

Journal Pre-proof

General Sparse Risk Parity Portfolio Design via Successive Convex Optimization

Linlong Wu, Yiyong Feng, Daniel P. Palomar

PII: S0165-1684(19)30485-2
DOI: <https://doi.org/10.1016/j.sigpro.2019.107433>
Reference: SIGPRO 107433



To appear in: *Signal Processing*

Received date: 29 October 2018
Revised date: 1 August 2019
Accepted date: 13 December 2019

Please cite this article as: Linlong Wu, Yiyong Feng, Daniel P. Palomar, General Sparse Risk Parity Portfolio Design via Successive Convex Optimization, *Signal Processing* (2019), doi: <https://doi.org/10.1016/j.sigpro.2019.107433>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier B.V.

HIGHLIGHTS

- We conducted new simulations on S&P 500. From this new simulation, we can see that the portfolio behaves as we expected as it achieves a good tradeoff among different metrics of interest. However, since the initial set of assets changes and we have no preprocessing on the large scale data, the performance will deteriorate to some degree as we expect.
- We explained some technical points in terms of the proposed algorithm, especially in terms of the efficiency.
- We revised some illustrations of the focus of this paper to make our point more clear. This paper proposed a sparse risk parity portfolio and the corresponding fast solving numerical algorithm. The algorithm can design the portfolio weights well with the support of read data simulation.

General Sparse Risk Parity Portfolio Design via Successive Convex Optimization

Linlong Wu, Yiyong Feng, and Daniel P. Palomar

Abstract—Since the 2008 financial crisis, risk management has become more important and portfolio approaches, such as the minimum-variance and equally weighted portfolios, have gained popularity. However, such portfolios still do not diversify the risk in the true sense. Recently, risk parity portfolios have been receiving significant interest from both the theoretical and practical perspectives due to its advantages in the diversification of (ex-ante) risk contributions among assets. However, this portfolio type usually results in nonzero weights in all the assets, which implies high transaction cost in practice. In addition, focusing only on the risk aspect can make this type of portfolio unsatisfactory if other performance factors, e.g., annual yield, are considered. In this paper, we jointly consider asset selection and risk diversification via imposing sparsity and risk parity regularization in the portfolio problem formulation, which turns out to be a general and flexible portfolio framework. Then we propose an efficient sequential algorithm based on the successive convex optimization framework. The numerical results on historical data show that our portfolio approach, compared with benchmark portfolios, can achieve a good balance among asset selection, risk diversification and other evaluation criteria, and achieves the best performance on profit and loss (P&L) and/or drawdown.

Index Terms—portfolio selection, risk diversification, sparsity, risk parity, successive convex optimization

I. INTRODUCTION

Although nearly ten years have passed since the 2007-2008 financial crisis, it is still casting a shadow and it has changed the world, especially the financial industry. The S&P 500 index during the financial crisis is presented in Figure 1. It shows the dramatic fluctuation and general recession of the equity market at that time. A predictable change brought by this crisis is that risk management has become more important, and is sometimes considered more significantly than performance management.

Both practitioners and researchers started more seriously to investigate less aggressive active portfolios in order to avoid possible catastrophic losses in a declining or volatile market. Two portfolio approaches, namely, the minimum-variance and equally weighted (EW) approaches, are widely used. A minimum-variance portfolio aims to minimize the portfolio variance and belongs to the mean-variance (MV) portfolio framework, which was pioneered by Henry Markowitz in 1959 [1]–[3]. Thus, minimum-variance portfolios inherit the drawback of the mean-variance type of portfolio, namely, its excessive sensitivity to the estimates of the covariance matrix.

Linlong Wu and Daniel P. Palomar are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST), Hong Kong (e-mail: {lwuag, palomar}@ust.hk). Yiyong Feng is with Three Stones Capital Limited (email: yiyong.feng@gmail.com). Yiyong Feng's work to this paper was done when he was with the HKUST.

This approach also tends to highly concentrated portfolios and spreads risk over only a few assets [4], which goes against the common sense approach of diversification as a way to diversify risk. Such concentrated portfolios could incur catastrophic losses in the case of extreme events, such as a financial crisis. For instance, equities registered negative returns of about -50% in the 2007-2008 financial crisis, with the consequent terrible performance of most hedge funds. Therefore, a portfolio that diversifies capital via minimizing the mean-variance optimization does not necessarily diversify risk [5].

In an EW portfolio, the capital is equally diversified among all available assets [6], [7]. It has been found by some empirical studies that no one portfolio based on optimization consistently outperforms the heuristic and simple EW portfolio approach in terms of the out-of-sample Sharpe ratio, return, or turnover [8]. The probable reason is that portfolios designed via optimization approaches are undermined by estimation error, which can be partially overcome by either shrinkage estimation [9], [10] or adding constraints on the portfolio weights [6], [7], [11]–[13]. Despite the good out-of-sample performance of EW portfolios, their risk diversification is dealt with in a very naive and indirect way.

A new paradigm, called a risk parity approach, was introduced to make portfolios, and hence their risk, truly more diversified. Around 2005, Qian [14], [15] first showed that uniform risk contributions actually lead to a diverse enough portfolio, and the (ex-ante) risk contributions (i.e., the risks computed using historical data) are not only a mathematical measurement of how diverse the risk is, but also good indicators of the (ex-post) loss contributions of the assets (i.e., the observed risks and losses in the future), especially when there exist large losses. According to this observation, a promising way to avoid a potential huge loss is to distribute the risk contributions over the selected assets. The typical risk parity portfolio is an equal risk contribution (ERC) portfolio, which makes all the available assets contribute equally to the total risk of the portfolio.

Risk parity portfolios, however, did not attract too much attention before the 2008 financial crisis, until Maillard et al. [16] first analyzed the properties of ERC portfolios and showed that they are a tradeoff between the MV and EW types of portfolios. Following that, works on different formulations [17]–[21] or numerical methods of computing the risk parity portfolios [22]–[25] were presented. Meanwhile, the risk parity approach found its way into various applications, e.g., risk-based indexation [26], [27], alternative assets management [19], equity portfolio optimization [28], [29], portfolio re-

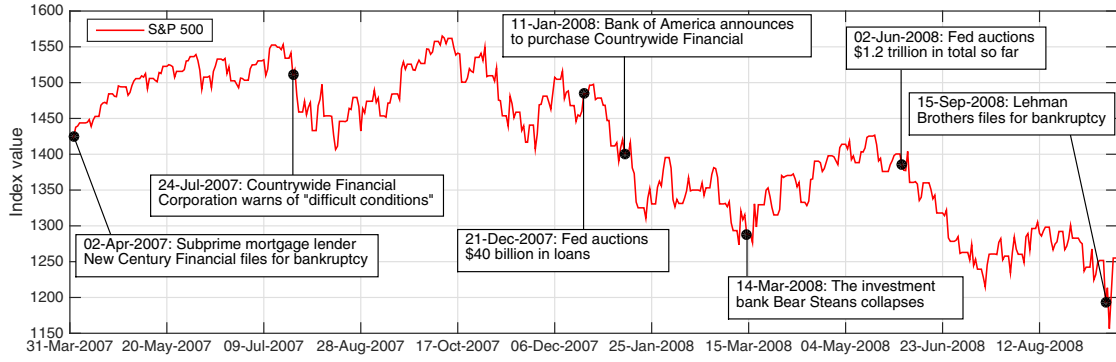


Figure 1. The S&P 500 index during the 2007-2008 financial crisis.

balancing [30], portfolio with multi-asset classes [19], [31]–[34], etc. The book [5] serves as a good summary of both the theoretical foundations and applications of risk parity portfolios.

Although the risk parity approach properly conducts risk diversification, some drawbacks limit its practical applications. One issue in portfolio design is the transaction cost, which in general is assumed to be concave [35]. For example, Lobo et al. [36] considered a constant plus linear cost function. However, the risk parity approach always results in a portfolio with nonzero weights in all the assets [5], [16], which implies a considerable transaction cost and may attenuate the portfolio performance significantly. Recall that MV portfolios tend to be highly concentrated, which at the same time can be interpreted as asset selection, no matter whether the selection is good or not. Thus, to reduce the transaction cost, it is beneficial to invest only in part of all the assets with a proper capital allocation. Another important issue is that the risk parity approach purely focuses on the risk without consideration of other investment goals. For example, if a portfolio aims at tracking a market index, designing a risk parity portfolio completely departs from this initial goal, although it will diversify the risk of the portfolio.

Given these drawbacks of risk parity portfolios and motivated by other existing portfolios, in this paper, we jointly conduct asset selection and risk diversification while at the same time taking other investment goals into consideration. An efficient algorithm based on successive convex optimization techniques is proposed. Detailed experiments on historical stock data are conducted with corresponding analysis. The major contributions of this paper are summarized as follows:

- We jointly consider risk parity-like diversification and asset selection in portfolio design. Specifically, the ℓ_0 -norm¹ is considered as a regularization imposing sparsity for the asset selection. The risk diversification is formulated as a penalty term based on the concepts of risk contribution and risk parity. To the best of our knowledge, the combination of risk parity-like diversification and

sparse asset selection in designing a portfolio has never before been investigated in the open literature.

