust framework from Chapter 3 with the derivation of regime-dependent estimated parameters from this chapter.

The out-of-sample investment period ranges from Jan–2003 until Dec–2016. We conduct three individual trials over this period, each with a different rebalancing policy. The portfolios are rebalanced every three, six, or twelve months. Our choice of rebalancing frequency is motivated by typical choices made by investors [76]. For all three trials, the in-sample calibration window is rolled forward every time a portfolio is rebalanced in order to use the most recently available data.

The in-sample calibration of both the nominal and robust portfolios is based on ordinary least squares regression and use two years of historical data (24 observations) immediately preceding the current investment period. All estimated parameters are re-calibrated every time the portfolios are rebalanced. In addition, the two robust portfolio uses a value of $\omega = 0.1$ to size their uncertainty sets (see Chapter 3 for information on the parameter ω).

The estimation of the regime-dependent parameters requires additional historical data to properly calibrate the Markov model. We use historical factor returns starting from Jan-1983, 20 years before the start of the out-of-sample test period. We apply the Baum–Welch algorithm [7] to find both the transition probabilities and the smoothed probabilities, with a practical implementation of this algorithm derived from [62]. As discussed in Section 5.1.1, we use the 'excess market return' factor as the sole input to the Baum–Welch algorithm and we assume a two-regime model. We use the resulting smoothed probabilities to differentiate between bullish and bearish time periods. This data-driven approach allows us to select periods of historical factor and asset returns corresponding to each regime. For consistency, the most recent 24 observations corresponding to each regime are used to perform the regression. We note that these 48 observations stretch beyond the immediately preceding two years of historical data, and go as far back in time as necessary to ensure the availability of sufficient data for both regimes. We favour this approach in order to guarantee a consistent number of observations per regime, improving the stability of the regression model. The appropriate transition probabilities p_{i1} and p_{i2} are selected by identifying the 'current' regime, i.e., by inspecting the most recent observation of the smoothed probabilities and selecting the largest of the two probabilities. After the two regime-dependent regression models are calibrated, and together with the appropriate transition probabilities, the asset expected returns and covariance matrix are estimated as shown in (5.4). This calibration process is repeated every time the portfolios are rebalanced. Finally, we note

that the regime estimation window, which is originally 20 years long, expands as we move forward in time, i.e., the regime estimation window that we feed to the Baum–Welch algorithm grows as the experiment moves forward through time.

All experiments were conducted on an Apple MacBook Pro computer (2.8 GHz Intel Core i7, 16 GB 2133 MHz DDR3 RAM) running macOS 'Catalina'. The computer script was written in the Julia programming language (version 1.4.0) using the modelling language 'JuMP' [33] with IPOPT (version 3.12.6) as the optimization solver.

An example of the estimated changes in regime from this experiment is presented in Figure 5.2, which shows the smoothed probabilities of a two-regime model prior to the start of the first investment period (top plot) and the last investment period (bottom plot). We reiterate that, for the purpose of regime estimation, the Baum–Welch algorithm is applied only on the 'excess market return' factor. This data-driven process may sometimes result in inconsistent estimates as more data becomes available, i.e., a period that was once considered to be bullish may be considered bearish if new data exacerbates the differences between the two regimes.

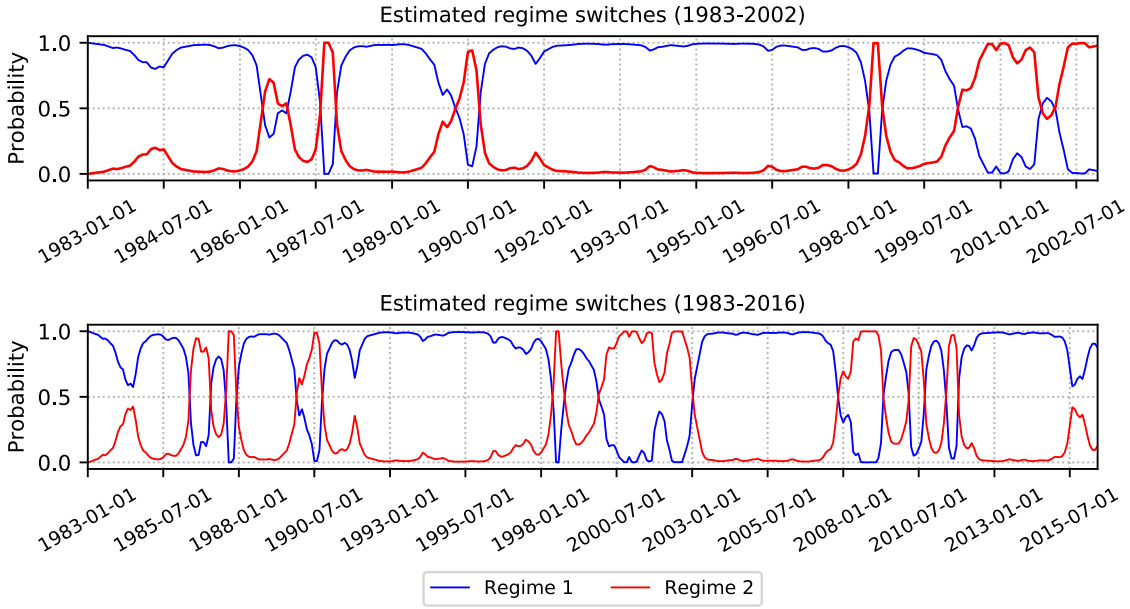

Figure 5.2: Estimated regime switches

Note: Regime 1 corresponds to a bullish market and regime 2 corresponds to a bearish market. The top plot corresponds to the calibration window before the first investment period. The bottom plot corresponds to the calibration window before the last investment period.

The portfolio wealth evolution is presented in Figure 5.3. Within the figure, a white or shaded background indicates whether the regime-switching portfolios were calibrated to align

with a bullish or bearish market regime, respectively. We note that both the white and shaded periods are chosen on an ex-ante basis by the Baum–Welch algorithm, i.e., the model determines the 'current' regime using only data available at the time of calibration, and is blind from any future data. We note that the regime-switching portfolios tend to align well with observed historical periods of market distress. For example, they are aligned to the bear market regime of the early 2000's recession and the financial crisis of 2008.

The plots of the three-, six-, and twelve-month rebalancing policies in Figure 5.3 show that the regime-switching portfolio is able to attain a higher terminal wealth than all other competing portfolios. On the other hand, the robust portfolio appears to fail to take advantage of the prolonged bull market periods to generate greater returns, which are approximately between 2003–2008 and 2012–2017.

Nevertheless, Table 5.3 shows that the robust portfolio has a lower volatility than either the nominal or regime-switching portfolios. However, even after adjusting for risk, the excess return of the robust portfolio is the lowest between all competing portfolios, as shown by the corresponding Sharpe ratios.

A more interesting behaviour is observed in the robust regime-switching portfolio. As the name entails, this portfolio is built by applying the robust framework from Chapter 3 to the regime-dependent partitions of the regime-switching factor model. This also allows us to derive the probability weighted perturbation of the covariance matrix. The results in Table 5.3 suggest that the robust regime-switching portfolio combines the best properties of both the regime-switching framework and the robust framework. Its alignment with the market allows it to attain a reasonable rate of return, while its robustness maintains a relatively low volatility over the entire out-of-sample period. Thus, the Sharpe ratio of the robust regime-switching portfolio is superior to all competing portfolio under every single rebalancing policy tested.

The average period-over-period turnover rate does highlight the weakness of the robust portfolios. As discussed in Chapter 3, the robust portfolios have a higher turnover rate when compared to the nominal. Table 5.3 shows the same is true of the robust regime-switching portfolio. On the other hand, the non-robust regime-switching portfolio shows a slightly lower average turnover rate when compared to the nominal, meaning it has the lowest turnover between all four portfolios. This stems from the stability of the regime-switching portfolio while we remain in the same market regime, i.e., our change in weights is small if we do not transition from one market regime to another. Nevertheless, we reiterate our previous argument

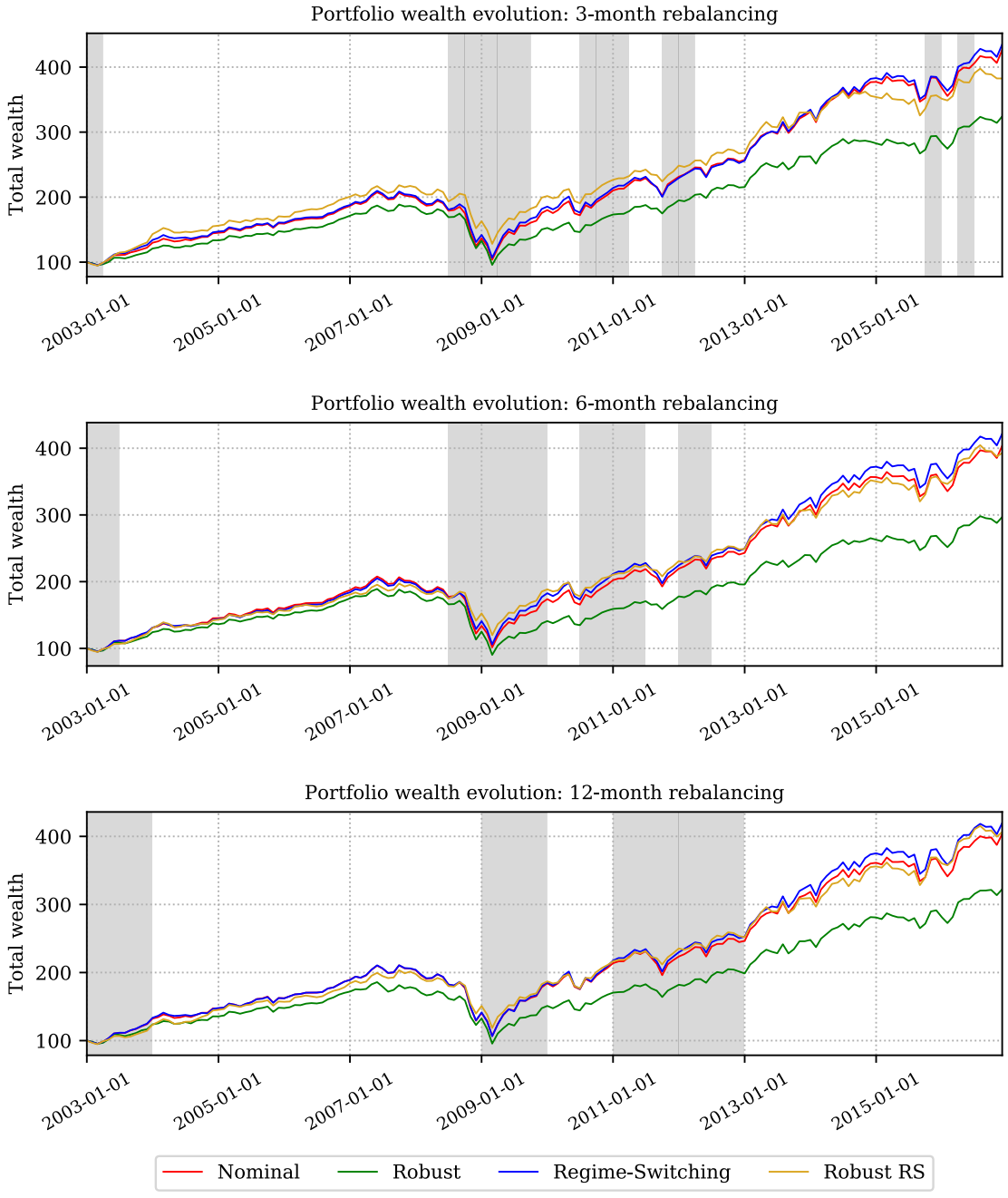

Figure 5.3: Portfolio wealth evolution

Notes: A shaded background indicates time periods where the regime-switching portfolios were calibrated for a bearish market regime. A white background indicates calibration for a bullish market regime.

that transaction costs in modern financial markets are very low when trading assets with high liquidity. Thus, for the most part, transaction costs are becoming increasingly negligible.

Table 5.3: Summary of results

	3-month				6-month				12-month			
	Nom.	Rob.	RS	RRS	Nom.	Rob.	RS	RRS	Nom.	Rob.	RS	RRS
Ann. Ex. Return (%)	9.7	7.5	9.8	8.8	9.3	6.9	9.6	9.0	9.3	7.5	9.6	9.3
Ann. Volatility (%)	14.9	13.6	15.0	13.0	14.8	14.0	14.7	12.6	14.8	13.6	14.9	13.0
Sharpe Ratio (%)	64.9	55.5	65.9	67.9	62.5	49.1	65.3	72.1	62.5	55.1	64.0	71.8
Avg. Turnover (%)	22.6	51.6	21.6	59.3	30.2	70.8	27.7	68.5	40.4	90.8	37.2	91.3

Notes: Ann, annualized. Ex, excess. Nom, nominal. Rob, robust. RS, regime-switching. RRS, robust regime-switching.

5.4 Conclusion

This chapter introduced a regime-switching factor model that reflects the cyclical nature of asset returns in modern financial markets. This model retains two important properties of a traditional factor model: the economic relevance of the factors, and the ability to easily derive the asset expected returns and covariance matrix. Deriving these parameters through the regime-switching factor model implicitly captures the cyclical information of the market through a probability-weighted approach, adapting the parameters to the current market regime.

In turn, using these parameters during risk parity portfolio optimization positions the resulting optimal portfolio to take advantage of this directional information, resulting in a portfolio with reduced ex-post volatility and increased returns. The novelty of the proposed framework is the seamless integration of the regime-dependency of the asset returns with the static nature of risk parity portfolio optimization, reconciling it with the dynamism of the market through periodic portfolio rebalancing.

The regime-switching factor model can be easily derived from any generic model where at least one of the factors observes the cyclical behaviour of the market. Explaining the transition between market regimes through a discrete-time Markov chain allows us to retain a data-driven modeling structure where we can use historical observations to calibrate our model and estimate the regime-dependent asset expected returns and covariance matrix. Thus, this aligns well with a single-period model such as risk parity portfolio optimization. Since regime-dependency is

already captured through the estimated parameters, we highlight that this framework comes at no additional computational cost during optimization. In general, this single-period approach is well suited for traditional asset management techniques, where portfolios are periodically rebalanced and there is a long-term investment horizon.

Moreover, since this factor model embeds the regime-dependency within the estimated parameters, we are able to use it in conjunction with other optimization frameworks. As exemplified in Section 5.3, we apply the robust framework proposed in Chapter 3 to create a robust regime-switching risk parity portfolio. This highlights both the flexibility and power of the different tools we are building throughout this thesis.

The experimental results show that a regime-switching portfolio is able to steadily outperform its nominal counterpart over a long investment horizon, exhibiting higher risk-adjusted returns. These results are consistent over three different rebalancing policies of varying lengths, showing that regime-switching portfolios are quite resilient. If rebalancing takes place periodically, our portfolio is able to adapt itself to the current market regime.

