Least-squares approach to risk parity in portfolio selection

Xi Bai*[1], Katya Scheinberg†[1], and Reha Tutuncu‡[2]

[1]Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, 18015, USA

[2]Goldman Sachs Asset Management, New York, NY, 10282, USA.

Original Publication: May 26, 2013

Last Revised: Oct. 21, 2013

Abstract

The risk parity optimization problem aims to find such portfolios for which the contributions of risk from all assets are equally weighted. Portfolios constructed using risk parity approach are a compromise between two well-known diversification techniques: minimum variance optimization approach and the equal weighting approach. In this paper, we discuss the problem of finding portfolios that satisfy risk parity of either individual assets or groups of assets. We describe the set of all risk parity solutions by using convex optimization techniques over orthants and we show that this set may contain exponential number of solutions. We then propose an alternative nonconvex least-square model whose set of optimal solutions includes all risk parity solutions, and propose a modified formulation which aims at selecting the most desirable risk parity solution (according to some criteria). When general bounds are considered, a risk parity solution may not exist. The nonconvex least-square model then seeks a feasible portfolio which is as close to risk parity as possible. Furthermore, we propose an alternating linearization framework to solve this nonconvex model. Numerical experiments indicate the effectiveness of our technique in terms of both speed and accuracy.

Keywords: Asset allocation, risk parity, alternating direction method, alternating linearization method.

1 Introduction

Portfolio construction has become a focus of many researchers in the past decades. One central goal in portfolio construction is to manage the tradeoff between return and risk. Since 1950s, many optimization models have appeared and they continue to play an important role in making

*E-mail: xib210@lehigh.edu,

†E-mail: katyas@lehigh.edu,

‡E-mail: reha.tutuncu@gs.com

investment decisions. However, some strategies resulting from optimal portfolio construction have been observed to be inefficient in practice. One of the the reasons is the lack of robustness of the optimal portfolios with respect to input parameters. When estimated from historical information, these parameters can be inaccurate predictors of future behavior of the security returns, and using them optimal portfolio construction can lead to inefficient portfolios.

One well-known example of an optimal portfolio construction strategy is the mean-variance optimization model. Proposed by Markowitz, mean-variance approach has been regarded as a fundamental framework in portfolio construction. It offered the first quantitative insight into the tradeoff between returns and risk. One persistent criticism of the mean-variance model has been its sensitivity to inputs. Among others, research by Best and Grauer [2] has shown that slight changes of input (especially in expected return estimate) often lead to dramatic changes in the optimal portfolio composition.

In practice, some investors apply much simpler portfolio construction strategies and might suffer from their limitations as a consequence. One of these simpler approaches is the well-known "60/40" strategy (60% equity, 40% bonds). This method usually has high expected portfolio return, but the portfolio volatility may be dominated by the equity risk. Another simple method is the "1/n" method, namely, portfolio construction with equal weights (EW) in all asset classes [6]. Equally-weighted method can be seen as a diversification strategy. But the diversification is achieved only at the capital allocation level and not in terms of risk contributions, simply because the approach does not utilize any information on the assets' volatility or their correlations.

In this paper, we investigate another portfolio strategy— the, so-called, risk parity approach. The idea of risk parity is not new and can be regarded as a special type of diversification strategy. Using volatility as the risk measure, the risk parity approach aims to create a portfolio with equal risk contributions from each of the assets in the portfolio. The past few years have witnessed a fair amount of risk parity research (for instance, [4, 5, 9]). Most of the work considers long-only portfolios and solve a risk budgeting problem which aims to equalize the total risk contribution for each asset. In this case, we demonstrate that a risk parity portfolio can be obtained from a solution of a convex optimization problem.

If long-short portfolios are considered, multiple solutions may exist. We show that a convex model still applies if the investor identifies in advance which assets should be shorted. In other words, it is possible to obtain a unique risk parity solution if we know the orthant in which the solution lies. However, if the orthant is not given, we show that there could be a combinatorially large number of solutions that satisfy risk parity and these solutions can be identified through an enumeration strategy. In practice, investors may consider adding general bounds on the individual weights of the assets, in which case the total number of risk parity solutions may be small. Moreover, if the bounds are sufficiently tight a risk parity solution satisfying such bounds may not exist at all.

In this paper we propose a generalized risk parity model which allows for short sales and applies to cases where risk parity solutions may not exist. Our model is similar to the model proposed in [9] in that we minimize a function that measures deviation from risk parity. However, our formulation has a simpler structure and allows for easier analysis and efficient algorithmic approaches. As in [9], the optimization model that we consider is not convex. Each risk parity solution is a global optimum of the model. Moreover, no local optima exist on the interior of the feasible set; in other words, all local optima of our proposed formulation occur due to the constraining of the feasible set by bounds on the weights. In the case when the bounds are tight and no feasible risk parity solution exists, then the global optimum is the solution which is "closest" to the risk parity, in the least-square sense. We develop an algorithmic framework based on alternating linearization methods (ALMs) to solve our generalized model. The framework is simple and convergent to a local optimum which is guaranteed to be a global optimum when no constraints are binding and relies on solving a sequence of convex quadratic subproblems. Our formulation also easily extends to the case of the multiple objectives - choosing the best risk parity solution according to some additional criterion, when multiple solutions exist.

The rest of the paper is organized as follows. After a brief discussion of the Markowitz framework in Section 2, we introduce the concept of risk parity and consider the convex log-barrier model as well as the new least-square model. In Section 3, we propose a class of algorithmic methods based on the ALM framework, to solve the least-square model. In Section 4, we discuss some extensions based on the least-square model, where we aim to choose the best risk parity solution. We also propose another extension of the risk parity problem to the case where we seek parity on risk contributions of a group of assets rather than individual assets. Experiments and computational results are discussed in Section 5, followed with conclusion remarks.

2 Risk parity problem

Numerous methods based on the famous Markowitz mean-variance framework have been proposed to overcome its drawbacks while maintaining its advantages (see, for instance, [11]). In this paper, we focus on risk based diversification strategies. Unlike the classic mean-variance approach, risk based strategies do not incorporate expected returns into the formulation. Motivations for not using expected returns in the portfolio construction include the difficulty of estimating these quantities accurately, and the well documented sensitivity of the optimal weights to small changes in expected returns. Since they do not rely on these parameters risk based strategies are considered to be more robust than approaches using expected returns [9].

One prominent example of risk-based strategies in portfolio selection is the minimum variance optimization approach. This approach aims to minimize the portfolio's volatility and can be formulated as a convex quadratic optimization problem. This problem can be solved efficiently using widely available optimization software and typically has a unique solution. Here we briefly introduce the minimum variance optimization model to compare it to the risk parity approach.

Suppose we have n risky assets. Their covariance is given by a symmetric matrix Σ which is assumed to be positive semidefinite. The following optimization problem minimizes the total variance of a fully-invested long-only portfolio:

$$\begin{array}{ll} \min_x & \frac{1}{2}x^T\Sigma x \\ \text{s.t.} & x_i \geq 0 \\ & \sum_{i=1}^n x_i = 1, \end{array} \tag{2.1}$$

where $x = [x_1, x_2, ..., x_n]^T$ is the vector of the weights of n assets. The factor $1/2$ in the objective is introduced to simplify the optimality conditions and has no impact on the optimal portfolio. From the first-order optimality conditions of the above problem we see that

$$\Sigma x - \lambda - \gamma \mathrm{e} = 0, \tag{2.2}$$

where e is an n-dimensional vector of all ones, $\lambda \in \mathbb{R}^n$ and $\gamma \in \mathbb{R}$ are the Lagrange multipliers corresponding to the long-only and full investment constraints, respectively. Note that, complementary slackness conditions imply that if some x_i is strictly larger than zero, then the corresponding λ_i must be zero. Combining this observation with (2.2), we obtain

$$(\Sigma x)_i = \gamma, \forall i \text{ s.t. } x_i \neq 0. \tag{2.3}$$

From (2.3) we obtain:

$$\left(\frac{\Sigma x}{\sqrt{x^T \Sigma x}}\right)_i = \left(\frac{\Sigma x}{\sqrt{x^T \Sigma x}}\right)_j, \forall i, j \text{ s.t. } x_i, x_j \neq 0. \tag{2.4}$$

Also, note that $\frac{\partial \sigma}{\partial x} = \frac{\Sigma x}{\sqrt{x^T \Sigma x}}$ is the vector of *marginal risk contributions* for the assets in the portfolio. Hence, finally we have

$$\frac{\partial \sigma}{\partial x_i} = \frac{\partial \sigma}{\partial x_j}, \forall i, j \text{ s.t.} x_i, x_j > 0. \tag{2.5}$$

The above condition implies that, as long as we invest in an asset, its marginal risk contribution should be the same as that of all other assets with positive weights in the portfolio. As such, minimum variance approach leads to portfolios with equal marginal risk contributions. In practice, while dominating other strategies from the perspective of low volatility, the minimum variance approach often leads to concentrated portfolios, i.e., encourages investors to concentrate on a small number of assets with lower risk profiles and to give up diversification. This behavior is often undesirable and this is exactly what risk parity optimization intends to overcome.

Risk parity portfolios can be motivated by considering Euler decomposition of a portfolio risk measure into contributions from each asset in the portfolio.

Theorem 2.1. *(Euler's theorem) Let a continuous and differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ be a homogeneous function of degree one. Then*

$$f(x) \quad = \quad x_1 \cdot \tfrac{\partial f}{\partial x_1} + x_2 \cdot \tfrac{\partial f}{\partial x_2} + \ldots + x_n \cdot \tfrac{\partial f}{\partial x_n} \ ,$$

where f is a homogeneous function of degree one if for any constant $c \in \mathbb{R}$, $f(cx) = c \cdot f(x)$.

Simply put, a risk parity portfolio is a portfolio where the total contribution of each asset to the total portfolio risk is equal. In this paper, we make the common choice of using volatility $\sigma(x) = (x^T \Sigma x)^{\frac{1}{2}}$ as our risk measure. In this context, the risk parity problem aims to find any portfolio that satisfies

$$x_i \cdot \frac{\partial \sigma}{\partial x_i} = x_j \cdot \frac{\partial \sigma}{\partial x_j}, \forall i, j. \tag{2.6}$$

We will refer to any solution of (2.6) as a *risk parity solution.* If the weight vector is normalized, then (2.6) becomes the normalized risk parity problem

$$\begin{array}{l} x_i \cdot \frac{\partial \sigma}{\partial x_i} = x_j \cdot \frac{\partial \sigma}{\partial x_j}, \forall i, j \\ \sum_{i=1}^n x_i = 1. \end{array} \tag{2.7}$$

We will refer to any solutions to (2.7) as *normalized risk parity solution.* Note that one can easily convert a solution of (2.6) into a solution of (2.7) through simple scaling, as long as the sum of the asset weights is non-zero.