- Our problem formulation is general and flexible. It can be applied to different investment scenarios, and many existing portfolios can be included in this formulation as special cases.
- For our general nonconvex portfolio problem, we approximate the original problem and then develop an efficient algorithm based on successive convex optimization techniques. The algorithm can deal with a large class of portfolio constraints and goals. Under some mild conditions, the proposed algorithms can also guarantee the convergence.
- Detailed and complete numerical experiments on historical stock data are conducted with corresponding analyses, and they reveal the unique advantages of our proposed portfolio approach. In particular, two investment scenarios, mean-variance optimization and index tracking, are considered in the numerical experiments. Many performance criteria and benchmark approaches (some existing mainstream portfolios) are taken into consideration.

The remainder of this paper is organized as follows. Section II briefly reviews the background of risk parity portfolios, which serves as the preliminaries for our problem formulation. Section III formulates the general portfolio framework and illustrates its flexibility and generality. The proposed efficient algorithm and the experimental results are presented in Section IV and Section V, respectively. Finally, we will conclude this paper in Section VI.

A conference version of this work has appeared at ICASSP 2016, Shanghai, Chian, Mar. 2016 [37]. This journal version includes detailed illustration of problem formulation, applications on two practical investment scenarios, comprehensive simulations, and a set of experiments using real financial data.

Notation: \mathbb{R}^n and $\mathbb{R}^{n \times n}$ denote the real $n \times 1$ vector and $n \times n$ matrix space, respectively. Boldface upper-case and lower-case letters stand for matrices and column vectors, respectively. Standard lower case letters stand for scalars. The notation $\mathbf{1}$ denotes a column vector with proper size and all elements being 1. The vector \mathbf{e}_i denotes the column vector with only the i -th element being one and zero elsewhere. The

¹The ℓ_0 -norm measures the number of nonzero elements of a vector. By using the terminology “norm” here, we are abusing notation.

transpose and inverse operators are denoted by the symbols $(\cdot)^T$ and $(\cdot)^{-1}$ respectively. $\text{Diag}(\mathbf{a})$ denotes a diagonal matrix with diagonal elements equal to those of \mathbf{a} and zero elsewhere.

II. BACKGROUND OF RISK PARITY PORTFOLIOS

In this section, we briefly introduce the background of risk parity portfolios, which will serve as the preliminaries for our problem formulation. We first review the risk contribution, which is used as the measurement of the risk contributed by each asset. Then we introduce the typical risk parity portfolio, namely, an equal risk contribution (ERC) portfolio.

A. Risk Contribution

Suppose there are n assets with random returns $\mathbf{r} \in \mathbb{R}^n$, and the mean vector and (positive definite) covariance matrix are denoted as $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$, respectively. We use $\mathbf{w} \in \mathbb{R}^n$ to denote the normalized portfolio (e.g., $\mathbf{w}^T \mathbf{1} = 1$) and it describes how the total capital budget is to be allocated over the assets. To study the risk parity approach, we need some well defined risk measurements $f(\mathbf{w})$ so that the “risk contribution” of each asset to the risk of the whole portfolio can be quantified. The following desired property is key in the risk parity literature [5].

Theorem 1 (Euler’s homogeneous function theorem). *Suppose that the function $f: \mathbb{R}^n \setminus \{0\} \rightarrow \mathbb{R}$ is continuously differentiable. Then, f is positively homogeneous function of degree one² if and only if*

$$f(\mathbf{w}) = \sum_{i=1}^n w_i \frac{\partial f}{\partial w_i}. \quad (1)$$

One observation from property (1) is that the component $w_i \frac{\partial f}{\partial w_i}$ can be regarded as the risk contribution from asset i to the total risk $f(\mathbf{w})$.

Interestingly and fortunately, most of the existing risk measurements do satisfy the Euler property (1) either directly (VaR and CVaR) [20], [21] or indirectly (variance) [16]. For example, note that the variance $\sigma^2(\mathbf{w}) = \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}$ does not satisfy (1) directly. Fortunately, it is easy to check that the volatility, i.e., the positive square root of the variance, $\sigma(\mathbf{w}) = \sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}$ does satisfy (1) as follows:

$$\begin{aligned} \sum_{i=1}^n w_i \frac{\partial \sigma}{\partial w_i} &= \sum_{i=1}^n w_i \left(\frac{\boldsymbol{\Sigma} \mathbf{w}}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}} \right)_i \\ &= \frac{1}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}} \sum_{i=1}^n w_i (\boldsymbol{\Sigma} \mathbf{w})_i \\ &= \frac{1}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \\ &= \sigma(\mathbf{w}). \end{aligned} \quad (2)$$

Thus the variance fits (1) indirectly via the volatility.

²A function $f(\mathbf{w})$ is a positively homogeneous function of degree one if $f(c\mathbf{w}) = cf(\mathbf{w})$ holds for any constant $c > 0$.

B. Equal Risk Contribution Portfolio

An equal risk contribution portfolio is a portfolio in which each asset has the same risk contribution. That is, given the risk measurement $f(\mathbf{w})$, the risk parity should satisfy

$$w_i \frac{\partial f(\mathbf{w})}{\partial w_i} = w_j \frac{\partial f(\mathbf{w})}{\partial w_j}, \quad \forall i, j. \quad (3)$$

Since $f(\mathbf{w})$ satisfies the Euler property (1), the relationship can be rewritten as

$$w_i \frac{\partial f(\mathbf{w})}{\partial w_i} = \frac{1}{n} f(\mathbf{w}), \quad \forall i. \quad (4)$$

When the volatility $\sigma(\mathbf{w}) = \sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}$ is taken as the risk measurement, the relationships (3) and (4) take the following forms:

$$w_i (\boldsymbol{\Sigma} \mathbf{w})_i = w_j (\boldsymbol{\Sigma} \mathbf{w})_j, \quad (5)$$

$$w_i (\boldsymbol{\Sigma} \mathbf{w})_i = \frac{1}{n} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}. \quad (6)$$

Given the above relationships (5) and (6), we can define some representative functions $g_i(\mathbf{w})$ to denote the risk contributions, as listed in Table I, where $\mathbf{M}_i \in \mathbb{R}^{n \times n}$ is a predefined sparse matrix with the i -th row being the same as that of the covariance matrix $\boldsymbol{\Sigma}$ and 0 elsewhere. More risk contribution functions can be found in [25, Table II].

TABLE I
DIFFERENT $g_i(\mathbf{w})$ FUNCTIONS WHEN VOLATILITY IS TAKEN AS THE RISK MEASUREMENT.

$g_i(\mathbf{w})$	$\nabla g_i(\mathbf{w})$
$w_i (\boldsymbol{\Sigma} \mathbf{w})_i$	$(\mathbf{M}_i + \mathbf{M}_i^T) \mathbf{w}$
$\frac{w_i (\boldsymbol{\Sigma} \mathbf{w})_i}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}}$	$\frac{(\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w})(\mathbf{M}_i + \mathbf{M}_i^T) \mathbf{w} - (\mathbf{w}^T \mathbf{M}_i \mathbf{w}) \boldsymbol{\Sigma} \mathbf{w}}{(\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w})^{3/2}}$
$\frac{w_i (\boldsymbol{\Sigma} \mathbf{w})_i}{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}$	$\frac{(\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w})(\mathbf{M}_i + \mathbf{M}_i^T) \mathbf{w} - (\mathbf{w}^T \mathbf{M}_i \mathbf{w})(2\boldsymbol{\Sigma}) \mathbf{w}}{(\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w})^2}$

III. A GENERAL PORTFOLIO OPTIMIZATION FRAMEWORK

In this section, we first formulate our portfolio problem, and then describe the challenges and assumptions. The flexibility and generality are illustrated in the final subsection.

A. Problem Formulation

In this paper, we jointly consider risk parity and sparsity and propose the following portfolio design formulation:

$$\begin{aligned} \underset{\mathbf{w}, \theta}{\text{minimize}} \quad & U(\mathbf{w}, \theta) \triangleq F(\mathbf{w}) + \lambda_1 \|\mathbf{w}\|_0 + \lambda_2 R(\mathbf{w}, \theta) \\ \text{subject to} \quad & \mathbf{w} \in \mathcal{W}, \end{aligned} \quad (7)$$

where

- $F(\mathbf{w})$ is assumed to be a convex function that represents an objective for a specific investment scenario, e.g., mean-variance optimization.
- $\|\mathbf{w}\|_0 \triangleq \sum_{i=1}^n 1_{\{w_i \neq 0\}}$ is the ℓ_0 -norm that regularizes the cardinality of the portfolio weights and it measures the degree of sparsity of the portfolio.