Chapter 6

A generalized risk parity framework

This chapter introduces a generalized risk parity (GRP) framework. This framework blends the desirable properties of MVO and risk parity. Specifically, this optimization problem maintains the MVO objective function, allowing us to optimize a desirable risk–return profile. However, we avoid the concentration of risk by enforcing a user-defined level of risk-based diversification.

The GRP framework can be interpreted from two different points of view. From a risk parity perspective, we relax the risk parity condition so that we can partially control the level of portfolio risk and expected return. From a MVO perspective, we impose a predefined limit on the portfolio risk dispersion, maintaining a desired level of risk-based diversification while giving the optimization problem flexibility to attain a more desirable objective value.

We extend this framework by considering a robust formulation of the asset expected returns in an effort to mitigate the impact of estimation error. Additionally, our framework does not impose any restrictions on short sales. However, allowing for 'long–short' portfolios while constraining individual risk contributions means this is a non-convex optimization problem, which is an issue we must address.

Managing the non-convexity of both risk parity and risk budgeting is still the subject of ongoing research. Feng and Palomar [40] propose an algorithm that sequentially solves a series of convex problems by using a first-order approximation to the original non-convex risk parity problem. Their research is centred around the nominal risk parity problem, and their first-order formulation does not allow for the type of non-convex quadratic constraint we will need to bound the asset risk dispersion. Moreover, their formulation is able to guarantee global convergence to a stationary point, but cannot make any claims about the quality of this local

optimum relative to all other extant optimal risk parity solutions.

Haugh et al. [57] introduce a generalized risk budgeting framework that seeks to maximize the risk–return utility of an investor while having binding risk contribution constraints. However, an investor must predetermine the exact risk budgets assigned to each asset (or basket of assets) before optimization takes place, and this restricts flexibility since the risk budgets are imposed exactly through equality constraints. To overcome the issue of non-convexity, Haugh et al. introduce an algorithm that combines the augmented Lagrangian and Markov chain Monte Carlo methods. This heuristic approach relies on sampling different starting points from the feasible set, and then proceeding to solve the non-convex problem repeatedly using traditional non-linear optimization algorithms. The best available solution is then chosen from the subset of optimal solutions.

This chapter proposes a different approach, which relies on sequentially tightening a semidefinite program (SDP) relaxation of the original non-convex problem. By definition, SDPs are convex. However, the solution to a SDP relaxation can only provide a lower bound to the global optimum. In the case of risk parity, the solution of the SDP relaxation generally violates any constraints pertaining to the asset risk contributions. A feasible solution can be attained if we impose a non-convex rank-1 constraint on the SDP. Imposing a non-convex constraint on our SDP might seem counterproductive, but proceeding in this fashion allows us to decompose this into two sub-problems: a convex SDP, and a non-convex rank-1 approximation. The first step involves solving a modified version of our SDP relaxation, while the second step requires us to find the closest rank-1 solution to the first step. Although non-convex, the second step has a closed-form expression that allows us to efficiently find an exact solution. We proceed to implement the alternating direction method of multipliers (ADMM) algorithm [46, 48, 18]. We favour the application of ADMM because, through the tightening of our relaxation, we approximate the global optimum from its lower bound. Relaxing a non-convex into a convex SDP and using ADMM to sequentially impose a rank-1 constraint is not a new approach, and has been successfully applied in optimal power flow problems [99].

This chapter proposes a similar method based on an ADMM algorithm to solve the non-convex GRP problem. This algorithm sequentially tightens the relaxed SDP, converging to a rank-1 constrained SDP solution, which is equivalent to our non-convex formulation. The sequential tightening of the relaxed feasible set ensures we approach our optimal solution from a lower bound, converging to a higher quality optimal solution. Thus, the contribution of this

chapter is twofold. First, we propose a GRP framework with an objective function that reflects an investor's risk–return profile while maintaining a desired degree of risk-based diversification. The resulting non-convex formulation can then be solved sequentially via the ADMM algorithm to attain a higher quality solution when compared to local optimization techniques.

6.1 Generalized risk parity

This section presents the GRP framework. This 'generalization' arises from our motivation to design a model that satisfies MVO users wishing to embed the desirable risk-based diversification properties of risk parity while still retaining control over their risk–return objective. Moreover, this also allows risk parity users to relax their risk-based diversification objective, enabling them to account for the overall risk and expected return of their respective portfolio.

Specifically, our proposed GRP framework is designed so investors can attain an optimal risk–return profile while maintaining a desirable degree of risk-based diversification. Moreover, it provides the flexibility to construct long–short portfolios. This falls in line with other generalized risk budgeting frameworks [57], but we note an important distinction proposed in this thesis. This framework allows the investor to preset a desired range of risk-based diversification instead of having to preset fixed risk budgets for its assets. This provides a degree of control over the dispersion of risk within the assets while offering greater flexibility to attain a more desirable risk–return profile.

Our objective is to allow an investor to impose a limit on the dispersion of the individual asset risk contributions, thereby promoting risk-based diversification throughout the portfolio. As discussed in Section 2.2, we have that the individual asset risk contributions are corresponding to the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$,

$$\hat{r}_i \triangleq x_i[\hat{\boldsymbol{\Sigma}}\boldsymbol{x}]_i = \boldsymbol{x}^\top \hat{\boldsymbol{R}}^i \boldsymbol{x}.$$

where the collection of matrices $\hat{\boldsymbol{R}}^i \in \mathbb{S}^n$ for $i = 1, \ldots, n$ are partitions of the estimated asset covariance matrix, as shown in (2.17).

A simple yet tractable method to achieve this is to impose a limit on the difference between the highest and the lowest values of $\hat{r}_i$. We can do this by imposing an additional $2n$ constraints

over the nominal MVO problem in (2.13). Thus, the GRP problem is the following

$$\min_{\boldsymbol{x},\zeta} \quad \boldsymbol{x}^\top \hat{\boldsymbol{\Sigma}} \boldsymbol{x} - \lambda \hat{\boldsymbol{\mu}}^\top \boldsymbol{x} \tag{6.1a}$$

$$\text{s.t.} \quad (1+c)\zeta - \boldsymbol{x}^\top \hat{\boldsymbol{R}}^i \boldsymbol{x} \geq 0, \quad i = 1, \ldots, n, \tag{6.1b}$$

$$\boldsymbol{x}^\top \hat{\boldsymbol{R}}^i \boldsymbol{x} - (1-c)\zeta \geq 0, \quad i = 1, \ldots, n, \tag{6.1c}$$

$$\mathbf{1}^\top \boldsymbol{x} = 1, \tag{6.1d}$$

where $\hat{\boldsymbol{\mu}} \in \mathbb{R}^n$ are the estimated asset expected returns, $\zeta \in \mathbb{R}$ is an auxiliary variable that serves as a placeholder for the midpoint between the highest and the lowest asset risk contributions, and $c \in \mathbb{R}_+$ is a preset risk diversification coefficient that serves to bound the dispersion of the asset risk contributions. We note that setting $c = 0$ enforces perfect risk parity, while $c \geq 1$ collapses the problem to the nominal MVO problem. Finally, the last constraint ensures all available wealth is invested. To the best of our knowledge, this is the first model to propose bounds on the dispersion of the asset risk contributions such that we relax the risk parity condition. This enforces a desirable level of risk-based diversification in our portfolio.

Constraints (6.1b) and (6.1c) are the $2n$ constraints that restrict our risk dispersion, i.e., these constraints ensure we retain our desired level of risk-based diversification. However, since the matrices $\boldsymbol{R}^i$ are indefinite, the GRP problem in (6.1) is non-convex. In turn, this means the problem is likely to converge to a local optimum. The non-convexity of this model is addressed in Section 6.2, where we propose an algorithm to overcome this problem. For now, the remainder of this section presents two extensions of our GRP framework.

6.1.1 Robust generalized risk parity

As we have discussed in Section 2.1.3, the impact of estimation errors in the expected returns have an impact ten times larger than estimation errors in the covariance matrix during optimization [25]. In turn, this is one of the drivers that motivates risk parity, a modern asset allocation framework that does not necessitate the expected returns as an input, avoiding the pitfalls arising from uncertainty in estimated expected returns.

However, our proposed GRP framework allows an investor to consider both portfolio risk and return. To mitigate the impact of uncertainty, we impose a robust structure on the estimated expected returns. We start by introducing an ellipsoidal uncertainty set around the true (but

unknown) expected returns as

$$\mathcal{U}_{\boldsymbol{\mu}} \triangleq \left\{ \boldsymbol{\mu} \in \mathbb{R}^n : \ (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\top \boldsymbol{\Theta}^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \theta^2 \right\},$$

where $\boldsymbol{\Theta} \in \mathbb{S}_+^n$ is the covariance matrix of estimation errors and $\theta \in \mathbb{R}_+$ is a measure of distance that scales the uncertainty set in proportion to some probabilistic guarantee. This uncertainty set states that the sum of squared deviations between the nominal estimate $\hat{\boldsymbol{\mu}}$ and any other point in the set cannot be greater than θ^2. Instead of having individual confidence intervals for each estimated expected return, $\hat{\mu}_i$, an ellipsoidal uncertainty set describes a joint confidence region [38].

The parameters $\boldsymbol{\Theta}$ and θ quantify the error from estimation and the size of the uncertainty set, respectively. Different methods are available to calibrate these two parameters. One example on how to do this is shown in [38] and is the method we will use for the remainder of this chapter. The matrix $\boldsymbol{\Theta}$ can be defined as the diagonal matrix of squared standard errors, i.e.,

$$\boldsymbol{\Theta} \triangleq \frac{1}{T} \text{diag}(\hat{\boldsymbol{\Sigma}}),$$

where T is the number of observations used during estimation and the operator $\text{diag}(\cdot)$ creates a diagonal matrix by setting every off-diagonal element to zero. If we approximate the joint confidence region as a sum of the squares of independent standard normal variables, we can use a χ^2-distribution with n degrees of freedom to determine θ^2, i.e.,

$$\theta^2 \triangleq \chi_n^2(\delta),$$

where δ is the confidence level that the true vector of expected returns, $\boldsymbol{\mu}$, is within the set $\mathcal{U}_{\boldsymbol{\mu}}$. For example, a popular choice in finance is to assign $\delta = 0.95$ to have a 95% confidence interval. However, in practice, using a χ^2-distribution may be prohibitively conservative. Additional detail on this calibration method, as well as other alternative methods, are shown in [38].

The robust counterpart of the GRP problem is

$$\min_{\boldsymbol{x},\zeta} \quad \boldsymbol{x}^\top \hat{\boldsymbol{\Sigma}} \boldsymbol{x} - \lambda\big(\hat{\boldsymbol{\mu}}^\top \boldsymbol{x} - \theta \|\boldsymbol{\Theta}^{1/2}\boldsymbol{x}\|_2\big) \tag{6.2a}$$

$$\text{s.t.} \quad (1+c)\zeta - \boldsymbol{x}^\top \hat{\boldsymbol{R}}^i \boldsymbol{x} \geq 0, \quad i = 1, \ldots, n, \tag{6.2b}$$

$$\boldsymbol{x}^\top \hat{\boldsymbol{R}}^i \boldsymbol{x} - (1-c)\zeta \geq 0, \quad i = 1, \ldots, n, \tag{6.2c}$$

$$\mathbf{1}^\top \boldsymbol{x} = 1. \tag{6.2d}$$

The ℓ_2-norm term in (6.2a) quantifies the ellipsoidal uncertainty around the estimated portfolio return, giving us the robust expression of the portfolio return.

The proposed ellipsoidal uncertainty structure is only one example of how to incorporate robustness. Indeed, many other alternatives exist in portfolio optimization, such as the use of box constraints [96], or even robustness derived from the estimation errors arising from a factor model structure [49]. Many of these examples can be incorporated into the GRP framework, provided the complexity of the robust structure is linear or quadratic.

6.1.2 Lowest-variance risk parity

Thus far, the GRP framework has emphasized two benefits over the nominal risk parity problem. First, it allows an investor to optimize with respect to a predetermined risk–return profile rather than simply equalizing the asset risk contributions. Second, it allows for the construction of long–short portfolios.

However, in certain scenarios, an investor who lacks confidence or is unable to produce reliable estimates of the asset expected returns might wish to ignore portfolio returns altogether and focus solely on risk-based diversification. Such an investor would revert back to the nominal risk parity framework. As we saw in Section 2.2.1, there exist up to 2^{n-1} risk parity solutions, except some of these solutions may be more desirable than others if they have a lower portfolio variance. Indeed, the concept of the lowest variance risk parity (LVRP) portfolio is not new. For example Bai et al. [5] propose a heuristic algorithm that iteratively moves from the minimum variance portfolio towards the nearest risk parity portfolio in an effort to find a risk parity solution with low variance.

Alternatively, we can use the GRP framework to solve this problem. We do this by setting the trade-off coefficient and the risk diversification coefficient to zero in (6.1), i.e., we set $\lambda = c = 0$.

Therefore, our LVRP framework is

$$\min_{\boldsymbol{x},\zeta} \quad \boldsymbol{x}^\top \hat{\boldsymbol{\Sigma}} \boldsymbol{x} \tag{6.3a}$$

$$\text{s.t.} \quad \boldsymbol{x}^\top \hat{\boldsymbol{R}}^i \boldsymbol{x} = \zeta, \quad i = 1, \ldots, n, \tag{6.3b}$$

$$\mathbf{1}^\top \boldsymbol{x} = 1. \tag{6.3c}$$

After setting $c = 0$, we can see that the $2n$ inequality constraints that bound the individual risk contributions collapse to n equality constraints, enforcing the risk parity condition. As with (6.1) and (6.2), the LVRP problem is also non-convex. Nevertheless, the issue of non-convexity can be overcome through an ADMM algorithm as we will see in the section.

6.2 Sequential tightening via ADMM

This section presents a heuristic algorithm that sequentially approximates a global optimal solution to the non-convex GRP problem. We start by relaxing the problem into a SDP. We then re-introduce a non-convex rank-1 constraint into the SDP, and proceed to decompose the problem into a convex sub-problem and a non-convex sub-problem. Finally, we present the complete ADMM algorithm for our application.