We will also consider situations where risk parity solutions may not exist because of the presence of additional restrictions on the portfolio weights. In such cases, we will look for portfolios that are "close to risk parity" and for this purpose it will be important to quantify the deviation from risk parity. We address these concerns in Section 2.3.

Let us assume for a moment that the correlation between any two assets is a constant, that is $\rho_{ij} = \rho, \forall i, j$. It can be shown that a closed form solution can be deduced with this assumption [9], and such a solution coincides with the solution under minimum variance optimization approach. This solution is given by the following identity: $x_i = \frac{\sigma_i^{-1}}{\sum_{j=1}^n \sigma_j^{-1}}$. Some literature refers to this case as the "naive risk parity" solution. When the correlations are not constant, a closed-form solution for x does not exist in general and numerical approaches need to be applied.

2.1 Long-only risk parity via convex optimization

Next, we consider the problem of finding a long-only risk parity solution, i.e., finding a vector of weights of n assets $x = [x_1, x_2, ..., x_n]^T$ such that $x_i \cdot \frac{\partial \sigma}{\partial x_i} = x_j \cdot \frac{\partial \sigma}{\partial x_j}, \forall i, j$ and $x \geq 0$. In this case, in turns out that solving an artifical optimization problem that incorporates a logarithmic barrier term is equivalent to finding a risk parity solution:

$$\begin{array}{ll} \min_x & \frac{1}{2} x^T \Sigma x - c \sum_{i=1}^n \ln x_i \\ \text{s.t.} & x_i > 0, \end{array} \tag{2.8}$$

where Σ is the covariance matrix, c is an arbitrary positive constant. Our use of the logarithmic barrier term in the objective function of (2.8) is motivated by a related formulation in [9] that uses a constraint incorporating the sum of the logarithms. Other authors have also utilized the logarithmic barrier function for solving the risk parity problem in independently developed studies [8, 10].

Since Σ is positive semidefinite and the logarithm function is strictly concave, the objective function of (2.8) is strictly convex. We observe that this convex optimization problem has a unique solution at the point where the gradient of the objective function, $\Sigma x - c x^{-1}$ is zero (first-order condition), where $x^{-1} = [1/x_1, 1/x_2, ..., 1/x_n]^T$. Hence, at optimality we have $(\Sigma x)_i = \frac{c}{x_i}, \forall i$, which leads to

$$x_i (\Sigma x)_i = x_j (\Sigma x)_j, \forall i, j. \tag{2.9}$$

It is now easy to see that (2.9) is equivalent to $x_i \cdot \frac{\partial \sigma}{\partial x_i} = x_j \cdot \frac{\partial \sigma}{\partial x_j}, \forall i, j$ and that risk parity is achieved at the unique optimal solution of (2.8).

There is no guarantee that the weights in the solution of (2.8) will sum to one, so the result may represent a levered or an under-invested portfolio. Fortunately, we have the following result showing the existence and uniqueness of the risk parity solution in the long-only case if we impose the additional constraint that the sum of all weights equals to one.

Lemma 2.1. *Let Σ be a positive semidefinite covariance matrix. Then there exists a unique solution x which satisfies:*

$$x_i (\Sigma x)_i = x_j (\Sigma x)_j, \forall i, j \tag{2.10}$$

$$\sum_{i=1}^n x_i = 1, \quad x_i > 0, \quad \forall i. \tag{2.11}$$

In fact, any two long-only risk parity solutions differ by a constant factor.

Proof. Since (2.8) is strictly convex, it has a unique solution, say, x^c for any given $c > 0$. Consider the following optimization problem:

$$\begin{array}{ll} \min_x & \frac{1}{2}x^T\Sigma x - \alpha c\sum_{i=1}^n \ln x_i \\ \text{s.t.} & x_i > 0, \end{array} \tag{2.12}$$

where α is a positive scalar. It is easy to verify that $\sqrt{\alpha}x^c$ is the unique solution for this problem since $\sqrt{\alpha}x_i^c\,[\Sigma(\sqrt{\alpha}x^c)]_i = \alpha c$. Hence, as α varies from 0 to ∞, $\sum_{i=1}^n \sqrt{\alpha}x_i^c$ varies from 0 to ∞. Further, for $\alpha^* = \frac{1}{(\sum_{k=1}^n x_k^c)^2}$, the solution of (2.12), denoted as x^*, satisfies $\sum_{i=1}^n x_i^* = 1$. It is easy to see that $x_i^* = \frac{x_i}{\sum_{k=1}^n x_k^c}$ and x^* is unique and independent of initial choice of c.

□

2.2 Risk parity solutions over orthants

In this subsection, we explore the set of risk parity solutions when the long-only restriction on weights is removed. The approach discussed in Section 2.1 identifies only the solution in the nonnegative orthant. To find solutions in other orthants, we consider the following modified log-barrier approach:

$$\begin{array}{ll} \min_x & \frac{1}{2}x^T\Sigma x - c\sum_{i=1}^n \ln \beta_i x_i \\ \text{s.t.} & \beta_i x_i > 0, \end{array} \tag{2.13}$$

where $\beta = [\beta_1, \beta_2, \ldots, \beta_n]^T \in \{-1, 1\}^n$, defines the orthant in which we are seeking a solution. If $\beta_i = 1$, then $x_i \in (0, +\infty)$; otherwise, $x_i \in (-\infty, 0)$. For each choice of β, (2.13) is a convex optimization problem and thus it has a unique solution x^β. Since there are 2^n such different β, there are 2^n such solutions. Let $\bar{\beta} = -\beta$ define the complementary orthant of the orthant defined by β. Then, it is easy to see that $x^\beta = -x^{\bar{\beta}}$. We are primarily interested in scaled solutions to have the following lemma.

Lemma 2.2. *Let Σ be a positive semidefinite covariance matrix. Then there exist at most 2^{n-1} solutions which satisfy $x_i(\Sigma x)_i = x_j(\Sigma x)_j, \forall i, j;$* and $\sum_{i=1}^n x_i = 1$.

Proof. For each β and x^β, $x_N^\beta = \frac{x^\beta}{\sum_{i=1}^n x_i^\beta}$ is a normalized risk parity solution satisfying (2.7) as long as $\sum_{i=1}^n x_i^\beta \neq 0$. Note that $\frac{x_i}{\sum_{i=1}^n x_i} = \frac{-x_i}{\sum_{i=1}^n (-x_i)}$. Also note that for all (unnormalized) risk parity solutions in the same orthant we either have $\sum_{i=1}^n x_i^\beta = 0$ or $\sum_{i=1}^n x_i^\beta < 0$ or $\sum_{i=1}^n x_i^\beta > 0$. Hence, for every choice of β, $\bar{\beta} = -\beta$ generates the same scaled risk parity solution x_N^β, and thus there are at most 2^{n-1} such solutions. In other words, for each orthant, the normalized risk parity solution exists if and only if $\sum_{i=1}^n x_i^\beta \neq 0$ for any of the (unnormalized) risk parity solutions in that orthant. In case when $\sum_{i=1}^n x_i^\beta < 0$ then the normalized risk parity solution lies in the complementary orthant.

□

Remark. Note that a scaled risk parity solution does not exist if $\sum_{i=1}^n x_i^\beta = 0$ in Lemma 2.2 (hence there may be fewer than 2^{n-1} normalized solutions). However, when $\sum_{i=1}^n x_i^\beta = 0$, the portfolio is "market-neutral" i.e., the aggregate exposure to the market is zero. This case is independently interesting and can sometimes be desirable in portfolio selection.

To illustrate Lemma 2.2 with a simple example, let us consider the simple case when the covariance matrix is diagonal, i.e.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \dots & \\ & & & \sigma_n^2 \end{bmatrix}.$$

Then risk parity is equivalent to the following system which contains $(n-1)$ linearly independant equations:

$$\begin{aligned} \beta_1\sigma_1 x_1 &= \beta_2\sigma_2 x_2 \\ \beta_1\sigma_1 x_1 &= \beta_3\sigma_3 x_3 \\ &\dots \\ \beta_1\sigma_1 x_1 &= \beta_n\sigma_n x_n, \end{aligned} \tag{2.14}$$

where $\beta_i \in \{-1, 1\}, \forall i \in \{1, ..., n\}$. Suppose $\sigma_i^{-1} + \sum_{j\neq i} \frac{\beta_j}{\beta_i}\sigma_j^{-1} \neq 0$. Then (2.14), together with $\sum_{i=1}^n x_i = 1$, leads to a set of closed form solutions

$$x_i = \frac{\sigma_i^{-1}}{\sigma_i^{-1} + \sum_{j\neq i} \frac{\beta_j}{\beta_i}\sigma_j^{-1}}. \tag{2.15}$$

Note that there are at most 2^{n-1} different solutions (since there are 2^n different β, and each pair $(\beta, -\beta)$ give the same system of equations (2.14)). Additionally, in (2.14), if $\beta_i = 1, \forall i \in \{1, ..., n\}$, then (2.15) becomes $x_i = \frac{\sigma_i^{-1}}{\sum_j \sigma_j^{-1}}$, which is simply the "naive risk parity" solution discussed previously.

Given that there may be exponentially many different risk-parity solutions in the long-short case, additional preferences (e.g., the risk-parity solution with the least volatility) or restrictions on the weights (e.g., lower/upper bounds on weights) can be used to narrow down these choices. If, for example, investors know a priori which assets are desirable to short, then the log barrier approach of Section 2.1 can be easily extended to find the desired long-short risk-parity portfolio by selecting the appropriate β in (2.13). In the next two subsections, we discuss efficient methods for finding solutions satisfying additional preferences or restrictions without resorting to enumeration.

2.3 Least-square model with general bounds

The log-barrier approach to finding risk parity solutions in the long-only setting does not immediately extend to scenarios with additional constraints or preferences. In particular, when general bounds are added, risk parity solution may not exist, however, up to 2^{n-1} different instances of (2.13) may have to be solved before infeasibility can be established. Moreover, the log barrier formulation gives no guidance on how to produce feasible solution which may be "close to risk parity". Simple approaches, such as projecting infeasible normalized risk parity solutions onto the feasible region defined by the constraints may generate solutions that largely deviate from risk parity. In addition, it is not clear how to extend this approach to the cases when risk parity is desirable not for individual assets but for groups of assets (e.g., grouped by industry).