- $R(\mathbf{w}, \theta)$ measures the risk concentration and has the form

$$R(\mathbf{w}, \theta) \triangleq \sum_{i=1}^n (g_i(\mathbf{w}) - \theta)^2 1_{\{w_i \neq 0\}}, \quad (8)$$

where $g_i(\mathbf{w})$ is any smooth nonconvex differentiable function that measures the risk contribution of the i -th asset of the portfolio, as listed in Table I, and θ is a (scalar) variable denoting the average risk contributions of the selected assets (i.e., those with $w_i \neq 0$). The smaller the quantity $R(\mathbf{w}, \theta)$ is, the more uniformly the risk is distributed among the the selected assets.

- $\lambda_1, \lambda_2 \geq 0$ are the parameters that control the degrees of sparsity and risk concentration, respectively.
- In the set $\mathcal{W} = \{\mathbf{w} | \mathbf{w}^T \mathbf{1} = 1, \mathbf{w} \geq \mathbf{0}, \mathbf{w} \in \mathcal{W}_m\}$, $\mathbf{w}^T \mathbf{1} = 1$ denotes the capital budget constraint, which implies that the portfolio weights are normalized; $\mathbf{w} \geq \mathbf{0}$ denotes the long-only constraint (so that short selling is not allowed); and \mathcal{W}_m is a set that denotes the investor's profile, capital limitations, market regulations, etc..

For the sake of understanding, we can interpret our problem formulation from another perspective. Note that the objective function of (7) can be rewritten as

$$\begin{aligned} U(\mathbf{w}, \theta) &= F(\mathbf{w}) + \lambda_1 \|\mathbf{w}\|_0 + \lambda_2 \sum_{i=1}^n (g_i(\mathbf{w}) - \theta)^2 1_{\{w_i \neq 0\}} \\ &= F(\mathbf{w}) + \sum_{i=1}^n \left(\lambda_1 + \lambda_2 (g_i(\mathbf{w}) - \theta)^2 \right) 1_{\{w_i \neq 0\}} \\ &= F(\mathbf{w}) + \lambda_1 \sum_{i=1}^n \left(1 + \alpha (g_i(\mathbf{w}) - \theta)^2 \right) 1_{\{w_i \neq 0\}}, \end{aligned} \quad (9)$$

where $\alpha \triangleq \lambda_2 / \lambda_1 \geq 0$. The second term in (9) is a weighted ℓ_0 -norm of \mathbf{w} , with the weights $1 + \alpha (g_i(\mathbf{w}) - \theta)^2$ representing the deviation of the risk contribution of the i -th asset from the average level θ . Thus, the larger the risk deviation is, the heavier the weight that is put to the cardinality penalty. This interpretation indicates that our problem formulation makes sense since assets with large risk deviations should be penalized and assets with small risk deviations should be selected.

B. Challenges and Assumptions

Since each function $g_i(\mathbf{w})$ is highly nonconvex and the indicator function $1_{\{w_i \neq 0\}}$ is nonconvex and nondifferentiable, $R(\mathbf{w}, \theta)$ is clearly nonconvex and nondifferentiable. Thus, problem (7) is hard to deal with. For technical reasons, the following assumptions are made:

- (A1) \mathcal{W} is closed and convex;
- (A2) $F(\mathbf{w})$ is twice differentiable on \mathcal{W} ;

Note that the above assumptions are quite standard and are satisfied by a large class of practical portfolio optimization problems. For example, the turnover constraint, box holding position constraint are always considered in practice [38], which satisfy assumption A1. The objectives of mean-variance portfolio and index tracking can be expressed as twice differentiable functions, which can be seen clearly later in Section IV-C.

C. Flexibility and Generality

The flexibility of problem (7) comes from three aspects: First, the investment goal is expressed in an abstract function $F(\mathbf{w})$, which can be specified according to the investment scenarios in which the portfolio is applied. Second, the portfolio constraints, except the budget and long-only constraints, are summarized in the abstract convex set \mathcal{W}_m , which can also be specified according to the investor's preference. Third, the two parameters λ_1 and λ_2 can be tuned in time to either respond to changes in the market or to reflect changes in the investor's preferences and concerns. All three user-defined options provide the investor with high flexibility.

Since $F(\mathbf{w})$ can be the objective of many other portfolio problems, e.g., portfolio hedging, portfolio rebalancing, etc [11], many existing portfolios can be interpreted as special cases of the general formulation (7). For example, when $\lambda_1 = 0$ and $F(\mathbf{w}) = 0$, problem (7) becomes the basic risk parity portfolio, as in [25]. When $\lambda_2 = 0$ and $F(\mathbf{w})$ is the index tracking error, problem (7) becomes a sparse index tracking problem [39]. Likewise, when $\lambda_2 = 0$ and $F(\mathbf{w})$ is the mean-variance optimization, problem (7) becomes the sparse mean-variance optimization problem. Thus, this formulation is very general, which makes the proposed algorithm also general with respect to different application scenarios.

IV. EFFICIENT ALGORITHMS BASED ON SUCCESSIVE CONVEX OPTIMIZATION

Given the described challenges of the original problem (7), we first approximate it with a (nonconvex) differentiable problem and then derive a successive convex approximation method with the analysis of its convergence. At the end of this section, we briefly illustrate the application of the proposed algorithm to the mean-variance optimization and index tracking scenarios.

A. Smooth Approximation Problem

In the risk concentration measure $R(\mathbf{w}, \theta)$, the indicator function $1_{\{w_i \neq 0\}}$ is highly nonconvex and nondifferentiable, which will be approximated by a smooth approximation denoted by $\rho_p^\epsilon(x)$, where p and ϵ are two controlling parameters.

In Table II, three smooth approximations of the indicator function are listed without further detailed descriptions. For a detailed discussion of the approximations, please refer to [40]. For intuitive illustration, Figure 2 shows the three approximations when $p = 0.2$ and $\epsilon = 0.05$. As pointed out in [40], the smaller ϵ is, the better the indicator function is approximated around the point 0. Numerically, we can set ϵ to be very small, e.g., $\epsilon = 10^{-8}$, to achieve satisfactory approximation around the point 0.

Replacing the indicator function $1_{\{w_i \neq 0\}}$ in problem (7) with the approximation $\rho_p^\epsilon(x)$ yields the following approximation problem:

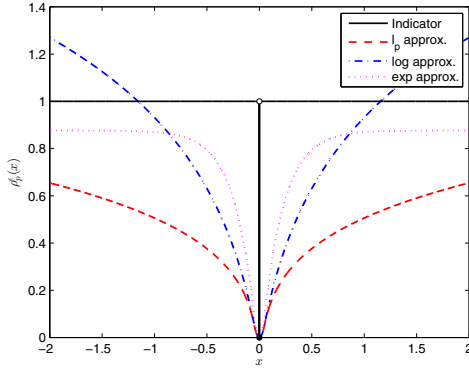
$$\begin{aligned} &\underset{\mathbf{w}, \theta}{\text{minimize}} \quad \tilde{U}(\mathbf{w}, \theta) \triangleq F(\mathbf{w}) + \lambda_1 \sum_{i=1}^n \rho_p^\epsilon(w_i) + \lambda_2 \tilde{R}(\mathbf{w}, \theta) \\ &\text{subject to} \quad \mathbf{w} \in \mathcal{W}, \end{aligned} \quad (10)$$

TABLE II

SMOOTH APPROXIMATIONS OF THE INDICATOR FUNCTION $\mathbf{1}_{\{x \neq 0\}}$ AND ITS DERIVATIVE AND THE GLOBAL QUADRATIC UPPER BOUND FUNCTIONS