6.2.1 SDP relaxation

We begin by relaxing the GRP problem in (6.1) into a SDP. To do so, we introduce a new variable $\boldsymbol{X} \in \mathbb{S}_+^n$ as a convex relaxation of the square of our asset weights $\boldsymbol{x}$ such that $\boldsymbol{X} \succeq \boldsymbol{x}\boldsymbol{x}^\top$. We then introduce a new decision variable $\boldsymbol{Y} \in \mathbb{S}_+^{n+1}$ as an expression of the Schur complement between $\boldsymbol{x}$ and $\boldsymbol{X}$. Finally, we reformulate the input parameters to align with the dimensions of $\boldsymbol{Y}$. Therefore, we have that

$$\boldsymbol{Y} \triangleq \begin{bmatrix} \boldsymbol{X} & \boldsymbol{x} \\ \boldsymbol{x}^\top & 1 \end{bmatrix}, \quad \boldsymbol{Q} \triangleq \begin{bmatrix} \hat{\boldsymbol{\Sigma}} & -\frac{\lambda}{2}\hat{\boldsymbol{\mu}} \\ -\frac{\lambda}{2}\hat{\boldsymbol{\mu}}^\top & 0 \end{bmatrix}, \quad \boldsymbol{C}^i \triangleq \begin{bmatrix} \boldsymbol{R}^i & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{bmatrix} \text{ for } i = 1, \ldots, n.$$

Expressing our variables and input parameters in this fashion allows us to relax the GRP

problem in (6.1) into the following SDP,

$$\min_{\boldsymbol{Y},\zeta} \quad \mathrm{Tr}(\boldsymbol{Q}\boldsymbol{Y}) \tag{6.4a}$$

$$\text{s.t.} \quad (1+c)\zeta - \mathrm{Tr}(\boldsymbol{C}^i\ \boldsymbol{Y}) \geq 0, \quad i = 1, \ldots, n, \tag{6.4b}$$

$$\mathrm{Tr}(\boldsymbol{C}^i\ \boldsymbol{Y}) - (1-c)\zeta \geq 0, \quad i = 1, \ldots, n, \tag{6.4c}$$

$$\sum_{i=1}^{n} Y_{i,n+1} = 1, \tag{6.4d}$$

$$Y_{n+1,n+1} = 1, \tag{6.4e}$$

$$\boldsymbol{Y} \succeq 0, \tag{6.4f}$$

where $\mathrm{Tr}(\cdot)$ is the trace operator. For reference, we will refer to the convex set formed by constraints (6.4b–6.4f) as the set $\mathcal{S}_{\mathrm{GRP}}$.

By definition, (6.4) is a convex relaxation of (6.1). However, it is not guaranteed to respect our predefined limit on the dispersion of the asset risk contributions. We can address this by imposing a rank-1 constraint on (6.4) to recover our original non-convex GRP problem. In other words, adding the non-convex constraint $\mathrm{rank}(\boldsymbol{Y}) = 1$ is equivalent to imposing that $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^\top$, ensuring that our risk dispersion constraints are not violated. Thus, essentially, the SDP relaxation works by removing the rank-1 condition, thereby relaxing the non-convex feasible set from (6.1) into the convex set $\mathcal{S}_{\mathrm{GRP}}$.

6.2.2 Problem decomposition and various ADMM steps

Our proposed approach attempts to attain a rank-1 solution to the original GRP problem in (6.1) by sequentially solving a series of convex problems based on the SDP relaxation in (6.4), with each iteration tightening our approximation to a rank-1 solution. We apply the ADMM algorithm to decompose our problem and sequentially solve it.

ADMM was originally proposed by Glowinski and Marroco [48] and Gabay and Mercier [46]. Since then there have been a plethora of literature to study this algorithm, with a non-exhaustive list provided here [23, 35, 36, 41, 44, 45, 95]. A literature survey of ADMM, its implementation and applications is provided by Boyd et al.[18].

The ADMM algorithm takes the form of a decomposition–coordination procedure, where the solutions to smaller sub-problems are coordinated to find the solution to a larger problem.

ADMM combines the benefits of the dual decomposition and augmented Lagrangian methods, where the problem is partitioned through dual decomposition, and reconciled through the augmented Lagrangian method. Depending on the application, the smaller sub-problems can reduce the complexity of the original problem, or at least reduce the computational burden of having to directly solve the larger problem. Iteratively coordinating and reconciling the sub-problems eventually leads to convergence, yielding a solution to the original problem. Our implementation of the ADMM algorithm is motivated by that of You and Peng [99], where they applied a similar method in the context of optimal power flow.

ADMM steps: generalized risk parity

To implement a suitable ADMM algorithm for GRP, we begin by transferring the rank-1 requirement to an auxiliary variable $\boldsymbol{Z} \in \mathbb{S}_+^{n+1}$, followed by the inclusion of two additional constraints: $\boldsymbol{Y} = \boldsymbol{Z}$ and $\operatorname{rank}(\boldsymbol{Z}) = 1$. Directly imposing these constraints would convert the SDP relaxation into a non-convex problem. However, the auxiliary variable $\boldsymbol{Z}$ enables us to decompose the problem into a convex sub-problem and a non-convex sub-problem.

The convex sub-problem is the same as the SDP relaxation in (6.4), except we replace the objective function (6.4a) with the augmented Lagrangian below

$$\begin{aligned} L_{\mathrm{GRP}}(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Lambda}) &\triangleq \operatorname{Tr}(\boldsymbol{Q}\boldsymbol{Y}) + \operatorname{Tr}\big(\boldsymbol{\Lambda}^\top(\boldsymbol{Y} - \boldsymbol{Z})\big) + \frac{\rho}{2}\|\boldsymbol{Y} - \boldsymbol{Z}\|_F^2 \\ &= \operatorname{Tr}(\boldsymbol{Q}\boldsymbol{Y}) + \frac{\rho}{2}\Big\|\boldsymbol{Y} - \Big(\boldsymbol{Z} - \frac{1}{\rho}\boldsymbol{\Lambda}\Big)\Big\|_F^2. \end{aligned}$$

The augmented Lagrangian adds both a Lagrangian term and a penalty term to our original objective function, thereby attempting to find an optimal solution $\boldsymbol{Y}^*$ that approximates the rank-1 constrained auxiliary variable $\boldsymbol{Z}$. These two additional terms in our objective function can be expressed as a single term by completing the square between the Lagrangian and penalty terms, as we have shown above. The parameter $\rho \in \mathbb{R}_+$ is a tuning parameter and $\boldsymbol{\Lambda} \in \mathbb{S}^{n+1}$ is the dual variable corresponding to our self-imposed constraint $\boldsymbol{Y} = \boldsymbol{Z}$.

Through this reformulation, we transfer the non-convexity of the problem to the auxiliary variable $\boldsymbol{Z}$. In turn, this allows us to maintain convexity in $\boldsymbol{Y}$, which leads to the convex and

non-convex sub-problems. The ADMM algorithm iterates through the following steps

$$\text{Convex } \boldsymbol{Y}\text{-minimization:} \qquad \boldsymbol{Y}^{k+1} = \underset{(\boldsymbol{Y},\theta)\in\mathcal{S}_{\mathrm{GRP}}}{\operatorname{argmin}} \; L_{\mathrm{GRP}}(\boldsymbol{Y}, \boldsymbol{Z}^k, \boldsymbol{\Lambda}^k), \tag{6.5}$$

$$\text{Non-convex } \boldsymbol{Z}\text{-minimization:} \qquad \boldsymbol{Z}^{k+1} = \underset{\operatorname{rank}(\boldsymbol{Z})\leq 1}{\operatorname{argmin}} \; L_{\mathrm{GRP}}(\boldsymbol{Y}^{k+1}, \boldsymbol{Z}, \boldsymbol{\Lambda}^k), \tag{6.6}$$

$$\text{Dual variable } \boldsymbol{\Lambda}\text{-update:} \qquad \boldsymbol{\Lambda}^{k+1} = \boldsymbol{\Lambda}^k + \rho(\boldsymbol{Y}^{k+1} - \boldsymbol{Z}^{k+1}), \tag{6.7}$$

for each iteration k. The $\boldsymbol{Y}$-minimization step minimizes L_{GRP} subject to the convex set $\mathcal{S}_{\mathrm{GRP}}$. For clarity, we present the complete convex $\boldsymbol{Y}$-minimization step below,

$$\min_{\boldsymbol{Y},\zeta} \quad \operatorname{Tr}(\boldsymbol{Q}\boldsymbol{Y}) + \frac{\rho}{2}\left\|\boldsymbol{Y} - \left(\boldsymbol{Z} - \frac{1}{\rho}\boldsymbol{\Lambda}\right)\right\|_F^2 \tag{6.8a}$$

$$\text{s.t.} \quad (1+c)\zeta - \operatorname{Tr}(\boldsymbol{C}^i\ \boldsymbol{Y}) \geq 0, \quad i = 1,\ldots,n, \tag{6.8b}$$

$$\operatorname{Tr}(\boldsymbol{C}^i\ \boldsymbol{Y}) - (1-c)\zeta \geq 0, \quad i = 1,\ldots,n, \tag{6.8c}$$

$$\sum_{i=1}^{n} Y_{i,n+1} = 1, \tag{6.8d}$$

$$Y_{n+1,n+1} = 1, \tag{6.8e}$$

$$\boldsymbol{Y} \succeq 0. \tag{6.8f}$$

For reference, we note that the convex set formed by constraints (6.8b–6.8f) is the same as the set $\mathcal{S}_{\mathrm{GRP}}$ from problem (6.4).

If necessary, we can rewrite the the square of the Frobenius norm term in the objective as a semidefinite constraint through its Schur complement, i.e.,

$$\left\|\boldsymbol{Y} - \left(\boldsymbol{Z} - \frac{1}{\rho}\boldsymbol{\Lambda}\right)\right\|_F^2 \leq \nu \iff \begin{bmatrix} \boldsymbol{I} & \operatorname{vec}\left(\boldsymbol{Y} - (\boldsymbol{Z} - \frac{1}{\rho}\boldsymbol{\Lambda})\right) \\ \operatorname{vec}\left(\boldsymbol{Y} - (\boldsymbol{Z} - \frac{1}{\rho}\boldsymbol{\Lambda})\right)^\top & \nu \end{bmatrix} \succeq 0.$$

where $\operatorname{vec}(\cdot)$ is an operator that vectorizes a matrix (i.e., it reshapes a matrix into a column vector), $\boldsymbol{I}$ is an identity matrix of appropriate size, and $\nu \in \mathbb{R}$ is an auxiliary variable. However, it is typically numerically more efficient to model the square of the Frobenius norm as a rotated second-order cone rather than using a semidefinite cone constraint.

The $\boldsymbol{Z}$-minimization step projects the matrix $\boldsymbol{Y}^{k+1} + \frac{1}{\rho}\boldsymbol{\Lambda}^k$ onto a rank-constrained non-convex set. Nevertheless, rank-constrained problems can be solved analytically through the Eckart–Young–Mirsky theorem [18, 99]. We can find the optimal rank-1 approximation through

a singular value decomposition (SVD) of the matrix

$$\boldsymbol{Z}^{k+1} = s \cdot \boldsymbol{v}\boldsymbol{v}^\top,$$

where $s \in \mathbb{R}_+$ and $\boldsymbol{v} \in \mathbb{R}^{n+1}$ are the top singular value and vector of the matrix $\boldsymbol{Y}^{k+1} + \frac{1}{\rho}\boldsymbol{\Lambda}^k$, respectively. This closed-form solution allows us to efficiently solve the non-convex $\boldsymbol{Z}$-minimization step. Finally, the $\boldsymbol{\Lambda}$-update step is straightforward and can be carried out as shown in (6.7).

The $\boldsymbol{Y}$-minimization SDP can be tightened towards a rank-1 solution by repeatedly iterating through steps (6.5–6.7). Once we converge to a rank-1 solution, the SDP is no longer considered a relaxation of the original GRP problem. In other words, any optimal rank-1 solution will also respect the original feasible set.

We favour this approach of solving our quadratically-constrained non-convex GRP problem because it allows us to iteratively work towards optimality from a lower bound. Starting with a relaxed model allows us to find a lower bound to our optimal solution. Although a lower optimal objective value is preferable, in practice this lower bound is seldom feasible. Nevertheless, after identifying the lower bound, we can proceed sequentially towards the first available feasible solution. By design, this attempts to find a feasible solution that is as close as possible to the lower bound. This solution may be the global optimum, or, at the very least, we often find this a high quality optimal solution.

We note that the GRP framework and its ADMM algorithm are able to accommodate additional convex constraints, provided that these constraints still allow us to reformulate the problem as a SDP. Thus, practical considerations such as specific limits on short selling or weight concentration limits can be easily added to the problem. However, the implementation of these variants is beyond the scope of this thesis.

ADMM steps: robust generalized risk parity

We can follow a similar approach for the robust GRP problem, relaxing the problem into a SDP. The objective function of the SDP relaxation uses the same decision variable $\boldsymbol{Y}$ and input parameter $\boldsymbol{Q}$, with the addition of the robust term that captures the ellipsoidal uncertainty around the estimated portfolio return. The corresponding augmented Lagrangian of the robust

problem is

$$L_{\mathrm{R}}(\boldsymbol{Y},\boldsymbol{Z},\boldsymbol{\Lambda}) \triangleq \mathrm{Tr}(\boldsymbol{Q}\boldsymbol{Y}) + \lambda\theta\left\|\boldsymbol{\Theta}^{1/2}\boldsymbol{Y}_{1:n,n+1}\right\|_2 + \frac{\rho}{2}\left\|\boldsymbol{Y} - \left(\boldsymbol{Z} - \frac{1}{\rho}\boldsymbol{\Lambda}\right)\right\|_F^2.$$

The corresponding ADMM algorithm follows the same steps as those outlined in (6.5–6.7), with the addition of the robust term in the augmented Lagrangian function L_{R}. Specifically, the robust $\boldsymbol{Y}$-minimization step is the following

$$\begin{aligned}
\min_{\boldsymbol{Y},\zeta} \quad & \mathrm{Tr}(\boldsymbol{Q}\boldsymbol{Y}) + \lambda\theta\left\|\boldsymbol{\Theta}^{1/2}\boldsymbol{Y}_{1:n,n+1}\right\|_2 + \frac{\rho}{2}\left\|\boldsymbol{Y} - \left(\boldsymbol{Z} - \frac{1}{\rho}\boldsymbol{\Lambda}\right)\right\|_F^2 && \text{(6.9a)}\\
\text{s.t.} \quad & (1+c)\zeta - \mathrm{Tr}(\boldsymbol{C}^i\ \boldsymbol{Y}) \geq 0, \quad i = 1,\ldots,n, && \text{(6.9b)}\\
& \mathrm{Tr}(\boldsymbol{C}^i\ \boldsymbol{Y}) - (1-c)\zeta \geq 0, \quad i = 1,\ldots,n, && \text{(6.9c)}\\
& \sum_{i=1}^{n} Y_{i,n+1} = 1, && \text{(6.9d)}\\
& Y_{n+1,n+1} = 1, && \text{(6.9e)}\\
& \boldsymbol{Y} \succeq 0. && \text{(6.9f)}
\end{aligned}$$

If necessary, we can rewrite the ℓ_2-norm stemming from the robust term as a semidefinite constraint through its Schur complement, i.e.,

$$\|\boldsymbol{\Theta}^{1/2}\boldsymbol{Y}_{1:n,n+1}\|_2 \leq \tau \iff \begin{bmatrix} \tau\boldsymbol{I} & \boldsymbol{\Theta}^{1/2}\boldsymbol{Y}_{1:n,n+1} \\ (\boldsymbol{\Theta}^{1/2}\boldsymbol{Y}_{1:n,n+1})^\top & \tau \end{bmatrix} \succeq 0.$$

where $\tau \in \mathbb{R}_+$ is an auxiliary variable. The $\boldsymbol{Z}$-minimization step (6.6) and $\boldsymbol{\Lambda}$-update step (6.7) pertain to the sequential approximation of a rank-1 solution, and are unaffected by the inclusion of the robust term in the augmented Lagrangian L_{R}. Thus, these two steps remain the same as before. We note that the convex set described by constraints (6.9b–6.9f) is the same as the set $\mathcal{S}_{\mathrm{GRP}}$.