In this section we propose a least-squares formulation for solving the risk parity problem. Our formulation is similar to the following formulation proposed in [9]:

$$\begin{array}{ll} \min_x & \sum_{i=1,j=1}^n (x_i(\Sigma x)_i - x_j(\Sigma x)_j)^2 \\ \text{s.t.} & a_i \le x_i \le b_i \\ & \sum_{i=1}^n x_i = 1. \end{array} \tag{2.16}$$

Above, a_i and b_i are arbitrary constants representing the bounds on the weight of ith asset (a_i can be less than zero if we allow short sales). The objective function of (2.16) introduces a penalty term for each pair of risk contribution terms $x_i(\Sigma x)_i$ and $x_j(\Sigma x)_j$ that are different from each other. Alternatively, one can consider using penalty terms for deviations of risk contributions from their average value:

$$\min_x \sum_{i=1}^n (x_i(\Sigma x)_i - \frac{\sum_{j=1}^n x_j(\Sigma x)_j}{n})^2.$$

Our formulation is based on this second objective function, but replaces the average risk contribution term with a free variable θ that is also optimized:

$$\begin{array}{ll} \min_{x,\theta} & \sum_{i=1}^n (x_i(\Sigma x)_i - \theta)^2 \\ \text{s.t.} & a_i \le x_i \le b_i \\ & \sum_{i=1}^n x_i = 1, \end{array} \tag{2.17}$$

For future reference, we denote by $F(x,\theta)$ the objective function of (2.17):

$$F(x,\theta) := \sum_{i=1}^n (x_i(\Sigma x)_i - \theta)^2. \tag{2.18}$$

If the optimization problem (2.17) has an optimal value of zero, then risk parity is achieved. Otherwise, the value of the objective function $F(x,\theta)$ can be regarded as a minimum variance measure towards our goal.

The two formulations (2.16) or (2.17) are equivalent in the sense that any risk parity solution is a solution to both optimization problems. However, our formulation (2.17) offers a much simpler form of the objective function containing only n elements in the sum, while the formulation from [9], contains an order of n^2 elements. Hence, our formulation is computationally less demanding, is easier to analyze and contains fewer nonlinearities. Moreover it allows us to develop efficient optimization approaches as will be seen in Section 5. The auxiliary variable θ can always be set to its optimal value based on the following lemma, however, allowing θ to be a free variable significantly simplifies the formulation.

Lemma 2.3. *Given a solution x, there exists one and only one θ^* such that* (2.17) *is minimized, and $\theta^* = (\sum_{i=1}^n x_i(\Sigma x)_i)/n$.*

Proof. With a given x, (2.17) is a strictly convex unconstrained function of θ. Further, the function is minimized when first-order optimality is satisfied, which implies $\theta^* = (\sum_{i=1}^n x_i(\Sigma x)_i)/n$. □

Example 2.1. Consider three assets with volatilities $\sigma_1 = 1, \sigma_2 = 1, \sigma_3 = 2$, respectively and correlation matrix is given by

$$Cor = \begin{bmatrix} 1.0 & & \\ & 1.0 & \\ & & 1.0 \end{bmatrix}.$$

Thus, the covariance matrix is

$$\Sigma = \begin{bmatrix} 1.0 & & \\ & 1.0 & \\ & & 4.0 \end{bmatrix}.$$

The normalized risk-parity solution in the positive orthant is $x^0 = [0.4; 0.4; 0.2]$. Now, suppose we have a restricted feasible region: $0.5 \le x_1 \le \infty, x_2, x_3 \in \mathbb{R}^+$. If we project x^0 onto the feasible region, we obtain $x^p = [0.5; 0.35; 0.15]$. This solution is not "optimal" from the perspective of risk parity solution and the objective of (2.17). To see that, consider another solution which is obtained by optimizing (2.17): $x^{opt} = [0.5; 0.333; 0.167]$. We see that the objective function value at x^p equals 0.0143, while at x^{opt} it equals 0.0128.

Hence, in terms of risk parity, x^{opt} is preferable to x^p. Let us compare the two solutions in terms of risk concentration. For instance, we can compare the highest risk contribution and Herfindahl index. The highest risk contribution is defined as

$$HRC(x) := \max_i \frac{x_i(\Sigma x)_i}{x^T \Sigma x}. \tag{2.19}$$

Then, $HRC(x^p) = 0.5405$, while $HRC(x^{opt}) = 0.5292$. Another widely used measure for risk concentration is Herfindahl index, which is a method measuring risk concerntration [9]. Herfindahl index is defined as

$$h(x) = \sum_{i=1}^{n} \left[\frac{x_i(\Sigma x)_i}{x^T \Sigma x} \right]^2. \tag{2.20}$$

Hence, if Herfindahl index is 1, it stands for a perfectly concentrated portfolio; if Herfindahl index is $\frac{1}{n}$, then risk is perfectly separated and parity is achieved. We observe that $h_{x^p} = 0.4002$, and $h_{x^{opt}} = 0.3909$. While the differences are small, both measures indicate that x^{opt} is a better solution in terms of risk concentration, hence optimizing the objection function of (2.17) is desirable.

Note that, unlike probelms (2.1) and (2.8), (2.17) is a non-convex problem, hence in theory it is harder to solve and may produce local solutions. However, as we will show it is a useful formulation, as simple practical and fast optimization schemes can be developed for this problem and this formulation can be extended to include additional optimization criteria and different variants of risk parity. Moreover, as we show in the next section, if the constraints of (2.17) are removed (as is done when (2.8) is applied), then any local optimal solution is a global one. Hence our model finds the global risk parity solution whenever approach using (2.8) can find it.

3 Local/global optima issues

Let us consider the first-order optimality conditions for (2.17). The Lagrangian can be written as

$$\mathcal{L}(x, \theta) = \sum_{i=1}^{n} (x_i(\Sigma x)_i - \theta)^2 - \lambda_a^T(x - a) - \lambda_b^T(b - x) + \gamma(\sum_{i=1}^{n} x_i - 1), \tag{3.1}$$

where $\lambda_a, \lambda_b \in \mathbb{R}^n_+, \gamma \in \mathbb{R}$. Now, we can write down the KKT conditions as follows

$$\begin{array}{l} \frac{\partial \mathcal{L}}{\partial x_i} = \frac{\partial F}{\partial x_i} - (\lambda_a)_i + (\lambda_b)_i + \gamma = 0,\ \forall i \\ \frac{\partial \mathcal{L}}{\partial \theta} = \frac{\partial F}{\partial \theta} = 0 \\ \sum_{i=1}^n x_i - 1 = 0 \\ a_i \le x_i \le b_i,\ \forall i \\ \lambda_a, \lambda_b \ge 0 \\ (\lambda_a)_i(x_i - a_i) = 0,\ (\lambda_b)_i(x_i - b_i) = 0. \end{array} \tag{3.2}$$

The gradient of F with respect to x is

$$\nabla_x F(x, \theta) = 2\sum_{i=1}^n (x_i(\Sigma x)_i - \theta)(e_i\Sigma_i + (e_i\Sigma_i)^T)x, \tag{3.3}$$

where $e_i \in \mathbb{R}^{n\times 1}$ is the ith column of the identity and $\Sigma_i \in \mathbb{R}^{1\times n}$ is the ith row of the covariance matrix.

We have the following lemma.

Lemma 3.1. *A solution pair $\{x, \theta\}$ is a global optimum with $F(x, \theta) = 0$ if and only if $\nabla_x F = 0, \frac{\partial F}{\partial \theta} = 0$.*

Proof. 1) If $F = 0$, then $x_i(\Sigma x)_i - \theta = 0,\ \forall i$, from which $\nabla_x F = 2\sum_{i=1}^n (x_i(\Sigma x)_i - \theta)(e_i\Sigma_i + (e_i\Sigma_i)^T)x = 0$ holds trivially.

2) If $\nabla_x F = 0$, then $x^T\nabla_x F$=0 ; hence

$$2\sum_{i=1}^n (x_i(\Sigma x)_i - \theta)x^T(e_i\Sigma_i + (e_i\Sigma_i)^T)x = 2\sum_{i=1}^n (x_i(\Sigma x)_i - \theta)(x_i(\Sigma x)_i + (x_i(\Sigma x)_i)^T) = 0. \tag{3.4}$$

Let $B_i = x_i(\Sigma x)_i$, note that $B_i = B_i^T$. Then, ignoring the constant factor, we have

$$\sum_{i=1}^n (B_i - \theta)B_i = 0. \tag{3.5}$$

Applying the second condition $\frac{\partial F}{\partial \theta} = 0$, which implies $\theta = \frac{\sum_{i=1}^n B_i}{n}$, we have

$$n\sum_{i=1}^n B_i^2 = (\sum_{i=1}^n B_i)^2. \tag{3.6}$$

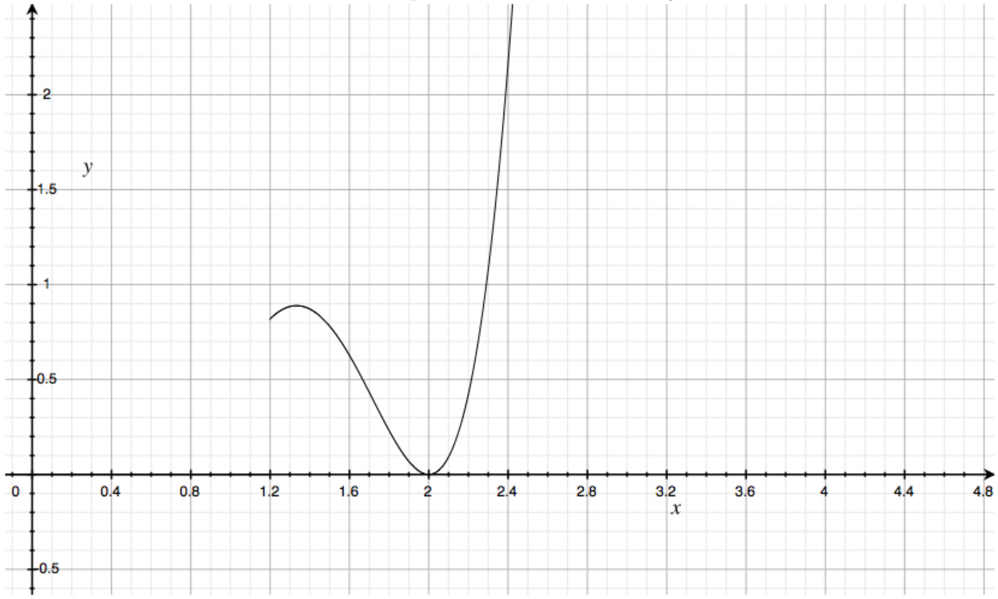
On the other hand, from Cauchy–Schwarz inequality we know that $n\sum_{i=1}^n B_i^2 \ge (\sum_{i=1}^n B_i)^2$. Furthermore, (3.6) holds only when $B_i = B_j$ for all $i, j \in \{1, ..., n\}$. Hence, $F(x, \theta) = 0$.