$$u(w_i; w_i^k) = d(w_i^k)(w_i)^2 + c_i^k.$$

Name	Parameters	Smooth approximation $\rho_p^\epsilon(x)$	Derivative $\nabla \rho_p^\epsilon(x)$	Weights $d(x)$ for upper bounds $u(w_i; w_i^k)$
ℓ_p	$\epsilon > 0$ $0 < p \leq 1$	$\begin{cases} \frac{p}{2} \epsilon^{p-2} x^2, & x \leq \epsilon \\ x ^p - (1 - \frac{p}{2}) \epsilon^p, & x > \epsilon \end{cases}$	$\begin{cases} p \epsilon^{p-2} x, & x \leq \epsilon \\ \text{sgn}(x) p x ^{p-1}, & x > \epsilon \end{cases}$	$\begin{cases} \frac{p}{2} \epsilon^{p-2}, & x \leq \epsilon \\ \frac{p}{2} x ^{p-2}, & x > \epsilon \end{cases}$
log	$\epsilon > 0$ $p > 0$	$\begin{cases} \frac{x^2}{\log(1+1/p)}, & x \leq \epsilon \\ \frac{2\epsilon(p+\epsilon) \log(1+1/p)}{\log(1+ x /p) - \log(1+\epsilon/p) + \frac{\epsilon}{2(p+\epsilon)}}, & x > \epsilon \end{cases}$	$\begin{cases} \frac{x}{\epsilon(p+\epsilon) \log(1+1/p)}, & x \leq \epsilon \\ \frac{\text{sgn}(x)}{(p+ x) \log(1+1/p)}, & x > \epsilon \end{cases}$	$\begin{cases} \frac{1}{2\epsilon(p+\epsilon) \log(1+1/p)}, & x \leq \epsilon \\ \frac{1}{2 x (x +p) \log(1+1/p)}, & x > \epsilon \end{cases}$
exp	$\epsilon > 0$ $p > 0$	$\begin{cases} \frac{e^{-\epsilon/p}}{2p\epsilon} x^2, & x \leq \epsilon \\ -e^{- x /p} + (1 + \frac{\epsilon}{2p}) e^{-\epsilon/p}, & x > \epsilon \end{cases}$	$\begin{cases} \frac{e^{-\epsilon/p}}{p\epsilon} x, & x \leq \epsilon \\ \text{sgn}(x) \frac{e^{- x /p}}{p}, & x > \epsilon \end{cases}$	$\begin{cases} \frac{e^{-\epsilon/p}}{2p\epsilon}, & x \leq \epsilon \\ \frac{e^{- x /p}}{2p x }, & x > \epsilon \end{cases}$

Figure 2. Three smooth approximations of the indicator function, $p = 0.2$ and $\epsilon = 0.05$.

where

$$\tilde{R}(\mathbf{w}, \theta) = \sum_{i=1}^n ((g_i(\mathbf{w}) - \theta) \rho_p^\epsilon(w_i))^2. \quad (11)$$

Now, the objective function of problem (10) is a continuous and differentiable approximation of $U(\mathbf{w})$. However, it is still nonconvex. In the following subsection, we mainly focus on the approximate problem (10) instead, and develop a fast numerical algorithm based on the successive convex optimization framework. Our empirical studies on historical data show that solving this approximate problem usually yields a good solution to the original problem.

B. An Efficient Algorithm via Successive Convex Optimization

The idea of successive convex optimization is to approximate the original (possibly nonconvex) function at each iteration point by a solvable convex approximation to get an updated iteration point. It is a very useful optimization tool and has found applications in many fields [25], [41]–[43]. Suppose that we are at the k -th iteration point (\mathbf{w}^k, θ^k) . We can update θ and \mathbf{w} in parallel as follows.

1) *Updating θ* : When \mathbf{w} is fixed to \mathbf{w}^k , problem (10) with respect to θ is equivalent to the following unconstrained scalar minimization problem:

$$\underset{\theta}{\text{minimize}} \quad \sum_{i=1}^n ((g_i(\mathbf{w}^k) - \theta) \rho_p^\epsilon(w_i^k))^2. \quad (12)$$

Setting the derivative with respect to θ to zero, we can find the optimal solution in closed-form as

$$\hat{\theta} = \frac{\sum_{i=1}^n g_i(\mathbf{w}^k) (\rho_p^\epsilon(w_i^k))^2}{\sum_{j=1}^n (\rho_p^\epsilon(w_j^k))^2} = \sum_{i=1}^n x_i^k g_i(\mathbf{w}^k), \quad (13)$$

which is a weighted sum of $g_i(\mathbf{w}^k)$ with

$$x_i^k \triangleq \frac{(\rho_p^\epsilon(w_i^k))^2}{\sum_{j=1}^n (\rho_p^\epsilon(w_j^k))^2}. \quad (14)$$

2) *Updating \mathbf{w}* : Since the first term $F(\mathbf{w})$ is already assumed to be convex, we focus on the last two nonconvex terms. Note that if $F(\mathbf{w})$ is not convex for some application scenario, we can use the same successive convex approximation techniques introduced in the following to deal with the nonconvex $F(\mathbf{w})$.

For the second term $\sum_{i=1}^n \rho_p^\epsilon(w_i)$, following [40], each nonconvex approximation $\rho_p^\epsilon(w_i)$ can be upper bounded globally by a quadratic convex function at the k -th point w_i^k as follows:

$$\rho_p^\epsilon(w_i) \leq u(w_i; w_i^k) = d(w_i^k)(w_i)^2 + c_i^k, \quad (15)$$

where $d(x)$ is a weight function depending on the approximation function, as listed in Table II, and c_i^k is a properly chosen constant so that the equality holds at w_i^k . Then the second term can be globally upper bounded by a convex quadratic function

$$\sum_{i=1}^n \rho_p^\epsilon(w_i) \leq \mathbf{w}^T \mathbf{D}^k \mathbf{w} + c^k, \quad (16)$$

where

$$\mathbf{D}^k \triangleq \text{Diag}([d(w_1^k), d(w_2^k), \dots, d(w_n^k)]), \quad (17)$$

$$c^k \triangleq \sum_{i=1}^n c_i^k. \quad (18)$$

For the third term $\tilde{R}(\mathbf{w}, \theta)$, we first define

$$\tilde{g}_i(\mathbf{w}, \theta) \triangleq (g_i(\mathbf{w}) - \theta) \rho_p^\epsilon(w_i), \quad (19)$$

for clarity of presentation. Then we have

$$\tilde{R}(\mathbf{w}, \theta) = \sum_{i=1}^n (\tilde{g}_i(\mathbf{w}, \theta))^2. \quad (20)$$

Next, following the idea of the Gauss-Newton method, we can linearize the functions \tilde{g}_i inside the square and get the following approximation:

$$P(\mathbf{w}, \theta) \triangleq \sum_{i=1}^n \left(\tilde{g}_i(\mathbf{w}^k, \theta) + (\nabla \tilde{g}_i(\mathbf{w}^k, \theta))^T (\mathbf{w} - \mathbf{w}^k) \right)^2, \quad (21)$$

where

$$\nabla \tilde{g}_i(\mathbf{w}, \theta) \triangleq \rho_p^\epsilon(w_i) \cdot \nabla g_i(\mathbf{w}) + ((g_i(\mathbf{w}) - \theta) \cdot \nabla \rho_p^\epsilon(w_i)) \cdot \mathbf{e}_i. \quad (22)$$

It is easy to check that $\tilde{R}(\mathbf{w}, \theta)$ and $P(\mathbf{w}, \theta)$ have the same derivative with respect to \mathbf{w} at point \mathbf{w}^k ; that is,

$$\nabla \tilde{R}(\mathbf{w}, \theta)|_{\mathbf{w}=\mathbf{w}^k} = \nabla P(\mathbf{w}, \theta)|_{\mathbf{w}=\mathbf{w}^k},$$

where $\nabla \tilde{R}(\mathbf{w}, \theta)$ and $\nabla P(\mathbf{w}, \theta)$ denote the partial gradient of $\tilde{R}(\mathbf{w}, \theta)$ and $P(\mathbf{w}, \theta)$ with respect to \mathbf{w} , respectively.

Thus, when θ is fixed to θ^k , replacing $\sum_{i=1}^n \rho_p^\epsilon(w_i)$ and $\tilde{R}(\mathbf{w}, \theta^k)$ with the right-hand side of (16) and $P(\mathbf{w}, \theta^k)$, respectively, and removing the constant, problem (10) can be approximated at \mathbf{w}^k by

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && F(\mathbf{w}) + \lambda_1 \mathbf{w}^T \mathbf{D}^k \mathbf{w} + \lambda_2 P(\mathbf{w}, \theta^k) \\ & && + \tau \|\mathbf{w} - \mathbf{w}^k\|_2^2 \\ & \text{subject to} && \mathbf{w} \in \mathcal{W}, \end{aligned} \quad (23)$$

where the proximal term $\|\mathbf{w} - \mathbf{w}^k\|_2^2$ with $\tau > 0$ is added for convergence reasons [41].