ADMM steps: lowest-variance risk parity

The LVRP problem in (6.3) can be formulated as a special case of the GRP problem with the trade-off coefficient and the risk diversification coefficient set to zero. Therefore, the corresponding ADMM steps are the same as those in (6.5–6.7), except we modify the $\boldsymbol{Y}$-minimization by setting $\lambda = c = 0$. However, we can take advantage of these conditions to further simplify the

problem. Setting $\lambda = 0$ allows us to ignore the estimated expected returns. In turn, we do this by introducing a new input parameter

$$\boldsymbol{M} \triangleq \begin{bmatrix} \hat{\boldsymbol{\Sigma}} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix},$$

which aligns the estimated covariance matrix with the dimensions of the variable $\boldsymbol{Y}$. It follows that the augmented Lagrangian of the LVRP problem is

$$L_{\mathrm{LV}}(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Lambda}) \triangleq \mathrm{Tr}(\boldsymbol{M}\boldsymbol{Y}) + \frac{\rho}{2}\left\|\boldsymbol{Y} - \left(\boldsymbol{Z} - \frac{1}{\rho}\boldsymbol{\Lambda}\right)\right\|_F^2.$$

We proceed to modify the $\boldsymbol{Y}$-minimization step by setting $c = 0$, allowing us to convert the risk dispersion inequality constraints into equality constraints. The LVRP $\boldsymbol{Y}$-minimization step is

$$\begin{aligned}
\min_{\boldsymbol{Y},\zeta} \quad & \mathrm{Tr}(\boldsymbol{M}\boldsymbol{Y}) + \frac{\rho}{2}\left\|\boldsymbol{Y} - \left(\boldsymbol{Z} - \frac{1}{\rho}\boldsymbol{\Lambda}\right)\right\|_F^2 && \text{(6.10a)} \\
\text{s.t.} \quad & \mathrm{Tr}(\boldsymbol{C}^i\ \boldsymbol{Y}) = \zeta, \quad i = 1, \ldots, n, && \text{(6.10b)} \\
& \sum_{i=1}^{n} Y_{i,n+1} = 1, && \text{(6.10c)} \\
& Y_{n+1,n+1} = 1, && \text{(6.10d)} \\
& \boldsymbol{Y} \succeq 0. && \text{(6.10e)}
\end{aligned}$$

The changes brought on by the LVRP problem have no impact on the $\boldsymbol{Z}$-minimization step (6.6) and $\boldsymbol{\Lambda}$-update step (6.7), i.e., these steps remain unchanged. We will refer to the convex set described by constraints (6.10b–6.10e) as the set $\mathcal{S}_{\mathrm{LV}}$.

6.2.3 ADMM algorithm

Here we describe a tractable implementation of the ADMM algorithm. To do so, we introduce several new parameters pertaining to the tuning of the iteration step size. The selection of parameters in ADMM is the source of ongoing research (e.g., see [47, 18, 85]), and there is no clear consensus on how to set these parameters precisely.

To set the algorithm stopping criteria and the penalty parameter ρ, we borrow some insight from Boyd et al. [18]. The stopping criterion is determined by the sizes of the primal and dual

residuals. Thus, for each iteration k, we have that

$$\varepsilon_p^{k+1} \triangleq \|\boldsymbol{Y}^{k+1} - \bar{\boldsymbol{Z}}^{k+1}\|_F,$$
$$\varepsilon_d^{k+1} \triangleq \rho_k \|\bar{\boldsymbol{Z}}^{k+1} - \bar{\boldsymbol{Z}}^{k}\|_F,$$

where the primal residual ε_p^k measures the difference between our solution to the convex step $\boldsymbol{Y}$ and its rank-1 approximation $\boldsymbol{Z}$, and the dual residual ε_d^k measures the progress attained by $\boldsymbol{Z}$ between each iteration. However, in practice, we use a relaxation of the variable $\boldsymbol{Z}$, denoted by $\bar{\boldsymbol{Z}}$. The definition of $\bar{\boldsymbol{Z}}$ requires us to define a few additional parameters, which we will discuss shortly.

The primal and dual residuals can measure our convergence and can be used as a stopping criteria if we define an acceptable tolerance level $\varepsilon \in \mathbb{R}_+$. In our case, we found that having $\varepsilon = 10^{-6}$ was sufficient to guarantee a good rank-1 approximation, i.e., we stopped the ADMM algorithm once $\varepsilon_p^k \leq 10^{-6}$.

We implement the 'fast ADMM' algorithm proposed in [47, 50] to improve our numerical performance. The 'fast ADMM' algorithm is applicable to convex problems, but can be applied to our problem given that our non-convex $\boldsymbol{Z}$-minimization step can be efficiently solved in closed-form. The fast ADMM algorithm is based on Nesterov's gradient decent method [79]. This variant of gradient descent is accelerated through an over-relaxation step. The fast ADMM algorithm applies a similar technique, and relaxes the $\boldsymbol{Z}$-minimization and $\boldsymbol{\Lambda}$-update steps. For clarity, we outline the relaxation parameters α and β from [47] below,

$$\beta^{k+1} \triangleq \frac{1 + \sqrt{1 + 4(\beta^k)^2}}{2},$$
$$\alpha^{k+1} \triangleq \begin{cases} 1 + \frac{\beta^k - 1}{\beta^{k+1}}, & \frac{\max(\|\varepsilon_p^k\|, \|\varepsilon_d^k\|)}{\max(\|\varepsilon_p^{k-1}\|, \|\varepsilon_d^{k-1}\|)} < 1, \\ 1, & \text{otherwise.} \end{cases}$$

In turn, α^{k+1} and β^{k+1} serve to relax the ADMM steps as follows

$$\bar{\boldsymbol{Z}}^{k+1} \triangleq \alpha^{k+1}\boldsymbol{Z}^{k+1} + (1 - \alpha^{k+1})\boldsymbol{Z}^k,$$
$$\bar{\boldsymbol{\Lambda}}^{k+1} \triangleq \alpha^{k+1}\boldsymbol{\Lambda}^{k+1} + (1 - \alpha^{k+1})\boldsymbol{\Lambda}^k.$$

The implementation of the fast ADMM algorithm is not necessary for convergence. However, it may improve the convergence rate of the algorithm, which we found particularly desirable

when dealing with large-scale problems.

We proceed to discuss the penalty parameter ρ. The penalty parameter serves to measure the trade-off between the primal and dual residuals during each iteration. Intuitively, a large penalty emphasizes our search for a solution that satisfies the rank-1 condition, reducing our primal residual. However, this has the effect of increasing our dual residual. From a practical standpoint, an aggressive penalty parameter has the drawback of forcing the algorithm to take large steps in pursuing a rank-1 solution, limiting our search of an optimal portfolio to a smaller space within the feasible set. Thus, in our case, we initialize the algorithm with a small penalty of $\rho_0 = 0.005$, and we update the penalty parameter after each iteration using the scheme outlined in [58] and [97],

$$\rho_{k+1} \triangleq \begin{cases} a\rho_k, & \varepsilon_p^{k+1} > b\varepsilon_d^{k+1}, \\ \rho_k/a, & \varepsilon_d^{k+1} > b\varepsilon_p^{k+1}, \\ \rho_k, & \text{otherwise}, \end{cases}$$

where $a \geq 1$ and $b \geq 1$ are two additional parameters that allow us to quantify the permissible level of variation of our penalty parameter. To provide a smooth change per iteration, we have found that the parameters $a = 1.05$ and $b = 10$ work well in practice.

Finally, we present the complete ADMM algorithm below in Algorithm 3. The algorithm outlines the steps necessary to solve our choice of problem between: (a) GRP, (b) robust GRP, or (c) LVRP. Moreover, if a user does not wish to use the fast ADMM algorithm, we can ignore the corresponding steps and parameters.

6.3 Numerical experiments

This section presents multiple numerical experiments that test the GRP problem and its two extensions, namely the robust GRP and LVRP problems. The experiments are divided into two categories: in-sample and out-of-sample. The purpose of the in-sample experiments is to assess the quality of the GRP problem and its corresponding ADMM algorithm from an optimization perspective. Thus, we focus on assessing the quality of the optimal solutions attained within the feasible region based solely on the in-sample parameters given to the algorithms. In other words, the quality of these optimization problems is measured by their optimal objective values.

On the other hand, the out-of-sample experiment serves to evaluate the GRP framework as a financial investment tool. The aim is to assess the financial performance of portfolios with

Algorithm 3: Fast ADMM for GRP portfolios

Input: Estimated covariance matrix $\hat{\boldsymbol{\Sigma}} \in \mathbb{S}_+^n$; Estimated expected returns $\hat{\boldsymbol{\mu}} \in \mathbb{R}^n$; Risk diversification coefficient $c \in [0, 1)$; Penalty parameters ρ_0, a, b; Primal and dual residuals ε_p^0, ε_d^0; Convergence tolerance ε; Fast ADMM parameters α^0, β^0; (Optional) Robust parameters $\theta \in \mathbb{R}_+$ and $\boldsymbol{\Theta} \in \mathbb{S}_+^n$

Output: Optimal portfolio $\boldsymbol{x}^*$

$k = 0$

Select the appropriate $\boldsymbol{Y}$-minimization step:

(a) GRP: $\boldsymbol{Y}^0 = \underset{(\boldsymbol{Y},\zeta)\in\mathcal{S}_{\mathrm{GRP}}}{\operatorname{argmin}} \ \operatorname{Tr}(\boldsymbol{Q}\boldsymbol{Y})$

(b) Robust GRP: $\boldsymbol{Y}^0 = \underset{(\boldsymbol{Y},\zeta)\in\mathcal{S}_{\mathrm{GRP}}}{\operatorname{argmin}} \ \operatorname{Tr}(\boldsymbol{Q}\boldsymbol{Y}) + \lambda\theta\tau$

(c) LVRP: $\boldsymbol{Y}^0 = \underset{(\boldsymbol{Y},\zeta)\in\mathcal{S}_{\mathrm{LV}}}{\operatorname{argmin}} \ \operatorname{Tr}(\boldsymbol{M}\boldsymbol{Y})$

$\boldsymbol{Z}^0 = s \cdot \boldsymbol{v}\boldsymbol{v}^\top$ through the SVD of $\boldsymbol{Y}^0$

$\boldsymbol{\Lambda}^0 = \rho_0(\boldsymbol{Y}^0 - \boldsymbol{Z}^0)$

$\bar{\boldsymbol{Z}}^0 = \boldsymbol{Z}^0$ and $\bar{\boldsymbol{\Lambda}}^0 = \boldsymbol{\Lambda}^0$

while $\varepsilon_p^k > \varepsilon$ **do**

  Select the appropriate $\boldsymbol{Y}$-minimization step:

  (a) GRP: $\boldsymbol{Y}^{k+1} = \underset{(\boldsymbol{Y},\zeta)\in\mathcal{S}_{\mathrm{GRP}}}{\operatorname{argmin}} \ L_{\mathrm{GRP}}(\boldsymbol{Y}, \bar{\boldsymbol{Z}}^k, \bar{\boldsymbol{\Lambda}}^k)$

  (b) Robust GRP: $\boldsymbol{Y}^{k+1} = \underset{(\boldsymbol{Y},\zeta)\in\mathcal{S}_{\mathrm{GRP}}}{\operatorname{argmin}} \ L_{\mathrm{R}}(\boldsymbol{Y}, \bar{\boldsymbol{Z}}^k, \bar{\boldsymbol{\Lambda}}^k)$

  (c) LVRP: $\boldsymbol{Y}^{k+1} = \underset{(\boldsymbol{Y},\zeta)\in\mathcal{S}_{\mathrm{LV}}}{\operatorname{argmin}} \ L_{\mathrm{LV}}(\boldsymbol{Y}, \bar{\boldsymbol{Z}}^k, \bar{\boldsymbol{\Lambda}}^k)$

  $\boldsymbol{Z}^{k+1} = s \cdot \boldsymbol{v}\boldsymbol{v}^\top$ through the SVD of $\left(\boldsymbol{Y}^{k+1} + \frac{1}{\rho_k}\bar{\boldsymbol{\Lambda}}^k\right)$

  $\boldsymbol{\Lambda}^{k+1} = \bar{\boldsymbol{\Lambda}}^k + \rho_k(\boldsymbol{Y}^{k+1} - \boldsymbol{Z}^{k+1})$

  $\beta^{k+1} = \frac{1+\sqrt{1+4(\beta^k)^2}}{2}$

  $\alpha^{k+1} = \begin{cases} 1 + \frac{\beta^k - 1}{\beta^{k+1}}, & \frac{\max(\|\varepsilon_p^k\|, \|\varepsilon_d^k\|)}{\max(\|\varepsilon_p^{k-1}\|, \|\varepsilon_d^{k-1}\|)} < 1, \\ 1, & \text{otherwise.} \end{cases}$

  $\bar{\boldsymbol{Z}}^{k+1} = \alpha^{k+1}\boldsymbol{Z}^{k+1} + (1 - \alpha^{k+1})\boldsymbol{Z}^k$

  $\bar{\boldsymbol{\Lambda}}^{k+1} = \alpha^{k+1}\boldsymbol{\Lambda}^{k+1} + (1 - \alpha^{k+1})\boldsymbol{\Lambda}^k$

  $\varepsilon_p^{k+1} = \|\boldsymbol{Y}^{k+1} - \bar{\boldsymbol{Z}}^{k+1}\|_F$

  $\varepsilon_d^{k+1} = \rho_k\|\bar{\boldsymbol{Z}}^{k+1} - \bar{\boldsymbol{Z}}^k\|_F$

  $\rho_{k+1} = \begin{cases} a\rho_k, & \varepsilon_p^{k+1} > b\varepsilon_d^{k+1}, \\ \rho_k/a, & \varepsilon_d^{k+1} > b\varepsilon_p^{k+1}, \\ \rho_k, & \text{otherwise.} \end{cases}$

  $k = k + 1$

Result: Optimal portfolio $\boldsymbol{x}^*$

risk-based diversification constraints. From an optimization perspective, imposing additional constraints reduces the size of the feasible region. Therefore, adding the risk-based diversification constraints is likely to worsen our in-sample optimal objective value. Nevertheless, there are practical benefits to having a risk-diverse portfolio [5, 26, 70], but these benefits are better assessed in an out-of-sample environment.

All experiments were conducted on an Apple MacBook Pro computer (2.8 GHz Intel Core i7, 16 GB 2133 MHz DDR3 RAM) running macOS 'Catalina'. The computer script was written in the Julia programming language (version 1.4.0) using the modelling language 'JuMP' [33] with MOSEK (version 9.0.1) as the optimization solver for SDPs and IPOPT (version 3.12.6) as the solver for the non-convex problems.