□

Lemma 3.1 implies that if constraints of (2.17) are not considered then first order optimality conditions determine the global optimal solution. On the other hand, when constrains are imposed local optima and local stationary points can occur.

The following simple example shows that a local stationary point can be caused by the equality constraint.

Figure 3.1: The function value in Example 3.2 with respect to x_1. Note that, in this 2×2 case, $x_2 = 1 - x_1$, and $\theta = \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2}{2} = \frac{\sigma_1^2 x_1^2 + \sigma_2^2 (1-x_1)^2}{2}$. Hence, the figure here shows $y = (x_1^2 - \frac{x_1^2 + 4(1-x_1)^2}{2})^2 + (4(1-x_1)^2 - \frac{x_1^2+4(1-x_1)^2}{2})^2$ when $x_1 \geq 1.2$. It shows that $x_1 = 1.2$ is a local optimum on the boundary.



Example 3.1. Consider an example of two assets with volatility $\sigma_1 = 1, \sigma_2 = 2$, respectively, and the covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}.$$

Suppose there are no bounds on x_1 and x_2 but we have $\sum_{i=1}^n x_i = 1$. Then it is easy to solve the system of equations that satisfies KKT conditions. There are three solutions, two of which, $x^1 = [\frac{2}{3}, \frac{1}{3}]$ and $x^2 = [2; -1]$ are risk parity solutions. The third stationary point is $x^3 = [\frac{4}{3}, -\frac{1}{3}]$, which is not a risk parity solution. However, this point is a local maximum. So far we have not seen an example where the equality constraint alone can cause a local minimum.

We now show that a local minimum can be caused by the bound constraints in (2.17), even if there is a risk parity solution (and thus, a global optimum) in the same orthant.

Example 3.2. Consider Example 3.1 but with additional bounds $1.2 \leq x_1 \leq \infty, -\infty \leq x_2 \leq -0.2$ (which enforces short sales of Asset 2). $x_1 = 2, x_2 = -1$ is the risk parity solution that satisfies the bounds. But as Figure 3.1 shows, $x_1 = 1.2, x_2 = -0.2$ is a local optimum which is not a global one.

4 Modified least-square models

In this section, we discuss several useful extensions of the risk parity problems that can be easily included in our least-square model.

Algorithm 1 Sequential min-variance risk parity algorithm

1. Choose $\rho^0 > 1$, $\beta \in (0,1)$ and x^0;
2. for $k = 0, 1, ...$
   (a) $x^{k+1} := \arg\min F(x)$, where $F(x)$ is defined as (4.1), given that x^k is a starting point.
   (b) If $\rho^k \le \epsilon$, then $x^{k+1} := \arg\min F(x)$ with $\rho^{k+1} = 0$, given that x^k is a starting point. STOP.
   Else $\rho^{k+1} := \rho^k \beta$.

4.1 Minimum variance with risk parity

As we discussed above, permissibility of short positions creates the possibility of finding multiple risk parity solutions. In such cases, the investors have the option to define additional criteria on their portfolio preferences to narrow down the choices for risk parity solutions, and possibly pick a "best" one. Introducing preferences about expected returns is one option, but we do not consider that here as we are focused on risk-based strategies. Instead, we focus on finding the risk-parity solution with the least variance. Hence we consider the following problem where the objective function is a weighted sum of total variance and least-squares risk parity term:

$$\begin{array}{ll} \min_{x,\theta} & \sum_{i=1}^{n} (x_i(\Sigma x)_i - \theta)^2 + \rho x^T \Sigma x \\ \text{s.t.} & a_i \le x_i \le b_i \\ & \sum_{i=1}^{n} x_i = 1, \end{array} \tag{4.1}$$

where $\rho \ge 0$ is the weight parameter. Note that, in the above we simply added a convex term to the objective function of (2.17). We propose an approach, described in Algorithm 1, of finding a risk parity solution with the smallest variance, where we simply solve a sequence of problems (4.1) with decreasing values of ρ.

As ρ grows towards infinity, problem (4.1) converges to the minimum variance portfolio problem. It is easy to show that for a large enough ρ problem (4.1) is convex in the feasible domain, if this domain is bounded. By setting initial ρ to a large value, we initiate Algorithm 1 with a potentially easy to solve problem and a solution that is close to the minimum variance solution but may be far from risk parity. Then, the algorithm solves a sequence of subproblems of the form (4.1) with decreasing values ρ, initializing each new subproblem with the solution of the previous subproblem. The goal is to converge to the risk parity solution that has the smallest variance among all risk parity solutions. Recall that each orthant contains at most one normalized risk parity solution. Algorithm 1 attempts to identify the correct orthant where the minimum variance risk parity solution lies. Hence, once ρ is small enough (for instance smaller than some tolerance, or once the convergence in terms of orthant is apparent), we can drop the minimum variance term and solve the problem in the right orthant, i.e., we have a risk parity problem and can obtain the exact risk parity solution that has small volatility.

Note that due to the nonconvexity of the objective function, there is no theoretical guarantee that Algorithm 1 will always converge to the minimum variance risk parity solution. However, in our experiments the performance of Algorithm 1 appears to be quite reliable. The numerical approach is much more efficient than an enumerative approach computing 2^{n-1} possible solutions.

Example 4.1. Often, the long-only portfolio is conservative in terms of risk, but it may not be the least risky one among all risk parity portfolios.

Consider three assets with volatility to be $\sigma_1 = 1, \sigma_2 = 1, \sigma_3 = 2$, respectively. Moreover, the correlation matrix is given by

$$Cor = \begin{bmatrix} 1.0 & -0.9 & 0.3 \\ -0.9 & 1.0 & -0.1 \\ 0.3 & -0.1 & 1.0 \end{bmatrix}.$$

Thus, the covariance matrix is

$$\Sigma = \begin{bmatrix} 1.0 & -0.9 & 0.6 \\ -0.9 & 1.0 & -0.2 \\ 0.6 & -0.2 & 4.0 \end{bmatrix}.$$

By setting ρ to be 1000, 10, 1, 0.01 and 10^{-6}, Algorithm 1 finds a risk parity solution $[0.574; 0.531; -0.105]$, after solving five subproblems. Since this is a small instance, it is not hard to empirically check that Algorithm 1 finds the risk parity solution that has the least volatility. There are 4 normalized risk parity solutions in total. We list these solutions in Table 4.1.

Table 4.1: *A comparison of strategies, with the lower and upper bounds to be $a = -1, b = 2$.*

Items for comparison	x_i	RC_i	Volatility
Risk parity portfolio (1)	[0.455;0.481;0.064]	[0.333;0.333;0.333]	0.289
Risk parity portfolio (2)	[-1.912;1.605;1.307]	[0.333;0.333;0.333]	3.840
Risk parity portfolio (3)	[1.784;-1.999;1.215]	[0.333;0.333;0.333]	4.805
Risk parity portfolio (4)	[0.574;0.531;-0.105]	[0.333;0.333;0.333]	0.238

In fact, our method is very robust with respect to the starting point. More results and larger instances are considered in numerical experiments in Section 6.

It is clear that similar strategy can be applied to select best risk parity portfolio based on another criterion, such as expected return or value at risk, etc.

4.2 Group risk parity

Another interesting extension of the risk parity problem is the case of group risk parity where we seek parity of risk contributions from groups of assets instead of individual assets. This variation is useful in the case when there are a large number of assets, for instance, in the context of equity investing. The assets can be grouped using a common risk factor such as industry membership or market capitalization and we look for equal risk contribution from each factor instead of each individual asset.

Another reason to apply group risk parity is to avoid fully dense solutions (which are enforced by individual risk parity) when the number of assets is large. This could be useful when investors would like to set an upper bound in the number of positions taken, or when transaction cost is of concern. To achieve that, one may consider adding a cardinality constraint in the group risk parity model. This approach brings additional computational difficulty and we do not discuss cardinality constraints here in the risk parity context. Interested readers can refer to [3] for more details.

For group risk parity, we solve the following nonconvex problem:

$$\begin{array}{ll} \min_{x,\theta} & \sum_{j=1}^{l}(\sum_{i\in\mathcal{G}_j} x_i(\Sigma x)_i - \theta)^2 \\ \text{s.t.} & a_i \le x_i \le b_i \\ & \sum_{i=1}^{n} x_i = 1, \end{array} \tag{4.2}$$

where $\mathcal{G}_j$ stands for the jth group, and l is the total number of groups. Here we make two assumptions: 1) we invest in all groups (i.e., we do not aim to pursue group sparsity); 2) there is no overlap (i.e., each asset can only lie in one of the groups). Risk parity between groups is achieved if the optimal value of the objective function is zero. In Section 6 we show how group risk parity may produce a desirable portfolio using sector membership for grouping different assets in an equity portfolio. In the next section we introduce an efficient algorithm that can handle problem (2.17), as well as the variations (4.1) and (4.2).

5 Algorithms solving second order least-square problems

In this section, we briefly introduce an algorithm for solving a class of second order least-square problems, namely Alternating Linearization Method (ALM). Details of this algorithm and others for this class of problems can be found in [1], including convergence analysis and additional numerical results. This algorithm was initially inspired by an alternating linearization algorithm in [7]. However, here it is applied to a nonconvex problem and in a substantially different setting.

Consider optimizing the following function

$$\min_{x\in\mathcal{X},\theta} F(x) = \sum_i ((A_i x)^T (B_i x) - \theta)^2, \tag{5.1}$$

where $x \in \mathbb{R}^n$, $A_i, B_i \in \mathbb{R}^{m\times n}$ and $\mathcal{X}$ is a set defined by linear constraints.

Note that our previous risk parity models can be embedded into this formulation. The objective function of risk parity problem, in the formulation of (5.1), has $A_i = \Sigma_i \in \mathbb{R}^{1\times n}$ as the ith row of the covariance matrix, and $B_i = e_i \in \mathbb{R}^{1\times n}$ is the ith column of the identity. In case of 4.2, $A_j \in \mathbb{R}^{m_j\times n}$ is defined by a submatrix of Σ which correspond to rows with indices from set $\mathcal{G}_j$. $B_j \in \mathbb{R}^{m_j\times n}$ is defined as follows: suppose the ith row in A_j is the corresponding (k_i)th row of Σ, then

$$(B_j)_{i,k} = \begin{cases} 1, & k = k_i \\ 0, & otherwise. \end{cases}$$

Note that (5.1) is equivalent to

$$\min_{x\in\mathcal{X},\theta} F(x,\theta) = \sum_{i=1}^{n} F_i(x) = \sum_{i=1}^{n} (x^T M_i x - \theta)^2, \tag{5.2}$$

where $M_i = A_i^T B_i \in \mathbb{R}^{n\times n}$. Clearly M_i is not generally symmetric or positive semidefinite. Hence we have a nonconvex function F in the form of (5.2). We consider a variable splitting approach which replaces $F(x,\theta)$ by $F(x,y,\theta) = \sum_{i=1}^{n}(x^T M_i y - \theta)^2,\ y = x$. For brevity, let us omit θ from the variables of F and use $F(x,y)$. Our method generates two sequences $\{x^k\}$ and $\{y^k\}$ in such a

Algorithm 2 Alternating linearization method (ALM)

1. Choose μ^0, and $x^0 = y^0$;
2. for $k = 0, 1, ...,$ do
   $x^{k+1} := \arg\min_x Q^1_{\mu^k_1}(x, y^k)$;
   $y^{k+1} := \arg\min_y Q^2_{\mu^k_2}(x^{k+1}, y)$;
   choose new penalty parameter $\mu^{k+1} \in (0, \mu^k)$.

way that $x^k \to x^*$ and/or $y^k \to x^*$ where x^* is a local optimal solution of (5.1).