After some mathematical manipulations, problem (23) can be further rewritten in a more compact form:

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && F(\mathbf{w}) + \mathbf{w}^T \mathbf{Q}^k \mathbf{w} + \mathbf{w}^T \mathbf{q}^k \\ & \text{subject to} && \mathbf{w} \in \mathcal{W}, \end{aligned} \quad (24)$$

where

$$\mathbf{Q}^k \triangleq \lambda_1 \mathbf{D}^k + \lambda_2 (\mathbf{A}^k)^T \mathbf{A}^k + \tau \mathbf{I}, \quad (25)$$

$$\mathbf{q}^k \triangleq 2\lambda_2 (\mathbf{A}^k)^T \mathbf{g}(\mathbf{w}^k, \theta^k) - 2 \left(\lambda_2 (\mathbf{A}^k)^T \mathbf{A}^k + \tau \mathbf{I} \right) \mathbf{w}^k \quad (26)$$

and

$$\mathbf{A}^k \triangleq [\nabla \tilde{g}_1(\mathbf{w}^k, \theta^k), \dots, \nabla \tilde{g}_n(\mathbf{w}^k, \theta^k)]^T, \quad (27)$$

$$\mathbf{g}(\mathbf{w}^k, \theta^k) \triangleq [\tilde{g}_1(\mathbf{w}^k, \theta^k), \dots, \tilde{g}_n(\mathbf{w}^k, \theta^k)]^T. \quad (28)$$

Under the assumption that $F(\mathbf{w})$ is convex, for a nonempty convex set \mathcal{W} and $\tau > 0$, problem (24) is strongly convex and can be solved by existing efficient solvers, e.g., MOSEK [44]. Moreover, if $F(\mathbf{w})$ is linear or convex quadratic, and \mathcal{W} only contains linear constraints, problem (24) reduces to a quadratic programming (QP).

3) *Iterative algorithm and convergence:* Algorithm 1 summarizes the derived iterative method. The following proposition gives the convergence of Algorithm 1.

Proposition 2. *Given the problem (10) under assumptions A1 and A2, suppose $\tau > 0$, $\gamma^k \in (0, 1]$, $\gamma^k \rightarrow 0$, $\sum_k \gamma^k = +\infty$ and $\sum_k (\gamma^k)^2 < +\infty$, and let $\{\mathbf{w}^k\}$ be the sequence generated by Algorithm 1. Then either Algorithm 1 converges*

Algorithm 1 convex optimization for general sparse risk-parity (GSRP) portfolios.

Input: $k = 0$, $\mathbf{w}^0 \in \mathcal{W}$, $\theta^0 = \sum_{i=1}^n x_i^0 g_i(\mathbf{w}^0)$, $\tau > 0$, $\{\gamma^k\} > 0$

Output: a stationary point of problem (10)

1: **repeat**

$$2: \quad x_i^k = \frac{(\rho_p^\epsilon(w_i^k))^2}{\sum_{j=1}^n (\rho_p^\epsilon(w_j^k))^2}$$

$$3: \quad \hat{\theta}^k = \sum_{i=1}^n x_i^k g_i(\mathbf{w}^k)$$

$$4: \quad \theta^{k+1} = \theta^k + \gamma^k (\hat{\theta}^k - \theta^k)$$

5: solve problem (24) to get the optimal solution $\hat{\mathbf{w}}^k$

$$6: \quad \mathbf{w}^{k+1} = \mathbf{w}^k + \gamma^k (\hat{\mathbf{w}}^k - \mathbf{w}^k)$$

$$7: \quad k \leftarrow k + 1$$

8: **until** convergence

in a finite number of iterations to a stationary point³ of (10) or every limit of $\{\mathbf{w}^k\}$ (at least one such point exists) is a stationary point of (10).

Proof: Assumption A1 in [41] is satisfied because \mathcal{W} is assumed to be closed and convex. Both the second term $\sum_{i=1}^n \rho_p^\epsilon(w_i)$ and the third term $\tilde{R}(\mathbf{w}, \theta)$ of the objective function of problem (10) are twice differentiable on the feasible set. Under assumption A2, assumptions A2 and A3 in [41] are satisfied. The closeness of \mathcal{W} and $\lim_{\theta \rightarrow +\infty} \tilde{U}(\mathbf{w}, \theta) = +\infty$ guarantees that assumption A4 in [41] is satisfied. Therefore, all the necessary conditions on the convergence are satisfied, and the rest of the proof follows directly from [41, Theorem 3]. ■

Remark 3. Note that the update of θ and \mathbf{w} in Algorithm 1 are in parallel. If we change the approximation $P(\mathbf{w}, \theta^k)$ in (24) to $P(\mathbf{w}, \theta^{k+1})$, then the update of θ and \mathbf{w} are in sequence and following [41, Theorem 3], it is easy to check that the convergence is still guaranteed.

Remark 4. As presented in Proposition 2, Algorithm 1 can be guaranteed to converge globally when the step-size parameter γ^k is properly chosen. One practical rule for choosing γ^k is as follows: given $\gamma^0 \in (0, 1]$, let

$$\gamma^k = \gamma^{k-1} (1 - \zeta \gamma^{k-1}), \quad k = 1, 2, \dots, \quad (29)$$

where $\zeta \in (0, 1)$ is a given constant [41], [42]. This rule has been applied in various numerical experiments, and in general it enjoys a very fast numerical convergence speed. See [25], [41]–[43] and [45].

C. Application to Mean-Variance Optimization and Index Tracking

As we illustrate in Section III, our problem formulation is very general and thus can be applied to many scenarios with different investment goals. In the subsection, we introduce two specific portfolio problems, which will be used as two

³The stationary point in our context is defined as follows: if the point \mathbf{z} satisfies the inequality $(\mathbf{x}_i - \mathbf{z}_i)^T \nabla_{\mathbf{z}_i} f(\mathbf{z}) \geq 0$ with $\nabla_{\mathbf{z}_i} f(\mathbf{z})$ being the gradient of $f(\cdot)$ on the direction $\mathbf{d} = \mathbf{z}_i$ at the point \mathbf{z} and \mathbf{x} being any point of the feasible set of $f(\cdot)$, then the point \mathbf{z} is the stationary point.

instances in the later experiments to evaluate our proposed portfolio.

1) *Mean-Variance Optimization*: Mean-variance optimization, pioneered by Markowitz, is widely used in portfolio design. For this investment scenario, $F(\mathbf{w})$ can be expressed as the combination of mean return and variance of the portfolio as follows:

$$F(\mathbf{w}) = \mathbf{w}^T \Sigma \mathbf{w} - \nu \mathbf{w}^T \boldsymbol{\mu}, \quad (30)$$

where $\boldsymbol{\mu}$ and Σ are denoted as the expected return vector and the covariance matrix of the available assets, respectively, and $\nu \geq 0$ is the tradeoff parameter.

Substituting the above expression of $F(\mathbf{w})$ back into the objective function of problem (24), the problem becomes

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && \mathbf{w}^T \tilde{\mathbf{Q}}^k \mathbf{w} + \mathbf{w}^T \tilde{\mathbf{q}}^k \\ & \text{subject to} && \mathbf{w} \in \mathcal{W}, \end{aligned} \quad (31)$$

where $\tilde{\mathbf{Q}}^k = \mathbf{Q}^k + \Sigma$, $\tilde{\mathbf{q}}^k \triangleq \mathbf{q}^k - \nu \boldsymbol{\mu}$, \mathbf{Q}^k and \mathbf{q}^k are defined in (25) and (26), respectively.

2) *Index Tracking*: As the name suggests, the goal of index tracking is to track a market index by setting a portfolio. For this investment scenario, $F(\mathbf{w})$ is commonly the empirical tracking error (ETE) between the target index and the constructed portfolio defined as

$$F(\mathbf{w}) = \|\mathbf{r}_c - \mathbf{R}\mathbf{w}\|_2^2, \quad (32)$$

where $\mathbf{r}_c \in \mathbb{R}^T$ denotes the T returns of the target index and $\mathbf{R} = [\mathbf{r}_1 \cdots \mathbf{r}_n] \in \mathbb{R}^{T \times N}$ contains the column-wise returns of the N individual assets over the same time period. Substituting (32) back into the objective function of problem (24), after some algebra manipulations, we still have a constrained quadratic problem having the same form as problem (31) with $\tilde{\mathbf{Q}}^k$ and $\tilde{\mathbf{q}}^k$ replaced by $\hat{\mathbf{Q}}^k \triangleq \mathbf{Q}^k + \mathbf{R}^T \mathbf{R}$ and $\hat{\mathbf{q}}^k \triangleq \mathbf{q}^k - 2\mathbf{R}^T \mathbf{r}_c$, respectively.

However, many other choices come naturally for index tracking. In our problem formulation, we will use the downside risk (DR) related to the index defined as

$$F(\mathbf{w}) = \|(\mathbf{r}_c - \mathbf{R}\mathbf{w})^+\|_2^2, \quad (33)$$

where $(\cdot)^+$ denotes the operation of elementwise positive projection. The DR means that we are interested in minimizing the tracking error only when the portfolio underperforms the index, which is reasonable because no one would mind his portfolio beating the target index. Note that the DR function is convex but nonsmooth. In order to use the successive convex approximation and guarantee the convergence of the algorithm, we need to approximate function (33) at each iteration. At the k -th iteration, function (33) can be approximated by the quadratic function

$$\tilde{F}(\mathbf{w}) = \|\mathbf{r}_c - \mathbf{R}\mathbf{w} - \mathbf{e}^k\|_2^2, \quad (34)$$

where

$$\mathbf{e}^k = -(\mathbf{R}\mathbf{w}^{k-1} - \mathbf{r}_c)^+. \quad (35)$$

A toy example is provided in Figure 3 for better intuition of this approximation function.