6.3.1 In-sample experiments

The in-sample experiments are designed to measure performance from an optimization perspective. In other words, the aim of the in-sample experiments is to determine whether the optimal objective value of the non-convex GRP problem can be improved by using the proposed ADMM algorithm.

We present the results of three in-sample experiments, one for each of the problems proposed: GRP, robust GRP, and LVRP. The three experiments share the following structure. There are four trials per experiment, with the number of assets $n = 33, 50$.

The trials use historical pricing data from a diverse pool of U.S. stocks belonging to the S&P 500 index. A list of these stocks is presented in Table 6.1. When $n = 33$ we use the stocks listed in the first three columns of the table. When $n = 50$ we use all the stocks listed in Table 6.1. The asset expected returns $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ are estimated using weekly excess returns from 01–Jan–2007 to 31–Dec–2009. The estimation process corresponds to a single-period estimate. In other words, we use all data from this three-year period to estimate a single pair of parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Given that it is not particularly relevant to the GRP framework, we do not use a factor model to estimate our parameters. The historical stock prices were obtained from Quandl.com [84].

The numerical results are evaluated primarily by three measures. First, we compare their objective values after optimization, with a lower value demonstrating better performance since these are minimization problems. The objective value is calculated on the risk–return profile alone, excluding the value of the Lagrangian penalty term brought on by the augmented La-

Table 6.1: List of assets

GICS Sector	**Company Tickers**				
Consumer Disc.	DIS	F	MCD	GPS	NKE
Consumer Staples	WMT	KO	KR	PG	CL
Financials	JPM	BAC	C	AON	WFC
Healthcare	BMY	PFE	JNJ	LLY	HUM
Industrials	BA	CAT	GE	LMT	MMM
Information Tech.	AAPL	HPQ	IBM	ORCL	QCOM
Materials	IP	MOS	NEM	PPG	PX
Energy	XOM	HAL	OXY	MRO	
Utilities	CNP	DTE	DUK	ED	
Real Estate	HCP	REG	UDR	WY	
Telecom.	S	T	VZ		

grangian method. In the case of the robust framework, we consider only the robust risk–return profile. For the lowest variance framework, we use the portfolio variance as the objective value.

The second measure of performance is the CV of the portfolios, which quantifies our risk-based diversification. The CV is calculated as shown in (2.21). A lower CV indicates a higher degree of risk-based diversification, with a CV equal to zero indicating risk parity. The user-defined risk diversification coefficient c controls the dispersion of the asset risk contributions, approximately bounding the CV. For example, if we have $c = 0.2$, we expect the CV of the resulting portfolios to be CV ≈ 0.2.

The third measure of performance is the runtime. This serves to compare the speed at which the different optimization models are able to produce an optimal solution. We consider this a relative measure due to the availability of our computing power. Thus, it is only useful in the context of the different portfolios tested during these experiments.

In-sample results: generalized risk parity

This experiment consists of two independent trials with $n = 33$ and $n = 50$. Both trials were conducted with a risk–return trade-off coefficient $\lambda = 0.1$ and with $c = 0.25$ and $c = 0.15$ as the risk diversification coefficient for the first and second trials, respectively. We tested the following competing optimization models:

- Nominal MVO: the convex quadratic problem in (2.13).
- SDP: the SDP in (6.4), which is the convex relaxation of the GRP problem.

- Non-convex (NC): the original non-convex formulation of the GRP problem in (6.1).
- NC-Warm: the same as the non-convex GRP problem in (6.1), except we warm-start the problem with the solution from the SDP relaxation.
- ADMM: the GRP problem prescribed by Algorithm 3 (a).

The nominal MVO model serves as a benchmark for our experiment. The nominal MVO problem can be viewed as a relaxation of the GRP problem, i.e., it is the optimization problem we would have if remove the risk-based diversification constraints. Therefore, the feasible region of the nominal MVO problem is larger than that of the GRP problem. In turn, the optimal objective value of the nominal MVO problem is guaranteed to be less than or equal to the rest of the competing problems. From our perspective, a high quality optimal solution from the competing problems should come as close as possible to that of the nominal MVO.

In practice, the GRP problem will not attain a better optimal objective value than the nominal MVO. However, the reason that the GRP problem may be appealing to an investor stems from a desire to maintain a degree of risk-based diversification because of the potential out-of-sample benefits.

Table 6.2 presents a summary of the results. Given that the SDP problem is a relaxation of our non-convex problem, it is not surprising to see that its optimal objective value comes as close as possible to that of the nominal MVO problem, with the caveat that it violates the risk diversification constraints. Nevertheless, the SDP relaxation serves as a lower bound on the GRP framework. The non-convex model is able to find a local optimal solution that complies with the risk dispersion constraints, but it appears to be far from the global optimal solution. This becomes apparent after we compare it to the warm-start non-convex problem. Nevertheless, the ADMM algorithm attains the most desirable objective value while still respecting the risk dispersion bounds. While we are not able to provide a theoretical guarantee on global optimality, we note that the ADMM algorithm is able to provide a better solution than its non-convex counterparts. We also note that this observation holds for both trials.

However, improving the optimal objective value through the ADMM algorithm comes at an increased computational cost. The runtime of the ADMM algorithm is significantly higher than the competing problems because of the following two reasons. First, the ADMM algorithm requires multiple iterations to converge to a desirable solution, meaning we must solve a sequence of convex optimization problems. Second, each iteration requires us to solve a SDP, which

Table 6.2: Summary of in-sample results for the GRP framework

	$n = 33$, $c = 0.25$, $\lambda = 0.1$				
	MVO	SDP	NC	NC-Warm	ADMM
Obj. Value	-0.0040	0.0157	0.0785	1.471	0.0376
CV	2.590	1.150	0.249	0.254	0.254
Runtime (sec)	0.003	0.023	0.045	0.057	16.04
	$n = 50$, $c = 0.15$, $\lambda = 0.1$				
	MVO	SDP	NC	NC-Warm	ADMM
Obj. Value	-0.0301	-0.0171	0.0977	0.0708	0.0186
CV	2.933	1.360	0.153	0.151	0.151
Runtime (sec)	0.006	0.054	0.090	0.085	256.6

Notes: NC, non-convex. NC-Warm, warm-started non-convex.

cannot be solved as efficiently as quadratic programs or second-order cone programs.

Figure 6.1 presents the convergence rate of the ADMM algorithm for both trials. The two plots on the left-hand side show the convergence of the CV. Both plots show how the CV converges towards the desired risk dispersion limit $c = 0.25$ (top) and $c = 0.15$ (bottom). The plots on the right-hand side display the objective value of the ADMM algorithm after each iteration. For comparison, these plots also show the optimal values of the nominal MVO problem, the SDP relaxation, and either the non-convex problem or its warm-started counterpart. Between the two non-convex problems, only the one with the most desirable objective value is shown in the plots. As a reminder, we note that the objective value of the ADMM algorithm is based solely on the risk–return profile, excluding the augmented Lagrangian terms.

The right-hand side plots in Figure 6.1 show that the ADMM algorithm initially departs from the same point as the SDP relaxation. Not surprisingly, the objective value increases as we tighten the solution to impose a rank-1 constraint. Nevertheless, we can also see that the algorithm converges to a more desirable point than either of the non-convex models. We note that during the first trial, with $n = 33$, the warm-start non-convex model actually converges to a worse local optimum when compared to the regular non-convex model. This, in turn, highlights that warm-starting the GRP problem does not always lead to a better solution. While the ADMM algorithm cannot guarantee global optimality, the path taken by the algorithm shows we start from a theoretical lower bound, sequentially sacrificing optimality while seeking feasibility in the original problem. The ADMM algorithm will stop as soon as it finds the first

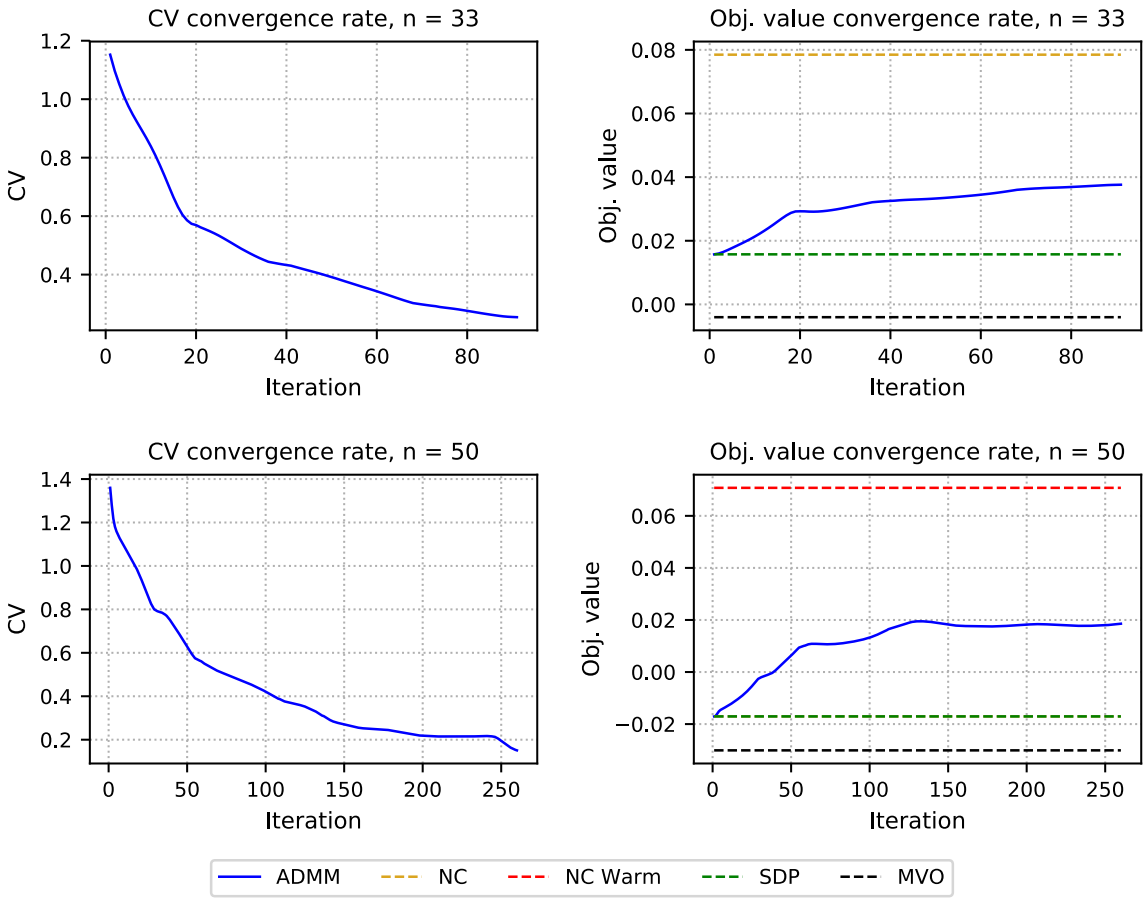

Figure 6.1: Convergence plots for the GRP framework

rank-1 solution, i.e., the first solution that satisfies feasibility of the original problem. Since we approach feasibility from the theoretical lower bound, this method often converges to a high quality, if not the global, optimal solution.

In-sample results: robust generalized risk parity

The second experiment evaluates the robust GRP framework. We construct the robust structure with a 90% confidence interval around the asset expected returns ($\delta = 0.9$). As with the previous experiment, we have five different optimization problems:

- Robust MVO: this model is the robust counterpart to the problem in (2.13), where we use the robust portfolio return with an ellipsoidal uncertainty set.
- SDP: the SDP in (6.9) without the augmented Lagrangian terms, which is the relaxation of the robust GRP framework.
- Non-convex (NC): the non-convex problem in (6.2).
- NC-Warm: the same as the non-convex problem in (6.2), except we warm-start the problem with the solution from the SDP relaxation.
- ADMM: the robust GRP problem prescribed by Algorithm 3 (b).

As with the previous experiment, the robust MVO problem serves as a benchmark since it is a less restrictive problem given that it excludes the risk-based diversification constraints. Therefore, the optimal objective value of the robust MVO problem is theoretically guaranteed to be less than or equal to that of any of the competing problems.

As noted in Table 6.3, both trials were conducted with $\lambda = 0.1$ and with $c = 0.25$ and $c = 0.15$, respectively. The results shown in Table 6.3 and Figure 6.2 follow the same pattern as those from the previous experiment. Once again, we see that the ADMM algorithm is able to attain a higher quality optimal solution when compared to the competing non-convex problems. However, we also find that the additional complexity from the robust formulation increases the runtime of the ADMM algorithm when compared to the non-robust ADMM algorithm from before (see Table 6.2). We found that the ADMM algorithm required more iterations than its non-robust counterpart to converge. Nevertheless, it is still able to deliver a high quality solution within reasonable time.

Table 6.3: Summary of in-sample results for the robust GRP framework

	$n = 33$, $c = 0.25$, $\lambda = 0.1$				
	Robust	SDP	NC	NC-Warm	ADMM
Obj. Value	0.099	0.110	0.536	0.203	0.146
CV	1.72	1.06	0.258	0.253	0.250
Runtime (sec)	0.026	0.026	0.049	0.061	7.16
	$n = 50$, $c = 0.15$, $\lambda = 0.1$				
	Robust	SDP	NC	NC-Warm	ADMM
Obj. Value	0.094	0.103	0.223	0.259	0.141
CV	1.75	1.12	0.150	0.150	0.151
Runtime (sec)	0.040	0.253	0.101	0.103	42.88

Notes: NC, non-convex. NC-Warm, warm-started non-convex.