Given y^k

$$F(x, y^k) \equiv \sum_{i=1}^{n} (x^T M_i y^k - \theta)^2, \tag{5.3}$$

and given x^k we have

$$F(x^k, y) \equiv \sum_{i=1}^{n} ((x^k)^T M_i y - \theta)^2, \tag{5.4}$$

Both $F(x, y^k)$ and $F(x^k, y)$ are convex functions of x and y, respectively, for any given y^k and x^k. Let $\nabla_i F$ denote the partial derivative of F with respect to either x ($i = 1$) or y ($i = 2$). In, particular, using the form of (5.2), we have

$$\begin{array}{l} \nabla_1 F(x, y) = \sum_{i=1}^{n} 2(x^T M_i y - \theta) M_i y \\ \nabla_2 F(x, y) = \sum_{i=1}^{n} 2(x^T M_i y - \theta) M_i^T x. \end{array} \tag{5.5}$$

We now construct the following two approximations of $F(x, y)$:

$$\begin{array}{l} Q^1_\mu(x, y^k) \triangleq F(x, y^k) + \left\langle \nabla_2 F(y^k, y^k), x - y^k \right\rangle + \frac{1}{2\mu} ||x - y^k||_2^2 \\ Q^2_\mu(x^{k+1}, y) \triangleq F(x^{k+1}, y) + \left\langle \nabla_1 F(x^{k+1}, x^{k+1}), y - x^{k+1} \right\rangle + \frac{1}{2\mu} ||x^{k+1} - y||_2^2, \end{array} \tag{5.6}$$

where μ is some choisen positive scalar. The following is the simple version of our ALM algorithm, shown in Algorithm 2.

In practice backtracking strategies should be applied to choose values of parameter μ at each iteration. A practical backtracking scheme is shown in Algorithm 3. Note that in each minimization step, we check whether a sufficient reduction has been obtained. If so, the minimization step is accepted and μ may be increased, otherwise is is decreased and a new candidate step is computed. Each minimization step in Algorithms 2 and 3 is a solution of a strictly convex quadratic programming problem, which can be done efficiently by many methods. We will discuss the implemented method in the next section.

Algorithm 3 Alternating linearization method with backtracking (ALM-bktr)

1. Choose μ^0, and $x^0 = y^0$;
2. for $k = 0, 1, ...,$ do
   (a) $x^{k+1} := \arg\min_x Q^1_{\mu_1^k}(x, y^k)$;
   (b) if $F(x^{k+1}) \le Q^1_{\mu_1^k}(x^{k+1}, x^k)$ then
   Choose $\mu_1^{k+1} \ge \mu_1^k$;
   else
   find the smallest n s.t. $\bar{\mu} := \mu_1^k \beta^n, \bar{x} := \arg\min_x Q^1_{\bar{\mu}}(x, y^k)$ and $F(\bar{x}) \le Q^1_{\bar{\mu}_1}(\bar{x}, y^k)$;
   $\mu_1^{k+1} := \bar{\mu}/\beta$, $x^{k+1} := \bar{x}$;
   (c) $y^{k+1} := \arg\min_y Q^2_{\mu_2^k}(x^{k+1}, y)$;
   (d) if $F(y^{k+1}) \le Q^2_{\mu_2^k}(x^{k+1}, y^{k+1})$ then
   Choose $\mu_2^{k+1} \ge \mu_2^k$;
   else
   find the smallest n s.t. $\bar{\mu} := \mu_2^k \beta^n, \bar{y} := \arg\min_y Q^2_{\bar{\mu}}(x^{k+1}, y)$ and $F(\bar{y}) \le Q^2_{\bar{\mu}_2}(x^{k+1}, \bar{y})$;
   $\mu_2^{k+1} := \bar{\mu}/\beta$, $y^{k+1} := \bar{y}$;

6 Numerical results

6.1 A comparison between strategies

We compare several asset allocation strategies on a small data set to illustrate the benefits of the risk parity strategy. Consider the following example with 5 assets and the covariance matrix of percentage returns (returns multiplied by 100) given by:

$$\Sigma = \begin{bmatrix} 94.868 & 33.750 & 12.325 & -1.178 & 8.778 \\ 33.750 & 445.642 & 98.955 & -7.901 & 84.954 \\ 12.325 & 98.955 & 117.265 & 0.503 & 45.184 \\ -1.178 & -7.901 & 0.503 & 5.460 & 1.057 \\ 8.778 & 84.954 & 45.184 & 1.057 & 34.126 \end{bmatrix}.$$

The covariance matrix above suggests that Asset 4 is a low risk asset while Asset 2 is a high risk asset. The other assets are in the middle of the risk spectrum. First, we consider the long-only risk parity case, when the lower and upper bound of the weights are set to be 0 and 1, respectively. As we have proved in Lemma 2.1, in this case, there is a unique risk parity solution.

Table 6.1: *A comparison of strategies, with the lower and upper bounds to be $a = 0, b = 1$.*

Items for comparison	x_i	RC_i	Volatility
1/n rule	[0.200;0.200;0.200;0.200;0.200]	[0.119;0.524;0.219;-0.002;0.139]	7.07%
Minimum variance portfolio	[0.050;0.006;0.000;0.862;0.082]	[0.050;0.006;0.000;0.862;0.082]	2.16%
Risk parity portfolio	[0.125;0.047;0.083;0.613;0.132]	[0.200;0.200;0.200;0.200;0.200]	3.04%

Table 6.1 shows a comparison of different strategies in terms of volatility and risk concentration. We compare three strategies: $1/n$ rule, minimum variance portfolio, and risk parity portfolio. The

risk contribution factor RC_i is calculated as

$$RC_i := \frac{x_i(\Sigma x)_i}{x^T \Sigma x},$$

which can be viewed as the percentage contribution for the ith asset to the total risk (variance of the portfolio). By design, the minimum variance strategy has the least volatility. However, the distribution of the risk contributions for the minimum variance portfolio is very skewed – more than 86% of the total risk comes from Asset 4. From Table 6.1 it can be observed that risk parity strategy also puts a high weight on the Asset 4, but is much more balanced in terms of risk contribution (equally weighted risk contribution). Meanwhile, the portfolio volatility of risk parity portfolio is between that of minimum variance portfolio and equally weighted portfolio, which indicates that risk parity could be viewed as a compromise between the other two strategies.

Risk parity portfolio does not always exist, as discussed in previous sections, if there are binding constraints on the asset weights. For instance, if we change the bounds to be $a = 0.05, b = 0.35$, there is no risk parity solution. Asset 4, given its much lower risk profile, requires a much higher weight than other assets to match their risk contributions. Since the upper bound on all the assets prevents this, risk parity cannot be achieved. Instead, we seek approximate parity by solving (2.17).

Let us call the optimal solution to (2.17) the "optimal parity" solution. From Table 6.2, we observe that the risk contribution of Asset 4 is lower than other assets and the resulting excess is shared roughly evenly among the remaining assets. No asset has risk contribution more than 30%.

Table 6.2: *A comparison of strategies, with the lower and upper bounds to be* $a = 0.05, b = 0.35$.

Items for comparison	x_i	RC_i	Volatility
1/n rule	[0.200;0.200;0.200;0.200;0.200]	[0.119;0.524;0.219;-0.002;0.139]	7.07%
Minimum variance portfolio	[0.200;0.050;0.050;0.350;0.350]	[0.280;0.178;0.086;0.034;0.421]	4.13%
Optimal parity portfolio	[0.204;0.060;0.130;0.350;0.256]	[0.256;0.198;0.234;0.027;0.284]	4.44%

Next, let us consider the long-short case. We know that our method finds a stationary point for nonconvex risk parity optimization problem (2.17). In particular, long-short case is interesting because there might be multiple solutions that attain risk parity. In Lemma 2.2, we showed that there could be as many as 2^{n-1} risk parity solutions when the bounds are loose enough. Table 6.3 shows that, provided different x^0, our alternating direction algorithm may converge to different solutions. We observe that all of them are in fact risk parity solutions and global optima for problem (2.17).

Given only the covariance information, risk averse investors can pick the risk parity solution with the lowest variance. As discussed in Section 4.1, we can apply Algorithm 1 to achieve this. In Table 6.3, it can be seen that the portfolio computed by Algorithm 1 (denoted as Risk parity portfolio* in Table 6.3) has the least volatility (and, in this case, it happens to be the long-only portfolio). Other risk parity solutions are generated by solving (2.17), with different random starting points. If investors choose to solve (2.13) and enumerate through all orthants, they are able to obtain $2^4 = 16$ solutions.

Table 6.3: *A comparison of strategies, with the lower and upper bounds to be $a = -1, b = 2$.*

Items for comparison	x_i	RC_i	Volatility
1/n rule	[0.200;0.200;0.200;0.200;0.200]	[0.119;0.524;0.219;-0.002;0.139]	7.07%
Minimum variance portfolio	[0.050;0.006;-0.012;0.856;0.100]	[0.050;0.006;-0.012;0.856;0.100]	2.16%
Risk parity portfolio *	[0.125;0.047;0.083;0.613;0.132]	[0.200;0.200;0.200;0.200;0.200]	3.04%
Risk parity portfolio (2)	[-0.223;0.074;0.125;0.820;0.204]	[0.200;0.200;0.200;0.200;0.200]	4.26%
Risk parity portfolio (3)	[0.154;0.073;-0.285;0.717;0.340]	[0.200;0.200;0.200;0.200;0.200]	3.48%
Risk parity portfolio (4)	[0.165;-0.118;-0.255;0.537;0.671]	[0.200;0.200;0.200;0.200;0.200]	3.38%

6.2 Strategic asset allocation

In this subsection, we backtest four different static investment strategies using real-world data.