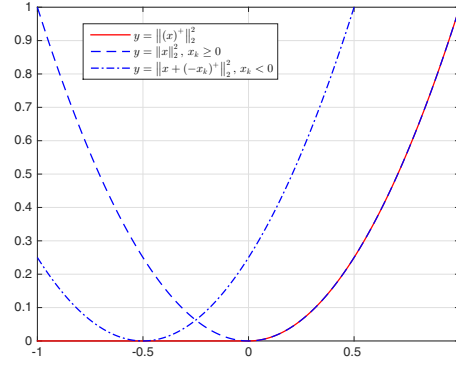


Figure 3. A toy example for illustrating the approximation of the DR.

Thus, at the k -th iteration, the term $F(\mathbf{w})$ of problem (24) is replaced by the approximation function (34). After some algebraic manipulations, we still have a constrained quadratic problem having the same form as problem (31) with $\tilde{\mathbf{Q}}^k$ and $\tilde{\mathbf{q}}^k$ replaced by $\hat{\mathbf{Q}}^k \triangleq \mathbf{Q}^k + \mathbf{R}^T \mathbf{R}$ and $\hat{\mathbf{q}}^k \triangleq \mathbf{q}^k - 2\mathbf{R}^T (\mathbf{r}_c - \mathbf{e}^k)$, respectively.

V. NUMERICAL EXPERIMENTS

Considering that financial engineering deals with complex real situations with real money, a qualified experiment for testing a portfolio strategy should satisfy three requirements: (1) it should use the historical market data; (2) the evaluation should be complete, comparable and realistic; and (3) the simulation details should be consistent with the reality and avoid backtracking.

In this section, we investigate the performance of the proposed portfolio in the context of mean-variance optimization and index tracking, respectively. In each investment scenario, we will introduce and explain our experiments in detail from the five aspects – source of data, simulation scheme, performance criteria, compared benchmarks and experimental results and analysis. Please also note that in all the following experiments, we only consider the budget and long-only constraints for simplicity.

A. Mean-Variance Optimization

1) *Data Set*: The data set consists of the daily returns of 50 randomly selected stocks of the S&P 500 from 1st, January, 2007 to 21st, September, 2008. Note that the investment period includes the 2007-2008 financial crisis, which is selected on purpose to test the portfolios' performance in an unstable investment environment. Consequently, their ability to resist dramatic fluctuations can be compared.

2) *Simulation Setup*: The simulation follows a rolling-window scheme, which is shown in Figure 4. Each rolling window can be divided into an in-sample and an out-of-sample one. In the in-sample window, three things are done: First, the mean returns are estimated via the exponentially weighted moving average, and the covariance matrix of the returns of these stocks is estimated via the regularized Tyler's

estimator [46]. Second, the parameter triplet $(\nu, \lambda_1, \lambda_2)$ is selected through cross validation. Third, the portfolio weights are designed based on these estimates, selected parameters and all of the in-sample data. Then in the next out-of-sample window, we evaluate the designed portfolios based on various criteria, which will be introduced in the next subsection. Note that there is no portfolio rebalancing in each out-of-sample window. For the missing records during the test period, we impute the latest available prices to them. For example, a stock may suffer a lack of liquidity resulting in no transaction and hence no price recorded on Friday. Then, we will impute the Thursday price to the missing one. It is reasonable because the return of a stock for barely jump a lot between several consecutive days. In our case, the missing data is bare and last for several days. However, if the missing period is longer, then some advanced imputation methods can be employed [47].

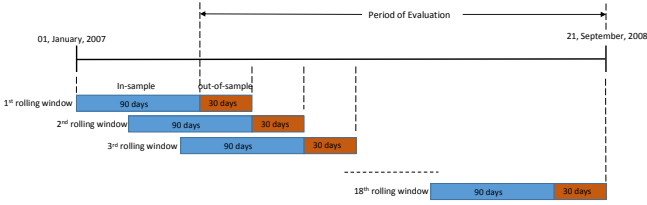


Figure 4. The rolling-window simulation scheme for mean-variance optimization.

3) *Performance Criteria*: Before introducing the performance criteria, we first define some variables that will be used in their definitions. Assume there are M rolling windows and n available assets and the length of each out-of-sample window is T . The return vector and the portfolio vector at time t in the m -th out-of-sample window are denoted by $\mathbf{r}_{tm} \in \mathbb{R}^{n \times 1}$ and $\mathbf{w}_{tm} \in \mathbb{R}^{n \times 1}$, respectively. Thus, the total return of the portfolio at time t will be $r_{tm}^p = \mathbf{r}_{tm}^T \mathbf{w}_{tm}$. Since there is no rebalancing in each window, the portfolio weights (which denotes capital allocation) changes slightly every day due to the price fluctuations. Thus, the average portfolio vector in the k -th out-of-sample window is defined as $\bar{\mathbf{w}}_k = \frac{1}{T} \sum_{t=1}^T \mathbf{w}_{tm}$. Now, the performance criteria can be introduced as follows:

a) *Mean of portfolio return*:

$$\bar{r}_m = \frac{1}{T} \sum_{t=1}^T r_{tm}^p. \quad (36)$$

b) *Volatility of portfolio return*:

$$s_m = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (r_{tm}^p - \bar{r}_m)^2}. \quad (37)$$

c) *Sharpe ratio*: Since there are no risk-free assets,

$$SR_m = \bar{r}_m / s_m, \quad (38)$$

which measures the risk premium per unit of volatility.

d) *Number of selected assets*:

$$\text{Card} = \|\bar{\mathbf{w}}_k\|_0. \quad (39)$$

e) *Gini index of normalized risk contributions (RCs)*:

The normalized risk contribution of the i -th asset is given by $\frac{\bar{w}_{m,i} (\bar{\Sigma}_m \bar{\mathbf{w}}_m)_i}{\bar{\mathbf{w}}_m^T \bar{\Sigma}_m \bar{\mathbf{w}}_m}$, where $\bar{\Sigma}_m$ the covariance matrix of the m -th rolling window. Suppose there are L selected assets and their corresponding nonzero risk contributions are sorted in an increasing order, which is denoted by π_1, \dots, π_L . The Gini index of normalized risk contributions is defined as

$$\text{Gini} = \frac{2 \sum_{l=1}^L m \pi_l}{L \sum_{l=1}^L \pi_l} - \frac{L+1}{L}, \quad (40)$$

which measures the degree of unbalance. The larger the value of the Gini index, the more concentrated are the risk contributions of the selected assets.

f) *Cumulative profit and loss (P&L)*:

$$W = W_0 \prod_{m=1}^M \prod_{t=1}^T (1 + r_{tm}^p), \quad (41)$$

where W_0 is the initial wealth.

g) *Drawdown and maximum drawdown (MDD)*: During the period $(0, t_0)$, the drawdown at time t_0 measures the decline from a historical peak in some variable $X(t)$ (wealth in our experiment) compared with $X(t_0)$, which is defined as

$$D(t_0) = \max \left\{ 0, \max_{t \in (0, t_0)} \{X(t) - X(t_0)\} \right\}. \quad (42)$$

The maximum drawdown up to time t_0 is the maximum of the drawdowns, which is defined as

$$MDD(t_0) = \max_{\tau \in (0, t_0)} \{D(\tau)\}. \quad (43)$$

Furthermore, the normalized maximum drawdown is defined as

$$NMDD(t_0) = \frac{MDD(t_0)}{X(t_h)}, \quad (44)$$

where $X(t_h)$ is the highest position of the variable related to $MDD(t_0)$.

h) *Return on turnover (ROT) and Profit/Cost*: Turnover, used to measure how actively a portfolio is managed, is defined as

$$\text{Turnover} = \sum_{m=2}^M W_m \|\mathbf{w}_m - \mathbf{w}_{m-1}\|_1, \quad (45)$$

where W_k is the total wealth at the beginning of the k -th out-of-sample window, \mathbf{w}_m is our designed portfolio vector for the m -th out-of-sample window, and \mathbf{w}_{m-1} is the last updated portfolio vector in the $(m-1)$ -th out-of-sample window. Turnover is certainly related to transaction cost, but the specific relationship between them depends on the regional market and the broker. Based on the turnover and related transaction cost, ROT and Profit/Cost are defined, respectively, as

$$\text{ROT} = \frac{\text{Net profit}}{\text{Total turnover}} = \frac{\text{Profit-Transaction cost}}{\text{Total turnover}}, \quad (46)$$

$$\text{Profit/Cost} = \frac{\text{Net profit}}{\text{Transaction cost}} = \frac{\text{Profit-Transaction cost}}{\text{Transaction cost}}. \quad (47)$$

4) *Benchmarks*: The benchmarks include MV, EW and ERC portfolios and the portfolio designed in [39]. This type of portfolio is formulated as

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && \mathbf{w}^T \Sigma \mathbf{w} - \nu \mathbf{w}^T \boldsymbol{\mu} + \lambda \|\mathbf{w}\|_2^2 \\ & \text{subject to} && \mathbf{w}^T \mathbf{1} = 1, \quad \|\mathbf{w}\|_0 \leq C_0, \end{aligned} \quad (48)$$

where C_0 is the parameter to control the number of selected assets, and can be interpreted as an ℓ_2 -penalized sparse mean-variance (SMV) portfolio. All parameters in these benchmarks are selected via cross validation.