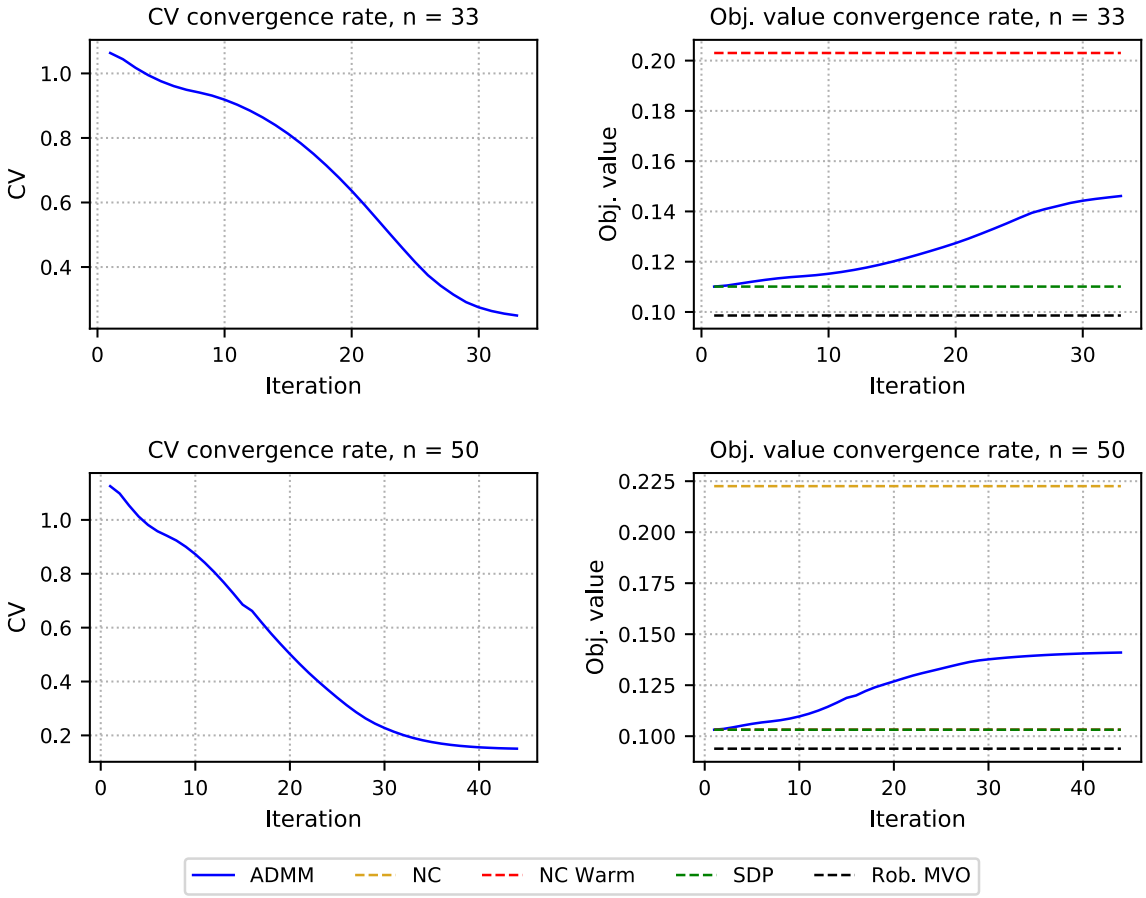

Figure 6.2: Convergence plots for the robust GRP framework

In-sample results: lowest variance risk parity

The third experiment evaluates the LVRP framework. As before, we tested five different optimization models:

- Risk parity: this is the nominal risk parity problem in (2.16), and it imposes the long-only restriction. The objective of this problem is to equalize risk contributions, and does not attempt to minimize variance explicitly.

- SDP: the SDP in (6.10) without the augmented Lagrangian terms, which is the relaxation of the LVRP framework.

- Non-convex (NC): the non-convex problem in (6.3).

- NC-Warm: the same as the non-convex problem in (6.3), except we warm-start the problem with the solution from the SDP relaxation.

- ADMM: the LVRP algorithm prescribed by Algorithm 3 (c).

Unlike our previous experiments, the nominal risk parity problem does not share the same objective as the rest of the competing models. Given that all other competing models seek to minimize variance, the nominal risk parity model is not expected to provide a comparable optimal objective value. Instead, its purpose is to show the portfolio variance attained if we attempt to equalize the asset risk contributions subject to long-only constraints, i.e., it highlights the opportunity forfeited when we disallow short sales.

A summary of the numerical results is shown in Table 6.4. Unlike previous experiments, this table displays the portfolio variance instead of the objective value. This allows for a comparison between the nominal risk parity problem and the LVRP problems.

The numerical results suggest that the LVRP ADMM algorithm can consistently deliver long-short risk parity portfolios with a lower portfolio variance when compared to competing problems. By definition, the SDP relaxation establishes a lower bound on the lowest variance attainable. Figure 6.3 shows how the algorithm sacrifices optimality while seeking to satisfy the risk parity condition. Initially departing from the lower bound, the portfolio variance increases as it seeks to find a rank-1 solution where the asset risk contributions are equalized.

Table 6.4: Summary of in-sample results for the LVRP framework

	$n = 33$				
	Nom. RP	SDP	NC	NC-Warm	ADMM
Variance	0.115	0.0465	0.0738	0.0699	0.0669
CV	2e-12	1.23	0.0005	0.0005	0.0122
Runtime (sec)	0.007	0.016	0.044	0.035	19.44
	$n = 50$				
	Nom. RP	SDP	NC	NC-Warm	ADMM
Variance	0.105	0.0322	0.0601	0.0597	0.0569
CV	2e-15	1.762	0.0008	0.0008	0.0122
Runtime (sec)	0.018	0.032	0.087	0.089	143.2

Notes: Nom. RP, nominal risk parity. NC, non-convex. NC-Warm, warm-started non-convex.

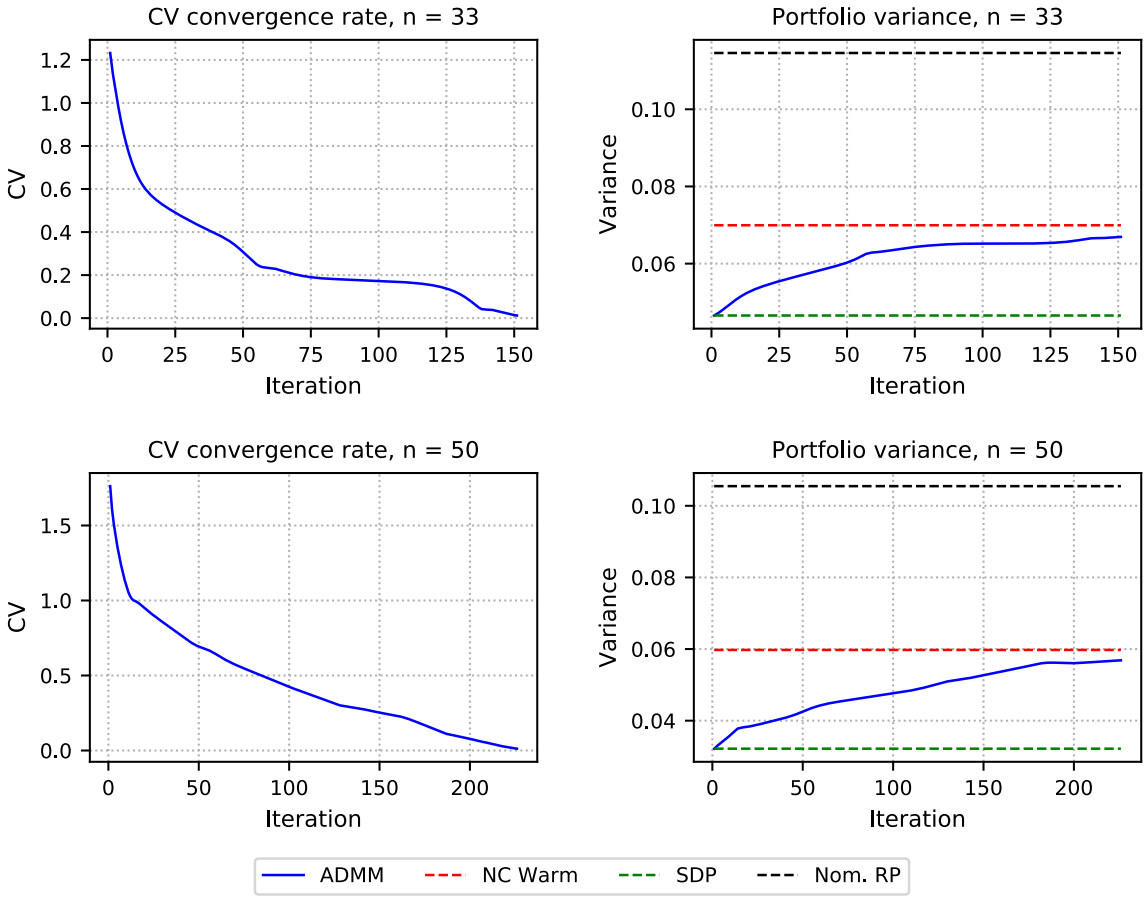

Figure 6.3: Convergence plots for the LVRP framework

6.3.2 Out-of-sample experiment

The out-of-sample experiment evaluates the financial performance of our three proposed frameworks. The experiment focuses solely on the ex post financial performance of the portfolios. An overview of the experimental setup follows. The universe of assets available for investment consists of the 50 assets listed in Table 6.1. The experimental data consists of weekly historical stock prices from 01-Jan-1997 to 31-Dec-2016. The data were obtained from Quandl.com [84]. The asset expected returns $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ are estimated using a three-year calibration window immediately preceding the first investment period. We do not use a factor model to estimate our parameters. The portfolios are rebalanced every six months. All parameters are re-estimated every time the portfolios are rebalanced. To elaborate, consider the first six-month investment period. We calibrate our parameters using data from 01–Jan-1997 to 31–Dec-1999, and then observe the out-of-sample portfolio performance from 01–Jan–2000 to 30–Jun–2000. We then roll the calibration window forward and re-estimate our parameters using the preceding 3-year period (01–Jul–1997 to 30–Jun–2000). The portfolios are then re-optimized and rebalanced, and their out-of-sample performance is observed from 01–Jul–2000 to 31–Dec–2000. We repeat these steps until the end of the investment horizon on 31–Dec–2016. This means we have a total of 34 six-month out-of-sample investment periods, for a total of 17 years, over which we record the wealth evolution of all portfolios.

We note that this experiment is non-exhaustive, as the portfolio performance is highly dependent on the choice of assets, as well as the choice of parameters c and λ. We only conduct this experiment for a single scenario with $c = 0.1$ and $\lambda = 0.05$. In addition, we use a 90% confidence interval around the asset expected returns for the robust models ($\delta = 0.9$). The out-of-sample experiment studies 12 portfolios constructed using the optimization problems listed below.

1. Nominal problems: $n = 50$, $c = 0.1$, $\lambda = 0.05$

    - Nominal MVO: the convex quadratic problem in (2.13).
    - Non-convex: the non-convex GRP problem in (6.1).
    - NC warm: the same as the non-convex GRP problem in (6.1), except we warm-start the problem with the solution to the corresponding SDP relaxation.
    - ADMM: the GRP framework prescribed by Algorithm 3 (a).

2. Robust models: $n = 50$, $c = 0.1$, $\lambda = 0.05$, $\delta = 0.9$

   - Robust MVO: the robust counterpart to the MVO problem in (2.13), where we use the robust portfolio return with an ellipsoidal uncertainty set.
   - Non-convex: the non-convex robust GRP problem in (6.2).
   - NC warm: the same as the non-convex robust GRP problem in (6.2), except we warm-start the problem with the solution to the corresponding SDP relaxation.

   iv) ADMM: the robust ADMM algorithm prescribed by Algorithm 3 (b).

3. Lowest variance models: $n = 50$

   - Risk parity: the nominal risk parity problem in (2.16), and it imposes the long-only restriction.
   - Non-convex: the non-convex LVRP problem in (6.3).
   - NC warm: the same as the non-convex LVRP problem in (6.3), except we warm-start the problem with the solution to the corresponding SDP relaxation.
   - ADMM: the algorithm prescribed by Algorithm 3 (c).

The portfolios are evaluated on their ex post performance, as measured by the following indicators. The portfolio return and risk are measured by the observed weekly rate of return and the corresponding standard deviation. Both the return and risk are shown as annualized measures. This is followed by the risk-adjusted rate of return, also known as the Sharpe ratio [94]. Our final performance is the average portfolio turnover rate, which provides some insight about the management and transaction costs that may be incurred during portfolio rebalancing. The average turnover rate is calculated as the average over the 17-year investment horizon.

Figure 6.4 shows the portfolio wealth evolution over the entire 17-year investment horizon. At first glance, we can see that the GRP and robust GRP ADMM portfolios attained the most wealth when compared to the competing portfolios. However, this is not the case for the LVRP portfolios, where the nominal risk parity portfolio attained the most wealth. Nevertheless, the LVRP ADMM portfolio still managed to attain more wealth than the two competing non-convex models, suggesting that the ADMM algorithm leads to better ex post financial performance than the naive non-convex optimization problems.

To better assess the measures of performance, Table 6.5 presents a summary of results. The results clearly suggest that the GRP portfolios are dominated by the ADMM portfolio.

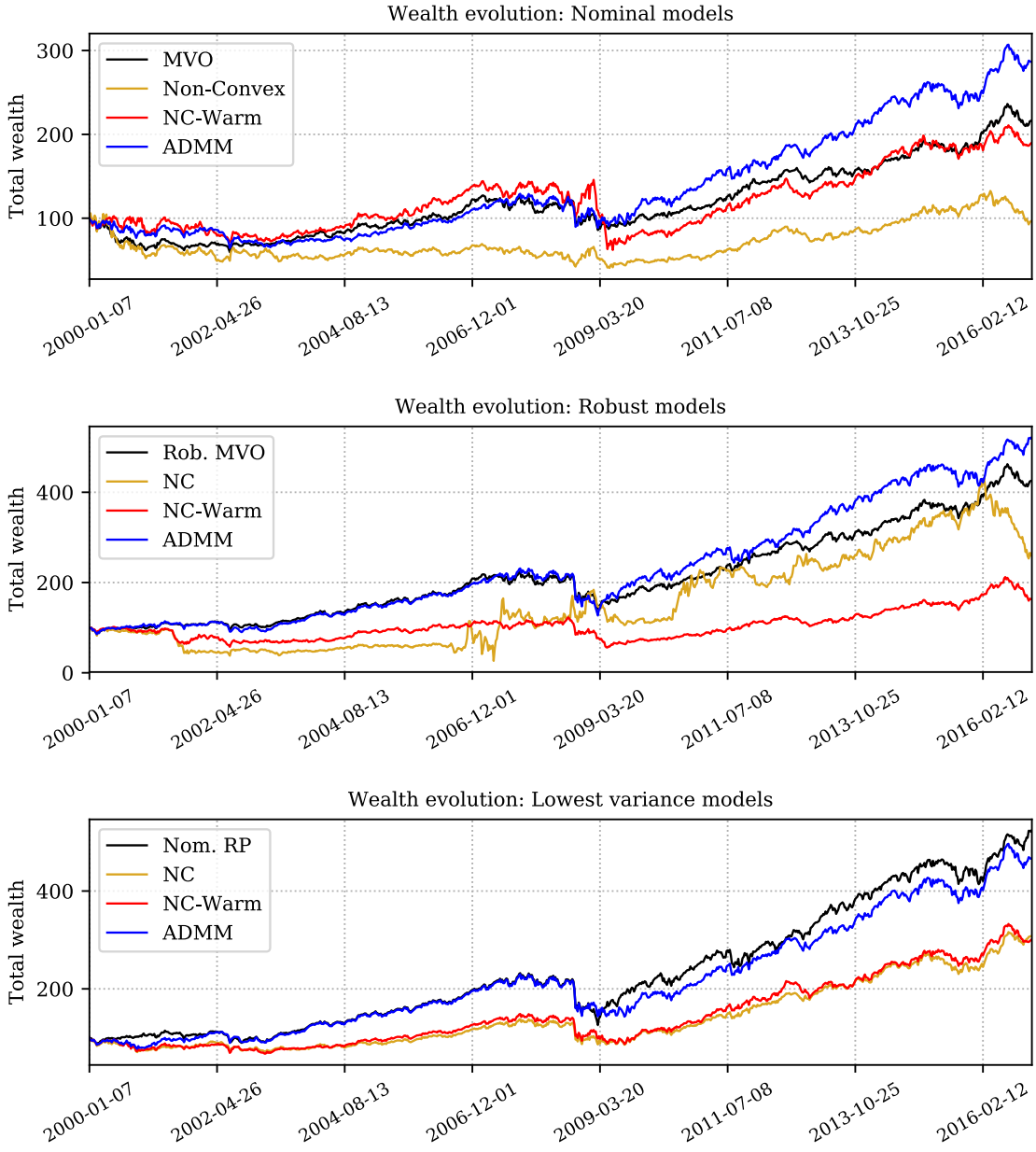

Figure 6.4: Portfolio wealth evolution from 07-Jan-2000 to 30-Dec-2016 with 6-month rebalancing

When compared to the benchmark MVO portfolio, the ADMM portfolio has a clear advantage in terms of realized return, even after adjusting for the risk incurred as shown by the corresponding Sharpe ratios. The average turnover does indicate that the ADMM portfolio does come with higher management and transaction costs, but these values are within a similar range to that of the nominal MVO portfolio.