We consider the following commonly used indices to represent different asset classes: S&P 500, MSCI World (Net), Russell 2500, Russell 2000 Growth, Russell 2000 Value, HFRI Equity Hedged Index, MSCI Emerging Markets (Net), HFRI Emerging MKTS Total, HFRI FoF (Conservative Index), BC Treasury 5-10 Yr, BC US Corporate High Yield Index, JPMorgan GBI-EM Index, JPMorgan EMBI+ Index, S&P Global Natural Resources - Energy Index. Here, we use a monthly sampling frequency, and the sampling period is from Nov. 2002 to Aug. 2012. The excess return and volatility are both annualized[1] and the risk-free rate is assumed to be the 3-month T-bill rate (the average is about 1.8% annually). For this test, we do not consider time varying covariance.

We compare the performance of 4 different strategies: 60/40 rule, $1/n$ rule, minimum variance portfolio, and risk parity portfolio. For each strategy, the weights and risk contribution of different asset classes are shown in Figure 6.1 and 6.2, respectively. We compare the excess return, volatility, Sharpe ratio, 5% nonparametric value-at-risk of these strategies. In addition, we compare two risk concentration metrics, namely the HRC and the Herfindahl index that are defined by (2.19) and (2.20), respectively.

It can be observed that for both of 60/40 rule and the equally weighted portfolio, the volatility and 5% VaR are high. For instance, for 1/n rule, there are 5% chances that we lose (more than) 5.80% of the money monthly. In terms of Sharpe ratio, risk parity portfolio is the best among all. For risk concentration, the traditional 60/40 rule is dominated by equity risk (more than 95%), as expected. As in the example of the previous section, the minimum variance portfolio, also has a high risk concentration: almost two third of the risk is contributed by the least risky asset, namely the HFRI FoF Conservative Index. This can be clearly observed from Figure 6.2, showing risk contribution of different assets. Overall, the risk parity portfolio provides a good compromise between balancing risk contributions and achieving reasonable amount of returns.

6.2.1 Asset level US equity portfolio

In this section, we consider US equity portfolio in the asset level. We porform a long-run simulation and compare different strategies. Similar experiments or simulated examples can be found in, for

[1] The return and covariance data can be found at the author's website: http://phd.ie.lehigh.edu/~xib210/

Figure 6.1: Weights of different asset classes in the asset allocation example (14×14)

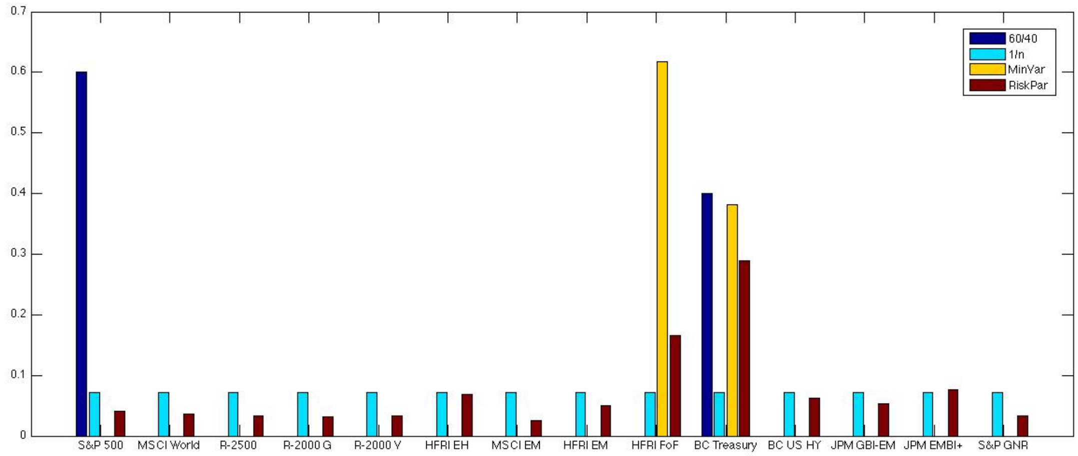


Figure 6.2: Risk contribution of different asset classes in the asset allocation example (14×14)

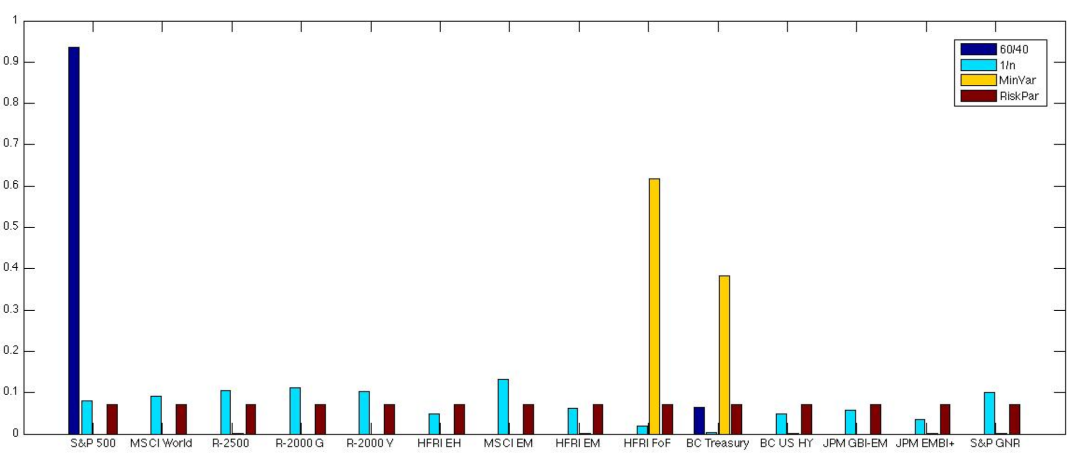


Table 6.4: *A comparison of strategies, with the lower and upper bounds to be $a = 0, b = 1$.*

Items for comparison	Excess return	Volatility	Sharpe Ratio	VaR 5%	Highest RC	Herfindahl index
60/40 rule	4.33%	9.36%	0.462	4.43%	95.49%	0.8783
1/n rule	8.59%	12.39%	0.693	5.80%	13.12%	0.0900
Minimum variance portfolio	2.59%	3.55%	0.730	1.17%	61.79%	0.5278
Risk parity portfolio	6.57%	7.72%	0.851	3.13%	7.14%	0.0714

instance, [5, 9].

Consider the equity universe with 17 industry sectors (food, mines, oil, clths, durbl, chems cnsum, cnstr, steel, fabpr, machn, cars, trans, utils, rtail finan, other) for the US market. The data is obtained from Kenneth French's data library[2].

We run the simulation based on the monthly return from Oct. 1972 to Sep. 2012. Since the sampling period is relatively long, we cannot assume the second moment parameters to be constant over time and we take into consideration the time varying of the risk parameters. For each year, we estimate the covariance matrix based on the returns of the previous j years (we thus do not account for the results in the beginning j years). Here we apply the rolling window to be 36, 60 or 120 months. The risk free rate is assumed to be the 3-month T-bill rate. Meanwhile, considering the rebalance cost, we apply annual rebalancing. The average annualized excess return, volatility, Sharpe ratio and risk contribution are reported. We also plot the excess return and cumulative excess return over time (Figure 6.3 - 6.5).

It can be observed from Table 6.5, 6.6 and 6.7 that minimum variance portfolio has less volatility (and sometimes enjoys a slightly higher Sharpe ratio) than the other two, while its risk concentration is much higher as well. As a comparison, very often the volatility of risk parity portfolio lies between the other two, which once again shows that risk parity is a compromise between equally weighted and minimum variance portfolio. Moreover, risk parity portfolio has higher returns than minimum variance portfolio, which can be clearly seen from Figure 6.3, 6.4 and 6.5.

Table 6.5: *A comparison of strategies, on asset level US equity portfolio (3 years rolling window)*

Items for comparison	Excess return	Volatility	Sharpe Ratio	Highest RC	Herfindahl index
1/n rule	8.04%	15.30%	0.525	9.76%	0.0674
Minimum variance portfolio	6.39%	11.67%	0.547	51.79%	0.4089
Risk parity portfolio	8.59%	13.77%	0.624	14.72%	0.2034

Table 6.6: *A comparison of strategies, on asset level US equity portfolio (5 years rolling window)*

Items for comparison	Excess return	Volatility	Sharpe Ratio	Highest RC	Herfindahl index
1/n rule	7.97%	15.46%	0.516	9.83%	0.0676
Minimum variance portfolio	6.65%	11.66%	0.570	49.67%	0.3738
Risk parity portfolio	7.08%	14.11%	0.507	14.11%	0.1329

[2]For more details, please go to http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Table 6.7: *A comparison of strategies, on asset level US equity portfolio (10 years rolling window)*

Items for comparison	Excess return	Volatility	Sharpe Ratio	Highest RC	Herfindahl index
1/n rule	8.90%	15.15%	0.588	9.98%	0.0684
Minimum variance portfolio	7.59%	11.39%	0.667	54.90%	0.4072
Risk parity portfolio	9.57%	14.54%	0.658	16.11%	0.1542

6.3 Group risk parity portfolios based on S&P 500

In this section, we study the investment strategy based on historical stock prices in S&P 500, from Aug. 21, 2009 to Aug. 20, 2010 (see: http://pages.swcp.com/stocks/). We ignore those stocks with less than 200 trading days of data. In total there are 482 stocks and 245 trading days considered. We now compare the simple "1/n" and minimum variance strategy with group risk parity.

The groups are determined by Global Industry Classification Standard (GICS Sector) and are found at the following site: http://en.wikipedia.org/wiki/List_of_S%26P_500_companies. Here, by GICS, we have ten sectors: Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Telecommunication Services and Utilities. Note that there is no overlap between groups, which means there exists one and only one group that each stock belongs to.

Since the sample size is small within a short period of time (only one year of data), we ignore the clustering with respect to the time series when testing the performance of the strategies. Here we have 244 daily returns in the sample data. Since, very often, we use much more data (5-10 times) in training the parameters than that is used to test our model, we divide the data into 2 groups, with the first 220 returns in the training set and the rest 24 as validation/testing data. The risk-free rate is assumed to be the 3-month t-bill rate.