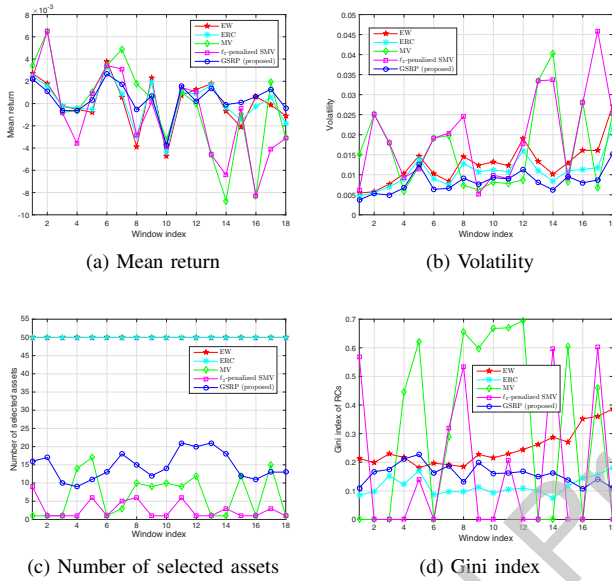


Figure 5. Evaluation the performance of portfolios in out-of-sample windows for mean-variance optimization.

5) *Numerical Results and Analysis*: In Figure 5, we compare the proposed general sparse risk-parity (GSRP) portfolio with the benchmarks in terms of some performance criteria for each out-of-sample window. Although the performances of these types of portfolios fluctuate along with the rolling windows, several interesting things can still be seen clearly: First, in terms of mean return and volatility, the GSRP portfolio is more stable and fluctuates in a small range compared with the other portfolios, especially the MV and ℓ_2 -penalized SMV portfolios. Second, the EW and ERC portfolios, as we described above, select all the stocks, while the GSRP portfolio selects 15 stocks on average. The MV and ℓ_2 -penalized SMV portfolios, as we described, easily become highly concentrated. Third, in the plot of the Gini index, our proposed portfolio is usually between the EW and the ERC portfolios and is very stable throughout all the windows, while the MV and the ℓ_2 -penalized SMV portfolios usually have high Gini index values if they select more than one stocks.

For clarity of evaluation, we summarize all the comparisons and show the average results in Figure 6, which fully demonstrates the main advantage of our method. Although our portfolio approach might not be the best in terms of one particular criterion, it achieves a superior balance among the various criteria. This versatility is desired to deal today's changeable

financial markets. In addition, the general formulation and the parameters of the proposed method make it possible to reduce it to a specific portfolio according to the market conditions.

All the evaluation results shown in Figure 6 and 5 are based on the statistics. In order to see how the proposed portfolio works over an investment period, we show the curves of the cumulative P&Ls and the corresponding drawdowns throughout the whole period in Figure 7. Observing the two sub-figures, we can see the following results: (1) Although the MV portfolio achieves the largest P&L around 14th, Oct, 2007, the MV and ℓ_2 -penalized SMV portfolios are not stable or cannot resist the market fluctuation. This can be seen clearly as both portfolios collapse rapidly after January, 2008. (2) The ERC and EW portfolios are very stable throughout the whole investment period. We can see clearly that both have small drawdowns of no more than 0.25, even as the market continues deteriorating through 2008. However, the P&L curves driven by the two portfolios expose their major drawback. The curves are in general very flat, without any obvious rising momentum; (3) Based on the previous two points, we can see that our proposed portfolio combines the advantages of the others. Its P&L curve keeps rising slowly when the market works well and stays stable, with small fluctuations, when the market declines. Its drawdown is also even better than that of the EW and ERC portfolios.

In addition, the maximum drawdowns of the P&Ls driven by these portfolios are provided in Table III. From this table, we find that the GSRP portfolio has the smallest maximum drawdown in both the absolute and normalized sense.

TABLE III
MAXIMUM DRAWDOWNS AND NORMALIZED MAXIMUM DRAWDOWNS

Portfolio	MDD	NMDD
EW	0.2502	21.29%
ERC	0.2067	17.70%
MV	0.7034	46.10%
SMV	0.6982	53.34%
GSRP	0.1255	10.59%

TABLE IV
TRANSACTION COST, NET PROFIT, ROT AND PROFIT/COST
(INITIAL BUDGET=1)

Portfolio	Cost	Net profit	ROT (bps)	Profit/Cost
EW	0.0014	0.0059 (0.59%)	10566	4.2143
ERC	0.0034	0.0242 (2.42%)	4190	7.1176
MV	0.0412	-0.2162 (-21.62%)	285	-5.2476
SMV	0.0444	-0.4317 (-43.17%)	192	-9.7230
GSRP	0.0161	0.1418 (14.18%)	1065	8.8075

In practice, transaction cost is an important issue because high transaction cost may attenuate the performance of a portfolio. In this experiment, we consider the following transaction rule: For each trade (selling and buying) of stock, the transaction cost is 0.15% of the trading volume. The transaction cost, net profit, ROT and Profit/Cost of these portfolios are provided in Table IV. Based on these numerical results, we can see that the EW portfolio has the smallest transaction cost and the largest ROT because it assigns the capital equally to all stocks so that the turnover is very small

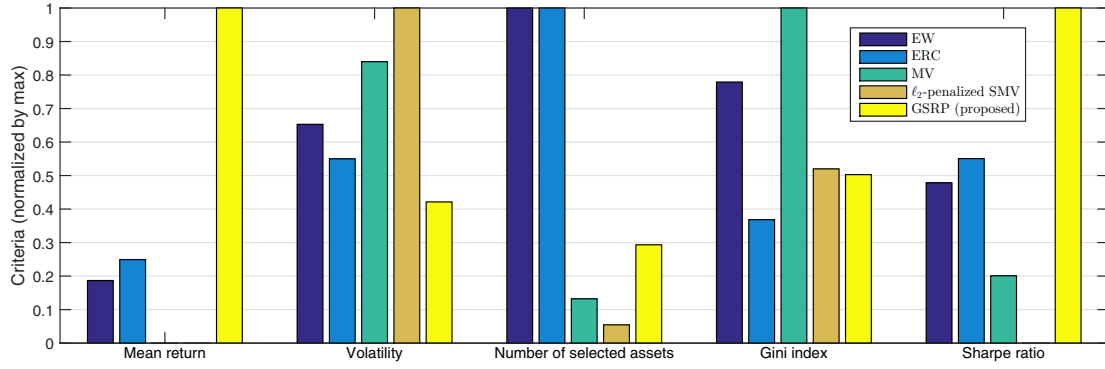
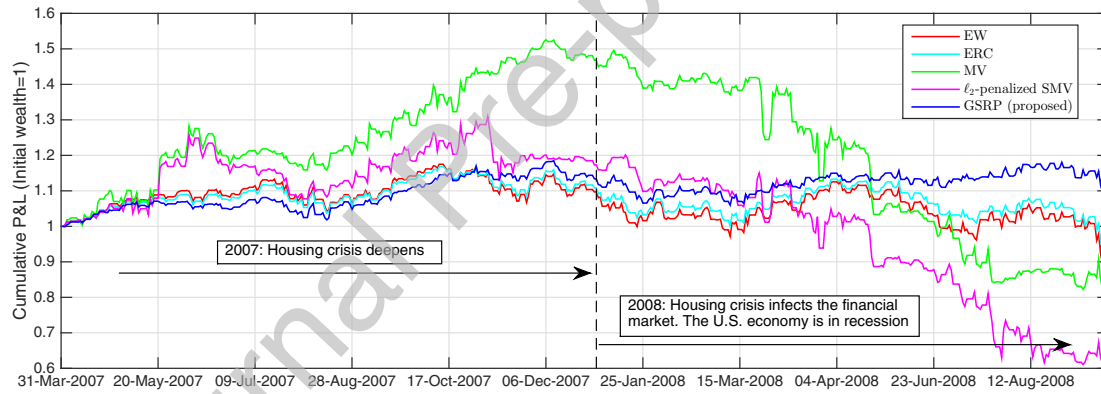
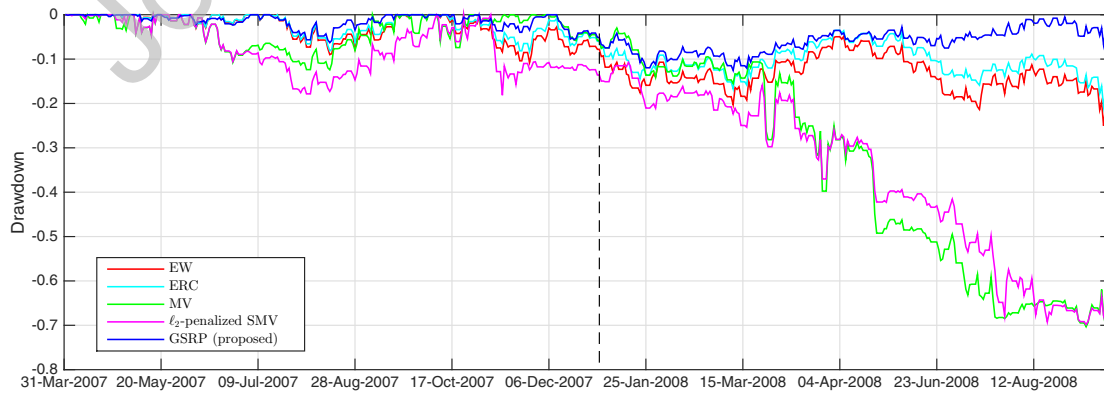


Figure 6. Summary of the performance evaluation for mean-variance optimization.