Comparing the ADMM portfolio against the competing non-convex portfolios, we can see that using a naive optimization approach is problematic. Not only did this lead to an erratic wealth evolution, but the tabulated results also show a meagre financial performance. Most noticeably, the average turnover rate corroborates the erratic behaviour of the non-convex portfolios when they are rebalanced. It appears that their tendency to converge to the nearest local solution leads to a very different set of optimal asset weights every time rebalancing takes place. In other words, the optimal solutions appear to be dramatically different every time the portfolios are re-optimized, leading to an unstable investment strategy and a large period-over-period turnover rate. From a pragmatic perspective, this highlights the importance of having consistency when dealing with real data.

The robust GRP portfolios share a similar trend to the previous ones. Once again, the realized return of the ADMM portfolio is greater than that of the robust MVO portfolio, even after adjusting for risk. Moreover, the average turnover rate is also lower for the ADMM portfolio, implying the ADMM portfolio would incur a lower transaction cost. Additionally, the low average turnover rate highlights the stability of the optimal portfolio weights period over period. As before, both non-convex portfolios displayed a high volatility and lower rate of return. Finally, the non-convex portfolios showed a very large average turnover rate.

Finally, the LVRP results show that the nominal risk parity portfolio, with its long-only constraint, has the best risk-adjusted rate of return. This advantage is compounded by its low turnover rate, which is typical of nominal risk parity portfolios [70]. When compared to the two non-convex portfolios, the ADMM portfolio was once again able to attain a higher rate of return and lower volatility, while maintaining a lower turnover rate.

6.4 Conclusion

This chapter introduced a GRP framework that allows an investor to optimize a predefined risk–return profile while maintaining a desirable degree of risk-based diversification. The pro-

Table 6.5: Summary of out-of-sample results

	GRP: $n = 50$, $c = 0.1$, $\lambda = 0.05$			
	MVO	NC	NC-Warm	ADMM
Ann. Ex. Return (%)	2.8	-1.9	2.0	4.4
Ann. Volatility (%)	15.5	23.3	19.7	16.6
Sharpe Ratio (%)	17.9	-8.3	10.2	26.6
Avg. Turnover (%)	59.9	207.0	142.7	66.2
	Robust: $n = 50$, $c = 0.1$, $\lambda = 0.05$, $\delta = 0.9$			
	Rob. MVO	NC	NC-Warm	ADMM
Ann. Ex. Return (%)	6.7	4.0	1.2	7.9
Ann. Volatility (%)	13.6	48.9	18.2	15.4
Sharpe Ratio (%)	49.5	8.1	6.3	51.5
Avg. Turnover (%)	17.1	807.2	192.2	11.7
	LVRP: $n = 50$			
	Nom. RP	NC	NC-Warm	ADMM
Ann. Ex. Return (%)	7.9	4.9	4.7	7.3
Ann. Volatility (%)	15.6	16.8	16.9	16.6
Sharpe Ratio (%)	50.8	28.9	28.0	44.0
Avg. Turnover (%)	11.9	73.3	63.5	42.5

Notes: Ann, annualized. Ex, excess. Rob., robust. NC, non-convex. NC-Warm, warm-started non-convex. Nom. RP, nominal risk parity.

posed framework also allows for short sales, thereby giving an investor increased flexibility when compared to the nominal risk parity problem. By design, the GRP framework combines the desirable properties of MVO while maintaining a desirable proximity to the risk parity condition. However, imposing the necessary constraints to limit the risk dispersion means the GRP framework leads to a non-convex optimization problem.

We remedy this issue with an ADMM algorithm designed to handle the non-convexity of the GRP framework. We begin by relaxing the original non-convex problem into a SDP. The algorithm operates by partitioning our original problem into two sub-problems: a convex SDP and a non-convex rank-constrained problem. The structure of the latter problem allows for a closed-form solution, thereby avoiding both the theoretical and numerical drawbacks associated with solving a non-convex problem. The algorithm proceeds to solve the problem sequentially through multiple iterations, with each iteration tightening the SDP relaxation towards a rank-1 solution. A rank-1 solution is no longer considered a relaxation, providing an optimal solution that complies with the original risk dispersion constraints.

In addition, we proposed two extensions to our GRP framework. First, we introduced an ellipsoidal uncertainty structure around the portfolio expected returns to formulate a robust GRP framework. Second, we reinstated the risk parity condition, thereby formulating a problem that seeks the risk parity solution with the lowest portfolio variance. We note that our GRP framework can accommodate additional convex constraints, provided these can be formulated as semidefinite constraints.

The numerical experiments show that the proposed ADMM algorithm is able to deliver a higher quality optimal solution when compared to the original non-convex problem, even after warm-starting the non-convex problem. Although the algorithm does require additional runtime, it is still able to converge within reasonable time. Since this algorithm is a heuristic, we are unable to make any claims regarding global optimality. Nevertheless, the design of the algorithm guarantees we start from a theoretical lower bound, slowly climbing upwards towards the nearest feasible solution. This, in turn, suggests we should find a high quality, if not global, optimal solution.

The proposed GRP framework and the implementation of the ADMM algorithm provide an investor with greater flexibility when considering investing in risk parity portfolios. Indeed, we believe this framework combines the most appealing aspects of both MVO and risk parity.

Chapter 7

Conclusion and future research

The objective of this thesis was to introduce multiple frameworks to address some of the most significant weaknesses of the risk parity asset allocation problem. Like many portfolio optimization problems based on forecasts, risk parity is sensitive to uncertainty and estimation errors in its input parameters. Moreover, risk parity promotes risk-based diversification at the expense of having control over a portfolio's risk and reward measure. In other words, the nominal risk parity framework forces the investor to ignore the pillars of MPT: portfolio risk and expected return. The final weakness of risk parity stems from the inherent non-convexity of the problem, which is typically addressed by disallowing short sales at the expense of limiting our investment choices.

This thesis addressed the aforementioned weaknesses of risk parity as follows. First, it addressed various forms of uncertainty associated with the estimation of parameters and their impact during optimization. Second, it introduced a generalized risk parity framework that combined the benefits of risk parity with those of MVO.

Chapter 3 addressed the issue of parameter uncertainty through robust optimization. We introduced a new robust framework specifically designed to target the nuances of risk parity. Traditional robust portfolio optimization problems target uncertainty to protect our optimal portfolio against the worst-case instance of the total portfolio risk. Instead, we propose to treat uncertainty in the same way that risk parity manages risk: we parametrize the individual asset risk contributions and address the uncertainty inherent to each partition of the portfolio risk. By design, the resulting robust risk parity portfolio reduces its investment in assets with greater uncertainty in their individual risk contributions. While this reduction means

the robust portfolio is unable to attain the risk parity condition, it still retains sufficient risk-based diversification while protecting the portfolio against uncertainty. To the best of our knowledge, this work is the first to introduce robustness to the risk parity problem by targeting the individual asset risk contributions, as opposed to targeting the overall portfolio risk measure. Finally, the contributions and findings from this research were published in [30].

The DRRP framework in Chapter 4 embeds robustness in a more standard fashion by targeting the worst-case estimate of the overall portfolio risk measure. However, our approach differs from traditional robust optimization by capturing the distributional information implied by the raw data themselves. In this sense, we do not assume any structure on the data used to estimate our asset covariance matrix, and instead take a scenario-based approach that breaks the assumption that all scenarios used for parameter estimation are equally likely. The problem is formulated as a game-theoretic minimax risk parity problem that sets the investor against a worst-case instance of the portfolio variance. Prior to our work, this particular approach to DRO, which targets the parameter estimation step by exploiting a discrete probability distribution, had only been applied to MVO. Thus, the first contribution of this work was to extend this application to risk parity. Moreover, we proposed a novel algorithm that efficiently solves this type of constrained convex–concave minimax problem. In turn, this means we can construct large DRRP portfolios within a reasonable amount of time. The contributions and findings from this research have been submitted for publication [27].

Instead of attempting to quantify uncertainty and using robust methods to mitigate the impact of parameter uncertainty in risk parity, the objective of Chapter 5 was to focus solely on improving the parameter estimation process. Unlike the previous chapter, we assume that the asset returns (and the corresponding observable data) follow a factor model structure. Moreover, we accept the well-known phenomenon that modern financial markets follow a cyclical behaviour. In turn, we propose a novel Markov regime-switching factor model of asset returns. Although other regime-switching frameworks have been proposed in the literature, these are often not tractable in practice. When they are tractable, they are often not scalable to portfolios with a realistic number of assets. On the other hand, our proposed factor model is designed to align with the best practices of the financial industry, retaining the intuitiveness associated with traditional factor models and allowing us to naturally derive the regime-dependent counterparts of the asset expected returns and covariance matrix. This covariance matrix embeds additional information about the current market regime; and, when used during optimization, it yields

a risk parity portfolio adapted to the current market regime. Our results show that having a better quality estimate of the current covariance matrix leads to portfolios with higher risk-adjusted returns. Retaining a factor model structure means we are able to construct a better risk parity portfolio without having to modify the underlying optimization problem. In other words, our regime-dependent risk parity optimization problem is as tractable and scalable as the original problem. The contributions and findings from this research were published as two separate manuscripts [26, 29].

The last part of this thesis departs from the nominal risk parity goal of being perfectly risk-diverse. Doing so allows us to design a generalized framework that incorporates the desirable elements of MVO. Specifically, the proposed generalized risk parity framework borrows the premise of MVO of attaining an optimal risk–reward trade-off. However, the framework embeds risk-based diversification constraints to ensure that the resulting optimal portfolio is still sufficiently risk-diverse. Mathematically modelling an asset's risk contribution is a non-convex activity, making it likely that our 'optimal' portfolio is, in reality, far from being desirably (i.e., globally) optimal. This motivated the development of an ADMM algorithm tailored to solve the generalized risk parity problem. The algorithm involves solving a sequence of convex relaxations that are tightened until we reach a feasible solution to the original problem. By approaching a feasible solution from a lower bound, we aim to find the global optimal solution. Our formulation is unable to provide theoretical guarantees on global optimality, but our experimental results suggest that the generalized risk parity problem with an ADMM solution is able to attain high quality optimal solutions that are stable when we rebalance our portfolio over multiple investment periods. To the best of our knowledge, this is the first risk parity problem that allows the investor to take short positions and to control the portfolio's risk and return while limiting the level of risk concentration to promote diversification. The contributions and findings from this research were published in [28].

The advances proposed by this thesis enhanced the resilience of risk parity against uncertainty and estimation error, and allowed the investor to construct risk-diverse portfolios while emphasizing other important financial attributes, namely risk, reward and the ability to take short positions. Moreover, this thesis accomplished its objective while respecting the maxim of maintaining interpretability and computational tractability. Each of the proposed frameworks addresses a fundamental deficiency of risk parity from a different perspective, and are meant to serve as tools for anyone wishing to construct modern risk parity portfolios. These frameworks

may be used independently or, in some cases, they may be combined to improve risk parity from multiple angles.

7.1 Future research

Although this thesis advanced risk parity in four different directions, there is still much research left to explore in the area of risk parity and, more broadly, risk-diverse asset allocation strategies. We finalize this thesis by presenting a non-exhaustive list of research directions that are of interest, some of which are currently being pursued.

As a mathematical optimization problem, risk parity seeks to be fully risk-diverse in its asset allocation over a predetermined basket of n assets. In other words, the problem seeks to equalize the risk contributions of all n assets. However, this implicitly makes an important assumption that requires special consideration from the investor. Consider the following scenario. An investor has a predetermined basket of n assets, all of which are U.S. stocks from the technology sector. If we use these stocks to construct a risk parity portfolio, is the resulting portfolio truly risk-diverse? The nominal risk parity model places the burden of identifying a sufficiently diverse basket of assets on the investor a priori. We propose two avenues that may remedy this problem.

Stemming from our work into factor models, we could use risk factors to systematically diversify the sources of risk in our portfolio. Instead of focusing on being risk-diverse from the asset perspective, we can attempt to do this from the factor perspective. Although expert judgement would still be required to select an appropriate basket of risk factors, the end result would be a portfolio that is risk-diverse from the perspective of the underlying drivers of risk. This is not a new idea and it has been previously studied by Roncalli and Weisang [86]. However, their approach focuses predominantly on the factors and misses the opportunity to be risk-diverse in both factors and assets alike. In this sense, we could use a nested clustered optimization approach to iteratively solve a larger problem with regards to the factors, while promoting risk-based diversification at the asset level through smaller subproblems. Such an approach is the subject of ongoing research.

The second avenue we consider is a cardinality-constrained risk parity optimization problem. Such an approach would lift the burden of asset selection away from the investor and transfer it to the optimization problem. The investor would be free to consider a very large basket

of n assets, such as the constituents of the S&P 500 index ($n = 500$), and simply impose a cardinality limit on the number of desired constituents k in the portfolio, where $k < n$. Such an approach brings a host of new challenges from an optimization perspective. First, the problem would become a mixed integer program, increasing the problem's complexity. Second, we can easily find a trivial risk parity solution to this problem, meaning we will need to model this as a multi-objective problem. These additional objectives may entail the minimization of variance or tracking the return of some underlying index. Moreover, we would have to ensure that the subset k is truly diverse (i.e., the sources of risk in the subset k must be representative of all the sources of risk in the broader basket of n assets). This problem is the subject of ongoing research.

An extension that we may consider in the future is the inclusion of higher moments in the generalized risk parity problem in Chapter 6, such as the portfolio skewness and kurtosis. Including higher moments would require a Lasserre-type relaxation [64] of the problem, followed up by an adaptation of the ADMM algorithm from Chapter 6 to iteratively impose a rank-1 constraint. However, the dimensions of the corresponding SDP subproblem may become prohibitively large, particularly since we must solve a sequence of SDPs until convergence. Thus, solving this problem would require special treatment to retain computational tractability, such as using first-order algorithms to approximate the solution to the SDP subproblem after every iteration. Developing a multi-objective risk parity problem with higher moments can be the subject of future research.