The results are shown in Table 6.8. The highest group risk contribution is defined as $HGRC := \max_j \frac{\sum_{i\in\mathcal{G}_j} x_i(\Sigma x)_i}{x^T\Sigma x}$. Also, the group Herfindahl index is defined as

$$h_{\mathcal{G}} = \sum_{j=1}^{p} \left[\frac{\sum_{i\in\mathcal{G}_j} x_i(\Sigma x)_i}{x^T\Sigma x} \right]^2 .$$

Since we are testing on the validation data, we cannot expect to achieve perfect risk parity, but it can be seen from Table 6.8 that the group Herfindahl index is not far from the perfect one (which is $1/10 = 0.1$ since we have 10 groups). Also, the group risk parity portfolio enjoys a higher Sharpe ratio than both 1/n rule and minimum variance portfolio, which indicates it might be a good compromise of the other two strategies. As a comparison, 1/n rule has less excess return and higher volatility, and thus is dominated. Moreover, minimum variance portfolio, while having the least volatility, suffers from risk concentration both on stock level and group level. Its highest group risk contribution is 48.34%: almost half of total risk lies in the group "Consumer Staples"

Figure 6.3: Excess returns and cumulative excess returns of three strategies on US equity market over time (estimation window with a length of 36 months)

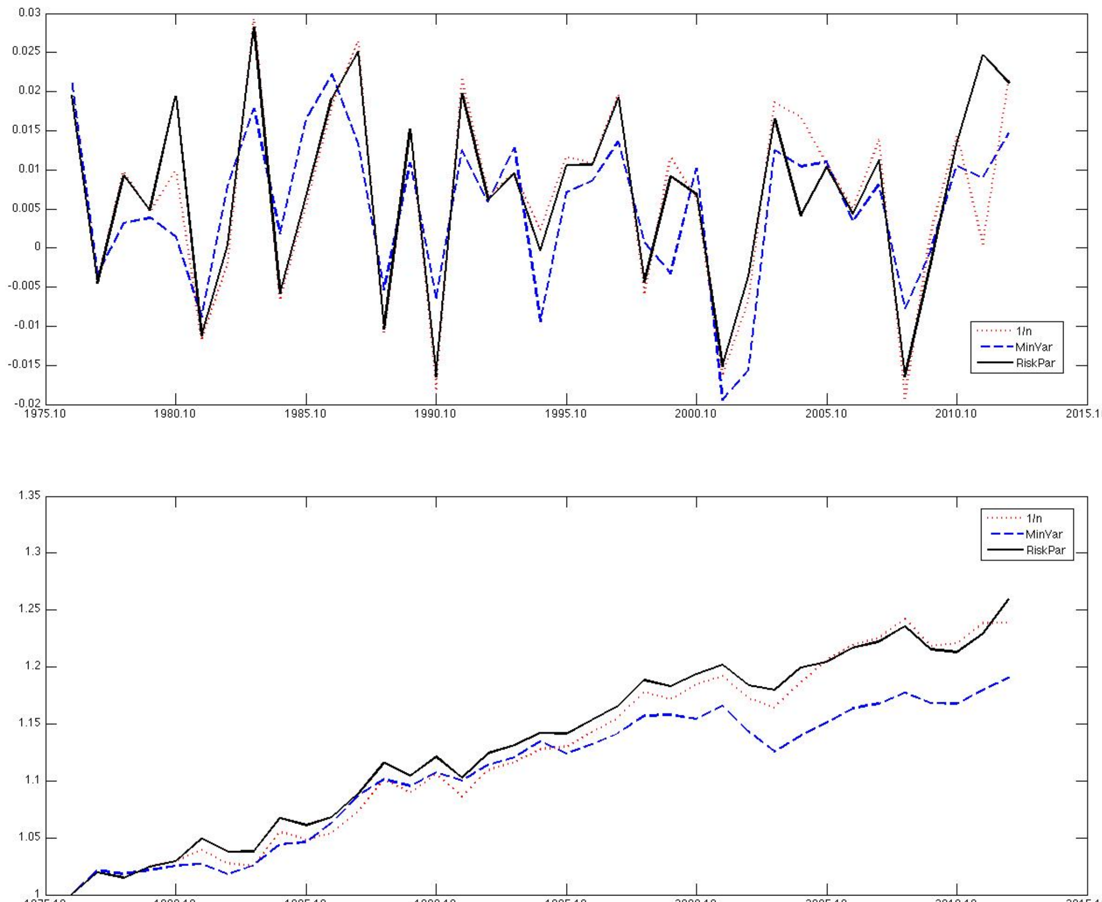
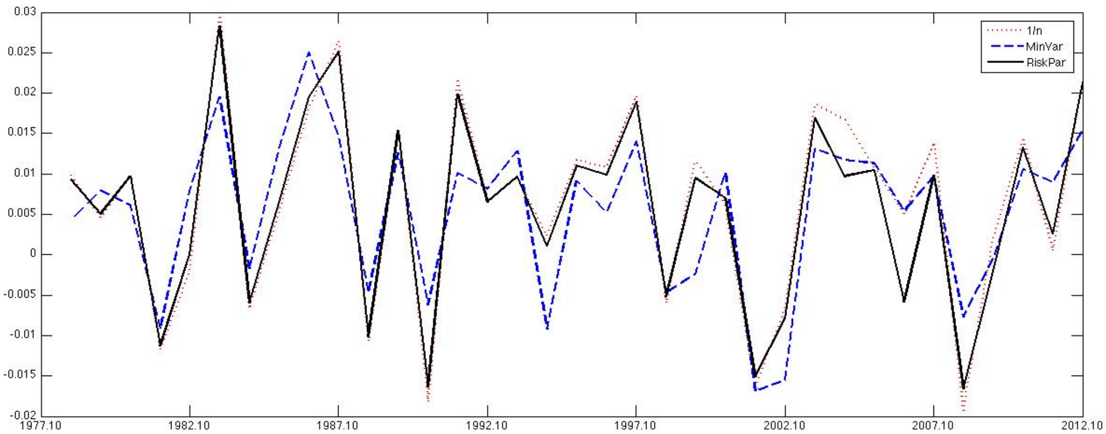


Figure 6.4: Excess returns and cumulative excess returns of three strategies on US equity market over time (estimation window with a length of 60 months)



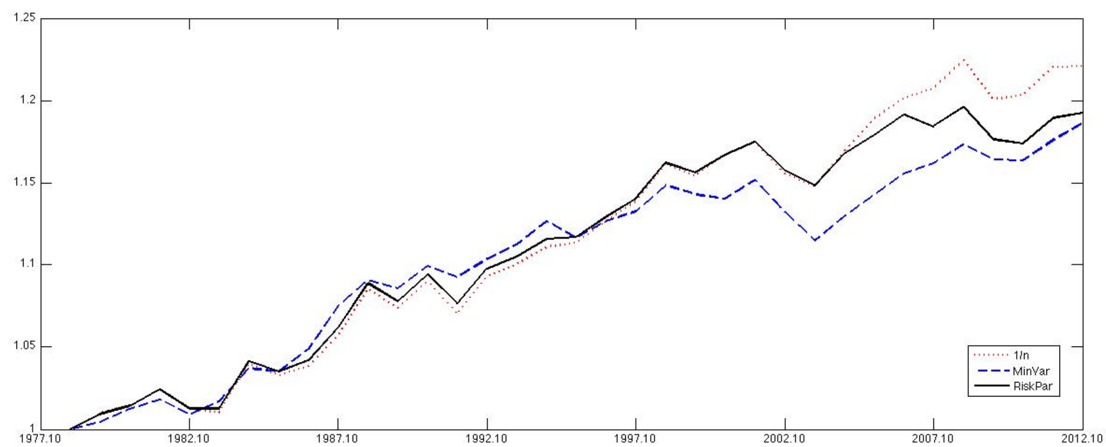


Figure 6.5: Excess returns and cumulative excess returns of three strategies on US equity market over time (estimation window with a length of 120 months)

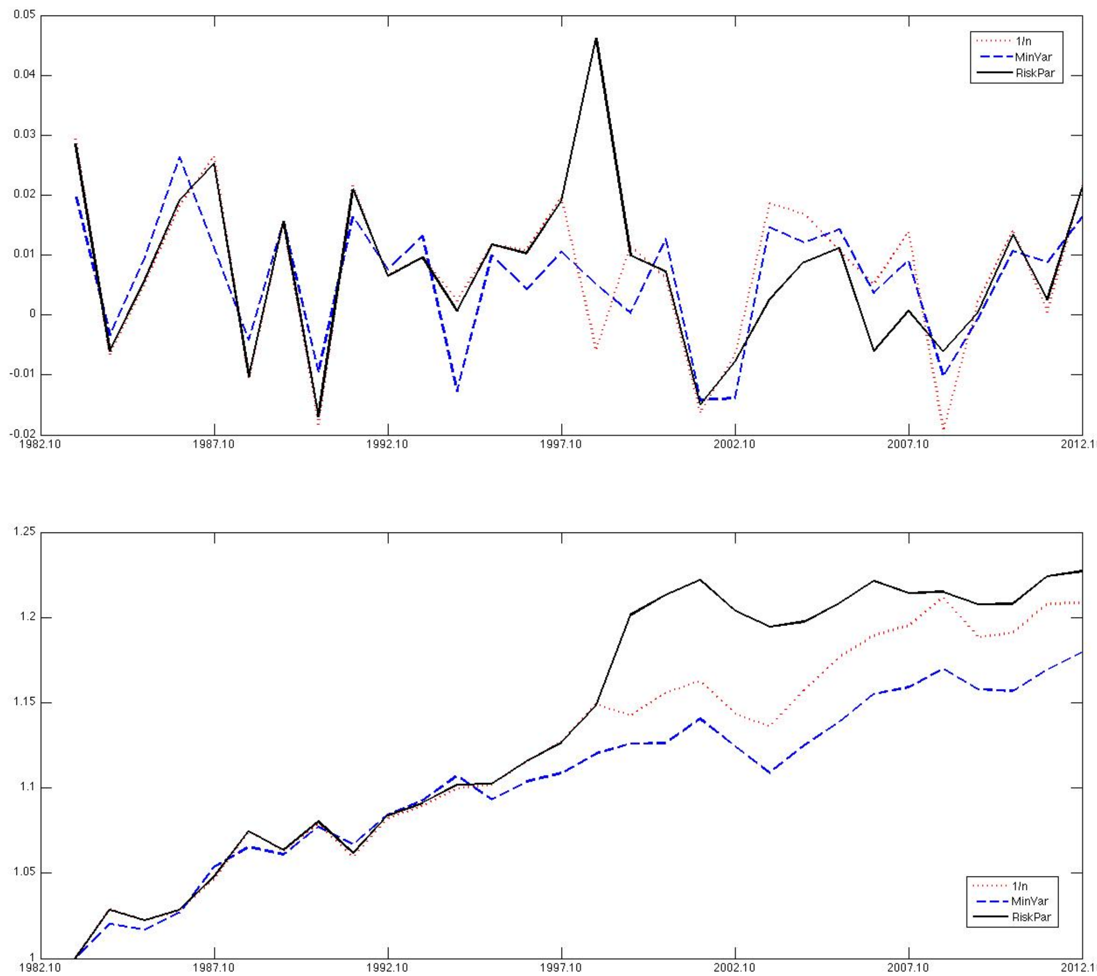
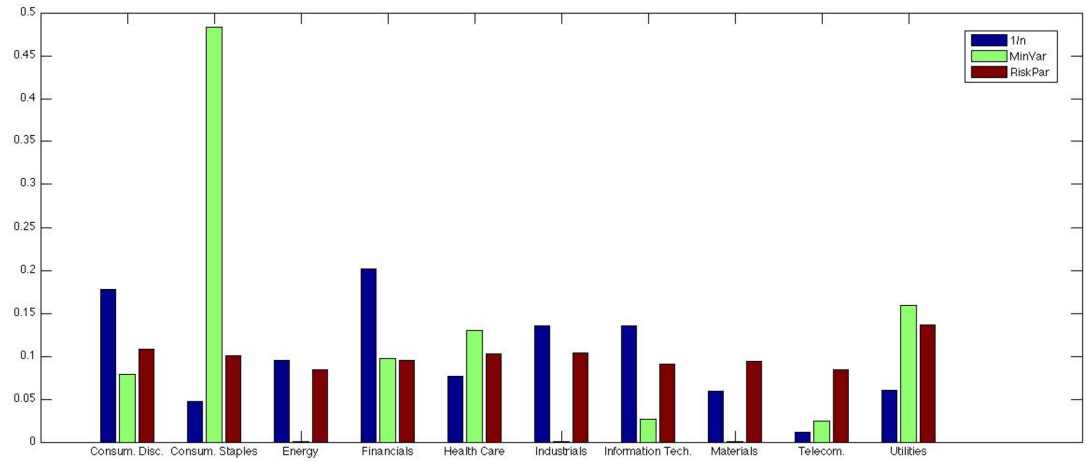


Figure 6.6: Risk contribution of different groups based on testing data



(this can be clearly seen in Figure 6.6). Furthermore, there are three stocks (out of 482), each of which has a risk contribution more than 10 percent (Johnson & Johnson - 10.47%, Southern Co. - 15.93%, and Wal-Mart Stores - 14.26%). Recall that we have hundreds of stocks in the investment pool: that is quite high risk concerntration.

Table 6.8: *A comparison of strategies, with the lower and upper bounds to be $a = 0, b = 1$.*

Items	Excess return	Volatility	Sharpe Ratio	Highest RC	Highest Group RC	Group Herfindahl index
1/n rule	12.42%	20.84%	0.596	1.62%	20.20%	0.133
Min. var.	7.70%	11.89%	0.648	15.93%	48.34%	0.293
Group risk par.	18.69%	19.09%	0.979	3.96%	13.60%	0.102

6.4 Efficiency of algorithms

In this section, we briefly discuss the efficiency of the algorithms shown in Section 5, in application to the instances discussed so far in this section. This paper is not aimed at algoritmic details, more discussion on algorithms and corresponding numerical experiments can be found in [1].

Our implementation is written in MATLAB and experiments performed in MATLAB R2010b on a laptop with Intel Core Duo 1.8 GHz CPU and 2GB RAM. We apply Mosek 6.0 to solve the QP subproblems in Algorithm 3, when optimizing Q_1 and Q_2.

In Table 6.9, we show the efficiency of the algorithms on different data sets. An arbitrary symmetric positive semidefinite matrix can be generated by $\Sigma = AA^T$, where A_{ij} is uniformly distributed within the interval $[0, 1]$. Recall that the main computational cost of each algorithm lies in the number of quadratic models it solves as the subproblem. Hence, in each table we report the number of iterations, the total number of QP solved. To show that our algorithms can find a stationary point, we compared the performance of the algorithms by recording the number of iterations they took when largest KKT violation (computed in absolute value) is below some thresholding ϵ's. We terminate the algorithm when the largest KKT violation is below 10^{-6}. It can be seen that, after the largest KKT violation is less than 10^{-6}, very often the function value is less than 10^{-14}, which is close to machine epsilon in MATLAB. Thus, since these are "long-only" cases, we can say that all algorithms can indeed find a global optimum with the objective function value to be zero.

Table 6.9: *A comparison of algorithm ALMs with backtracking on different instances. The starting point is chosen to be equally weighted portfolio, i.e., $x_i^0 = 1/n$. Due to testing data's scaling, we allow some parameter tuning for a large starting μ.*

Instance (size)	Starting μ	Bounds	k ($\epsilon = 10^{-4}$)	QP	F-value	k ($\epsilon = 10^{-6}$)	QP	F-value
Random I (20)	0.01	$a = 0; b = 1$	6	14	1.01×10^{-11}	7	17	8.86×10^{-15}
Random II (200)	0.0001	$a = 0; b = 1$	5	12	1.59×10^{-12}	7	18	6.58×10^{-18}
5 assets example (5)	1000	$a = 0; b = 1$	1	6	1.21×10^{-6}	11	16	1.01×10^{-9}
5 assets example (5)	1000	$a = 0.05; b = 0.35$	1	6	1.21×10^{-6}	6	20	1.63×10^{-7}
5 assets example (5)	1000	$a = -1; b = 2$	1	6	1.21×10^{-6}	11	16	1.01×10^{-9}
Asset allocation (14)	10000	$a = 0; b = 1$	1	2	2.68×10^{-8}	6	12	3.89×10^{-9}
US equity (482)	1000	$a = 0; b = 1$	1	2	1.11×10^{-11}	2	4	1.10×10^{-11}

Results showing lower and upper bounds to be 0 and 1, respectively, can be seen in Table 6.9. Moreover, Table 6.9 shows the case when much tighter bounds are considered. Recall that there might be no zero-parity solution and when that is the case, the global optimum does not lie in the interior of the box constraints.

Investors might use MATLAB optimization toolbox to solve (2.17), but may not obtain a satisfying result, especially when the number of assets becomes large. This may be due to the scale of the data or to the smooth feature of the function, which leads to a small function decrease per iteration. For instance, for the above 14×14 strategic asset allocation example, we can solve (2.17) using MATLAB *fmincon* with selected algorithms. We set the function tolerance (which relates to both the size of the latest change in the objective function value and the first-order optimality measure) as 10^{-8} and the maximum function evaluation number as 10000. Table 6.10 compares our ALM with the results MATLAB *fmincon* obtained. We can observe that, in this example, *fmincon*-SQP completely fails, as it does not goes far from the starting point. *fmincon*-interior point seems to perform better but still not as good as our algorithm (it takes longer to find a far worse solution than ALM). For ALM, given the tolerance, it finds a solution which is close to optimum by taking only 12 iterations and by solving 30 QPs. Again, such result shows the efficiency of the algorithm we proposed.

7 Conclusion and future work

In this paper, we discuss the problem of finding portfolios that satisfy risk parity as of either individual assets or groups of assets as closely as possible. We analyze the limitations convex

Table 6.10: *A comparison of algorithms strategic asset allocation instance* (14 × 14) *in Section 6.2. The starting point is chosen to be equally weighted portfolio, i.e.,* $x_i^0 = 1/n$.

Algorithm	ALM	*fmincon* - SQP	*fmincon* - interior point
Function value	1.69×10^{-11}	2.89×10^{-8}	4.99×10^{-9}
$x_1 \sim x_7$	[0.042 0.037 0.034 0.033 0.035 0.069 0.027]	[0.071 0.071 0.071 0.071 0.071 0.071 0.071]	[0.055 0.052 0.048 0.043 0.048 0.091 0.038]
$x_8 \sim x_{14}$	[0.052 0.159 0.284 0.064 0.054 0.077 0.034]	[0.071 0.071 0.071 0.071 0.071 0.071 0.071]	[0.068 0.119 0.113 0.089 0.075 0.109 0.050]
$RC_1 \sim RC_7$	[0.072 0.072 0.073 0.073 0.073 0.071 0.075]	[0.081 0.092 0.104 0.111 0.104 0.048 0.131]	[0.075 0.081 0.084 0.080 0.083 0.075 0.086]
$RC_8 \sim RC_{14}$	[0.071 0.067 0.067 0.071 0.071 0.071 0.073]	[0.063 0.019 0.004 0.049 0.057 0.036 0.101]	[0.075 0.039 0.013 0.076 0.076 0.072 0.086]
Succeed/Fail	Succeed	Fail	Partially succeed
CPU time (s)	0.026	-	5.061

optimization approach which was proposed in prior literature. We then propose an alternative nonconvex least-square model whose set of optimal solutions includes all risk parity solution, and propose a modified formulation which aims at selecting the most desirable risk parity solution (according to some criteria). Our model has many advantages, especially when general bounds are considered or when other constraints are considered to be added. Furthermore, we propose an alternating linearization framework to solve this nonconvex model. Numerical experiments indicate the effectiveness of our technique. Faster and more efficient algorithms remain a topic of future research.

References

[1] Xi Bai and Katya Scheinberg. Minimizing sum of squares with products of simple functions: a linearization approach. *Technical report available at phd.ie.lehigh.edu/ xib210/*, pages 1–15, September 2013.

[2] Michael Best and Robert Grauer. 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, April 1991.

[3] Pedro Brito and Luis Vicente. Efficient Cardinality / Mean-Variance Portfolios. *preprint available at http://www.optimization-online.org/DB_HTML/2012/03/3381.html*, pages 1–27, 2012.

[4] Denis Chaves, Jason Hsu, Feifei Li, and Omid Shakernia. Risk Parity Portfolio vs. Other Asset Allocation Heuristic Portfolios. *Journal of Investing*, 20(1):108–118, 2011.

[5] Denis Chaves, Jason Hsu, Feifei Li, and Omid Shakernia. Efficient Algorithms for Computing Risk Parity Portfolio Weights. *Journal of Investing*, 21(3-fall):150–163, 2012.

[6] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *Review of Financial Studies*, 22(5):1915–1953, December 2009.

[7] Donald Goldfarb, Shiqian Ma, and Katya Scheinberg. Fast alternating linearization methods for minimizing the sum of two convex functions. *Mathematical Programming*, pages 1–34, 2012.

[8] Hakan Kaya and Wai Lee. Demystifying Risk Parity. *Neuberger Berman*, March 2012.

[9] Sebastien Maillard, Thierry Roncalli, and Jerome Teiletche. On the properties of equally-weighted risk contribution portfolios. *Journal of Portfolio Management*, 36(4):60–70, 2010.

[10] Florin Spinu. An Algorithm for Computing Risk Parity Weights. *Available at SSRN: http://ssrn.com/abstract=2297383*, July 2013.

[11] Marc Steinbach. Markowitz Revisited: Mean-Variance Models in Financial Portfolio Analysis. *SIAM Review*, 43(1):31–85, January 2001.