Appendix A

Numerical implementation of statistical distances

This appendix describes how to numerically implement the squared Hellinger distance in (4.17d) and the TV distance in (4.17e). We use either of these two distance measures to define the ambiguity set $\mathcal{U}_{\boldsymbol{p}}$, and then use the set to construct the corresponding Euclidean projection optimization problem in (4.26). However, in their current form, most optimization solvers will reject them.

If we wish to use the squared Hellinger distance in (4.17d), then the projection optimization problem can be implemented as follows.

$$
\begin{aligned}
\min_{\boldsymbol{p},\boldsymbol{w}} \quad & \|\boldsymbol{u}-\boldsymbol{p}\|_2^2 \\
\text{s.t.} \quad & \mathbf{1}^T\boldsymbol{p} = 1 \\
& \frac{1}{2}\sum_{t=1}^{T} p_t - 2w_t\sqrt{q_t} + q_t \leq d_{\mathrm{H}}, \\
& p_t \geq w_t^2, \quad \text{for } t = 1,\ldots,T \\
& \boldsymbol{p},\ \boldsymbol{w} \geq 0,
\end{aligned}
$$

where $\boldsymbol{u} \in \mathbb{R}^T$ is some arbitrary vector that we wish to project onto the set $\mathcal{U}_{\boldsymbol{p}}$, while $\boldsymbol{w} \in \mathbb{R}^T$ is an auxiliary variable that serves as a placeholder for the square root of each element of $\boldsymbol{p}$. As before, $\boldsymbol{q} \in \mathcal{P}$ is the nominal probability distribution, while d_{H} is the maximum permissible distance in (4.21) and is defined by the number of scenarios T and the investor's subjective

confidence level δ.

On the other hand, if we wish to use the TV distance in (4.17e), then the projection optimization problem can be implemented as follows.

$$\begin{aligned}
\min_{\boldsymbol{p},\boldsymbol{w}} \quad & \|\boldsymbol{u}-\boldsymbol{p}\|_2^2 \\
\text{s.t.} \quad & \mathbf{1}^T\boldsymbol{p} = 1 \\
& \frac{1}{2}\sum_{t=1}^{T} w_t \leq d_{\mathrm{TV}}, \\
& w_t \geq p_t - q_t \quad \text{for } t = 1,\ldots,T \\
& w_t \geq q_t - p_t \quad \text{for } t = 1,\ldots,T \\
& \boldsymbol{p} \geq 0,
\end{aligned}$$

where $\boldsymbol{u} \in \mathbb{R}^T$ is some arbitrary vector that we wish to project onto the set $\mathcal{U}_{\boldsymbol{p}}$, while $\boldsymbol{w} \in \mathbb{R}^T$ is an auxiliary variable that represents the absolute value of the difference between the elements of $\boldsymbol{p}$ and $\boldsymbol{q}$. The maximum permissible distance d_{TV} is defined by the number of scenarios T and the investor's subjective confidence level δ as shown in (4.22).

Bibliography

[1] Ang, A. and Bekaert, G. (2002a). International asset allocation with regime shifts. *Review of Financial Studies*, 15(4):1137–1187.

[2] Ang, A. and Bekaert, G. (2002b). Regime switches in interest rates. *Journal of Business & Economic Statistics*, 20(2):163–182.

[3] Ang, A. and Timmermann, A. (2012). Regime changes and financial markets. *Annual Review of Financial Economics*, 4(1):313–337.

[4] Bae, G. I., Kim, W. C., and Mulvey, J. M. (2014). Dynamic asset allocation for varied financial markets under regime switching framework. *European Journal of Operational Research*, 234(2):450–458.

[5] Bai, X., Scheinberg, K., and Tütüncü, R. H. (2016). Least-squares approach to risk parity in portfolio selection. *Quantitative Finance*, 16(3):357–376.

[6] Barzilai, J. and Borwein, J. M. (1988). Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8(1):141–148.

[7] Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, 41(1):164–171.

[8] Bellman, R. E. (1961). *Adaptive control processes: a guided tour*. Princeton University Press.

[9] Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust optimization*. Princeton University Press.

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

[11] Bertsekas, D. P. (1976). On the Goldstein-Levitin-Polyak gradient projection method. *IEEE Transactions on Automatic Control*, 21(2):174–184.

[12] Bertsimas, D. and Sim, M. (2004). The price of robustness. *Operations Research*, 52(1):35–53.

[13] Bertsimas, D. and Sim, M. (2006). Tractable approximations to robust conic optimization problems. *Mathematical Programming*, 107(1-2):5–36.

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

[15] Birge, J. R. and Louveaux, F. (2011). *Introduction to stochastic programming*. Springer Science & Business Media.

[16] Birgin, E. G., Martínez, J. M., and Raydan, M. (2000). Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization*, 10(4):1196–1211.

[17] Booth, D. G. and Fama, E. F. (1992). Diversification returns and asset contributions. *Financial Analysts Journal*, 48(3).

[18] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122.

[19] Breton, M. and El Hachem, S. (1995). Algorithms for the solution of stochastic dynamic minimax problems. *Computational Optimization and Applications*, 4(4):317–345.

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

[21] Calafiore, G. C. (2007). Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, 18(3):853–877.

[22] Chaves, D., Hsu, J., Li, F., and Shakernia, O. (2011). Risk parity portfolio vs. other asset allocation heuristic portfolios. *Journal of Investing*, 20(1):108–118.

[23] Chen, G. and Teboulle, M. (1994). A proximal-based decomposition method for convex minimization problems. *Mathematical Programming*, 64(1-3):81–101.

[24] Chen, X., Sim, M., and Sun, P. (2007). A robust optimization perspective on stochastic programming. *Operations Research*, 55(6):1058–1071.

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

[26] Costa, G. and Kwon, R. H. (2019). Risk parity portfolio optimization under a Markov regime-switching framework. *Quantitative Finance*, 19(3):453–471.

[27] Costa, G. and Kwon, R. H. (2020a). Data-driven distributionally robust risk parity portfolio optimization. *Manuscript submitted for publication*.

[28] Costa, G. and Kwon, R. H. (2020b). Generalized risk parity portfolio optimization: An ADMM approach. *Journal of Global Optimization*, 78:207–238.

[29] Costa, G. and Kwon, R. H. (2020c). A regime-switching factor model for mean–variance optimization. *Journal of Risk*, 22(4):31–59.

[30] Costa, G. and Kwon, R. H. (2020d). A robust framework for risk parity portfolios. *Journal of Asset Management*, 21:447–466.

[31] Dai, Y.-H. and Fletcher, R. (2005). Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming. *Numerische Mathematik*, 100(1):21–47.

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

[33] Dunning, I., Huchette, J., and Lubin, M. (2017). Jump: A modeling language for mathematical optimization. *SIAM Review*, 59(2):295–320.

[34] Dupačová, J. (1987). The minimax approach to stochastic programming and an illustrative application. *Stochastics: An International Journal of Probability and Stochastic Processes*, 20(1):73–88.

[35] Eckstein, J. and Bertsekas, D. P. (1992). On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1-3):293–318.

[36] Eckstein, J. and Fukushima, M. (1994). Some reformulations and applications of the alternating direction method of multipliers. In *Large scale optimization*, pages 115–134. Springer.

[37] Endres, D. M. and Schindelin, J. E. (2003). A new metric for probability distributions. *IEEE Transactions on Information Theory*, 49(7):1858–1860.

[38] Fabozzi, F. J., Kolm, P. N., Pachamanova, D. A., and Focardi, S. M. (2007). Robust portfolio optimization. *Journal of Portfolio Management*, 33(3):40.

[39] Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

[40] Feng, Y. and Palomar, D. P. (2015). SCRIP: Successive convex optimization methods for risk parity portfolio design. *IEEE Transactions on Signal Processing*, 63(19):5285–5300.

[41] Fortin, M. and Glowinski, R. (1983). On decomposition-coordination methods using an augmented Lagrangian. In Fortin, M. and Glowinski, R., editors, *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*, volume 15. Elsevier.

[42] French, K. R. (2020). Data library. `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. [Online; accessed 20-May-2020].

[43] Fuglede, B. and Topsoe, F. (2004). Jensen-Shannon divergence and Hilbert space embedding. In *International Symposium on Information Theory, 2004. ISIT 2004. Proceedings.*, page 31. IEEE.

[44] Fukushima, M. (1992). Application of the alternating direction method of multipliers to separable convex programming problems. *Computational Optimization and Applications*, 1(1):93–111.

[45] Gabay, D. (1983). Applications of the method of multipliers to variational inequalities. In Fortin, M. and Glowinski, R., editors, *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*, volume 15, pages 299–331. Elsevier.

[46] Gabay, D. and Mercier, B. (1975). *A dual algorithm for the solution of non linear variational problems via finite element approximation*. Institut de recherche d'informatique et d'automatique.

[47] Ghadimi, E., Teixeira, A., Shames, I., and Johansson, M. (2015). Optimal parameter selection for the alternating direction method of multipliers (admm): quadratic problems. *IEEE Transactions on Automatic Control*, 60(3):644–658.

[48] Glowinski, R. and Marroco, A. (1975). Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. *Revue Française d'Automatique, Informatique, Recherche Opérationnelle. Analyse Numérique*, 9(R2):41–76.

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

[50] Goldstein, T., O'Donoghue, B., Setzer, S., and Baraniuk, R. (2014). Fast alternating direction optimization methods. *SIAM Journal on Imaging Sciences*, 7(3):1588–1623.

[51] Grippo, L., Lampariello, F., and Lucidi, S. (1986). A nonmonotone line search technique for Newton's method. *SIAM Journal on Numerical Analysis*, 23(4):707–716.

[52] Guastaroba, G., Mitra, G., and Speranza, M. G. (2011). Investigating the effectiveness of robust portfolio optimization techniques. *Journal of Asset Management*, 12(4):260–280.

[53] Guidolin, M. (2011). Markov switching models in empirical finance. In *Missing data methods: Time-series methods and applications*, pages 1–86. Emerald Group Publishing Limited.

[54] Guidolin, M. and Timmermann, A. (2007). Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control*, 31(11):3503–3544.

[55] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384.

[56] Hamilton, J. D. (2010). Regime switching models. In *Macroeconometrics and Time Series Analysis*, pages 202–209. Springer.

[57] Haugh, M., Iyengar, G., and Song, I. (2017). A generalized risk budgeting approach to portfolio construction. *Journal of Computational Finance*, 21(2):29–60.

[58] He, B., Yang, H., and Wang, S. (2000). Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities. *Journal of Optimization Theory and Applications*, 106(2):337–356.

[59] Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(3):279–292.

[60] Kapsos, M., Christofides, N., and Rustem, B. (2018). Robust risk budgeting. *Annals of Operations Research*, 266(1-2):199–221.

[61] Kim, S.-J. and Boyd, S. (2008). A minimax theorem with applications to machine learning, signal processing, and finance. *SIAM Journal on Optimization*, 19(3):1344–1367.

[62] Kritzman, M., Page, S., and Turkington, D. (2012). Regime shifts: Implications for dynamic strategies. *Financial Analysts Journal*, 68(3):22–39.

[63] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86.

[64] Lasserre, J. B. (2001). Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817.

[65] Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5):603–621.

[66] Levy, M. and Kaplanski, G. (2015). Portfolio selection in a two-regime world. *European Journal of Operational Research*, 242(2):514–524.

[67] Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151.

[68] Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, pages 13–37.

[69] Lobo, M. S. and Boyd, S. (2000). The worst-case risk of a portfolio. *Unpublished manuscript. Available from http://faculty. fuqua. duke. edu/% 7Emlobo/bio/researchfiles/rsk-bnd. pdf.*

[70] Maillard, S., Roncalli, T., and Teïletche, J. (2010). The properties of equally weighted risk contribution portfolios. *Journal of Portfolio Management*, 36(4):60–70.

[71] Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.

[72] Mausser, H. and Romanko, O. (2014). Computing equal risk contribution portfolios. *IBM Journal of Research and Development*, 58(4):5–1.

[73] Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361.

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

[75] Michaud, R. O. and Michaud, R. (2007). Estimation error and portfolio optimization: a resampling solution. Available at SSRN: https://ssrn.com/abstract=2658657 or http://dx.doi.org/10.2139/ssrn.2658657.

[76] Michaud, R. O. and Michaud, R. O. (2008). *Efficient asset management: a practical guide to stock portfolio optimization and asset allocation*. Oxford University Press.

[77] Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica: Journal of the Econometric Society*, pages 768–783.

[78] Nedić, A. and Ozdaglar, A. (2009). Subgradient methods for saddle-point problems. *Journal of Optimization Theory and Applications*, 142(1):205–228.

[79] Nesterov, Y. (2013). *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media.

[80] Neumann, J. v. (1928). Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320.

[81] Nystrup, P., Madsen, H., and Lindström, E. (2018). Dynamic portfolio optimization across hidden market regimes. *Quantitative Finance*, 18(1):83–95.

[82] Popescu, I. (2007). Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55(1):98–112.

[83] Qian, E. (2005). Risk parity portfolios: Efficient portfolios through true diversification. *Panagora Asset Management*.

[84] Quandl.com (2017). Wiki – various end-of-day stock prices. [Online; accessed 07-Nov-2017].

[85] Raghunathan, A. U. and Di Cairano, S. (2014). Alternating direction method of multipliers for strictly convex quadratic programs: Optimal parameter selection. In *American Control Conference (ACC), 2014*, pages 4324–4329. IEEE.

[86] Roncalli, T. and Weisang, G. (2016). Risk parity portfolios with risk factors. *Quantitative Finance*, 16(3):377–388.

[87] Rustem, B. and Howe, M. (2009). *Algorithms for worst-case design and applications to risk management*. Princeton University Press.

[88] Scarf, H. (1958). A min-max solution of an inventory problem. *Studies in the Mathematical Theory of Inventory and Production*, pages 201–209.

[89] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464.

[90] Shapiro, A. and Ahmed, S. (2004). On a class of minimax stochastic programs. *SIAM Journal on Optimization*, 14(4):1237–1249.

[91] Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2014). *Lectures on stochastic programming: modeling and theory*. SIAM.

[92] Shapiro, A. and Kleywegt, A. (2002). Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17(3):523–542.

[93] Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442.

[94] Sharpe, W. F. (1994). The Sharpe ratio. *Journal of Portfolio Management*, 21(1):49–58.

[95] Tseng, P. (1991). Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM Journal on Control and Optimization*, 29(1):119–138.

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

[97] Wang, S. and Liao, L. (2001). Decomposition method with a variable parameter for a class of monotone variational inequality problems. *Journal of Optimization Theory and Applications*, 109(2):415–429.

[98] Xiu, N. and Zhang, J. (2003). Some recent advances in projection-type methods for variational inequalities. *Journal of Computational and Applied Mathematics*, 152(1-2):559–585.

[99] You, S. and Peng, Q. (2014). A non-convex alternating direction method of multipliers heuristic for optimal power flow. In *Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on*, pages 788–793. IEEE.

[100] Žáčková, J. (1966). On minimax solutions of stochastic linear programming problems. *Časopis pro Pěstování Matematiky*, 91(4):423–430.