(a) Cumulative P&L



(b) Drawdown

Figure 7. Comparison in terms of cumulative P&L and drawdown.

and the transaction cost is mainly due to rebalancing. However, in terms of profit, the GSRP portfolio generates the largest net profit and also the largest Profit/Cost.

B. Index Tracking

1) *Data Set*: The data set consists of the daily returns of the Hang Seng index (HSI) and its component stocks from 30th, January, 2014 to 3rd, December, 2015. Note that since the index changes its constitutions intermittently to capture the market trends, we include all the related stocks during this experiment period in order to avoid the survivorship bias. Thus, the stock universe consists of 55 stocks (the HSI is calculated using 50 stocks).

2) *Simulation Setup*: The simulation also follows a rolling-window scheme, which is shown in Figure 8. The parameters λ_1 and λ_2 are also selected by cross validation.

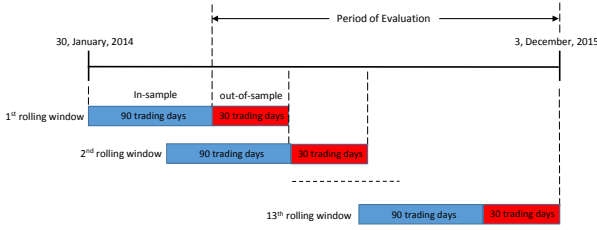


Figure 8. The rolling-window simulation scheme for index tracking.

3) *Performance Criteria*: We will also use some of the performance criteria described above. Note that the investment scenario is to track a market index. ETE and DR are also used to evaluate the portfolios.

4) *Benchmarks*: Besides the EW and ERC portfolios, we have another two benchmarks in the index tracking experiment. The first benchmark is an $\ell_{1/2}$ -penalized sparse index tracking (SIT) portfolio [48], the parameter of which is selected via cross validation. It is similar to the ℓ_2 -penalized SMV portfolio but the ℓ_2 -norm is replaced by the $\ell_{1/2}$ -norm. The second benchmark is referred to as an ideal index tracking (IIT) portfolio, and it is formulated as

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} \quad \|\mathbf{r}_c - \mathbf{R}\mathbf{w}\|_2^2 \\ & \text{subject to} \quad \mathbf{w}^T \mathbf{1} = 1. \end{aligned} \quad (49)$$

Note that this IIT portfolio is unrealistic in practice since it would likely selected most of the stocks and allows for short selling. But setting it as a benchmark provides us a lower bound in terms of empirical tracking error for a clear comparison.

5) *Numerical Results and Analysis*: In Figure 9, we compare the proposed portfolio with the benchmarks in terms of the number of selected stocks, Gini index, ETE and DR for the out-of-sample windows. For the number of selected stocks, the EW and ERC portfolios select all stocks, while the GSRP-type portfolios only select about eight stocks on average. For the Gini index, our portfolios are between the EW and ERC portfolios and approach the ERC portfolio closely. For the out-of-sample ETE, GSRP-ETE is very close the IIT and $\ell_{1/2}$ -penalized SIT portfolios, and on average is better than the EW

and ERC portfolios. The large tracking error of the GSRP-DR portfolio at the eighth window is because GSRP-DR uses the DR as the tracking criterion and it implies that the tracking curve of the GSRP-DR portfolio is above the index baseline. In terms of the out-of-sample DR, it is hard to tell which portfolio is the best, but we can see that the GSRP portfolios perform within an acceptable range.

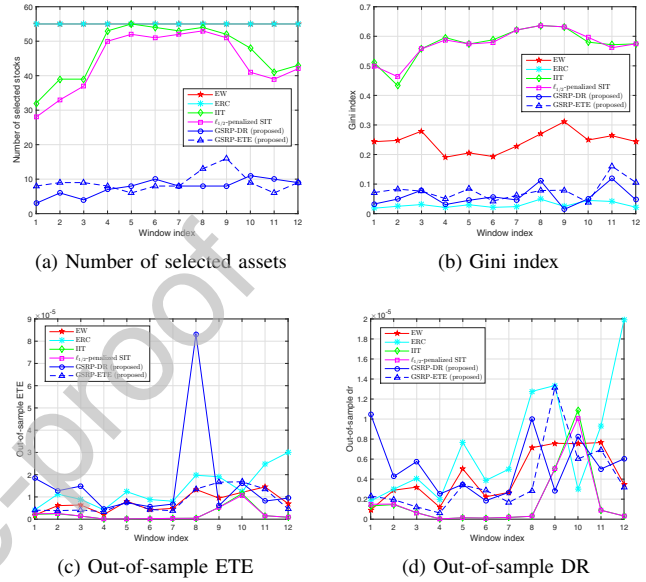


Figure 9. Evaluation of the performance of portfolios in out-of-sample windows for index tracking.

For clarity of evaluation and comparison, all the results in Figure 9 are averaged and summarized in Figure 10 and some further criteria are included. With the guarantee of a satisfactory tracking performance (DR is superior to ETE), our portfolios achieve a good balance among the further performance criteria compared with the other portfolios.

For index tracking, the primary goal is to track the index with an acceptable ETE or DR in a long-term investment period. Thus, we compare the cumulative P&Ls driven by the portfolios in Figure 11. Note that the starting index value is normalized to be 1M USD for the sake of the calculation of transaction cost. The rule of the transaction cost is exactly the same as that for the Asia-Pacific market set by Interactive Brokers⁴. Some important things can be seen clearly from this figure: First, the GSRP-DR curve is always above the GSRP-ETE curve. Recall that the only difference between the two portfolios is the tracking criterion. Thus, the superior performance of GSRP-DR implies that using DR instead of pure tracking allows the portfolio to beat the index. Second, we can see that the ERC portfolio tracks the index trend roughly and is slightly above the index curve in general. Recall that the ERC portfolio uses purely a risk parity strategy. This implies that risk parity control can make the portfolio more robust, which is very important when the market is in a dramatic fluctuation. Third, we can see clearly that the GSRP-ETE

⁴<https://www.interactivebrokers.com/en/index.php?f=commission&p=stocks1>

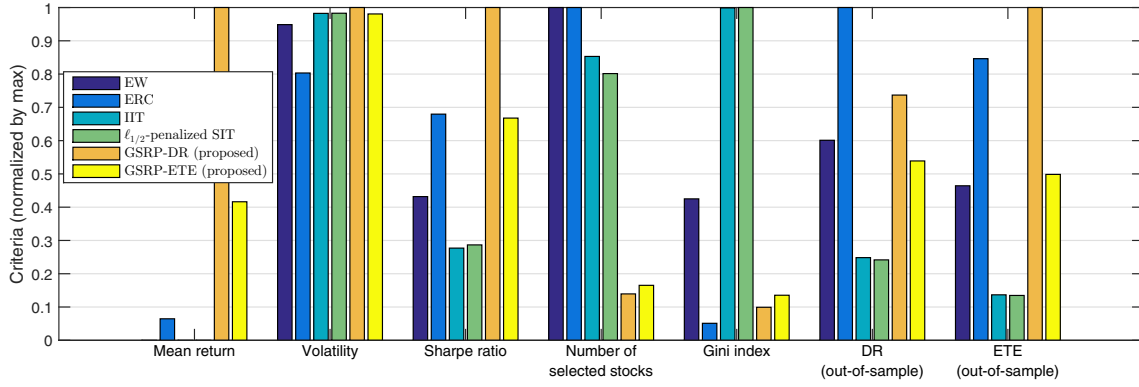


Figure 10. Summary of the performance evaluation for index tracking.

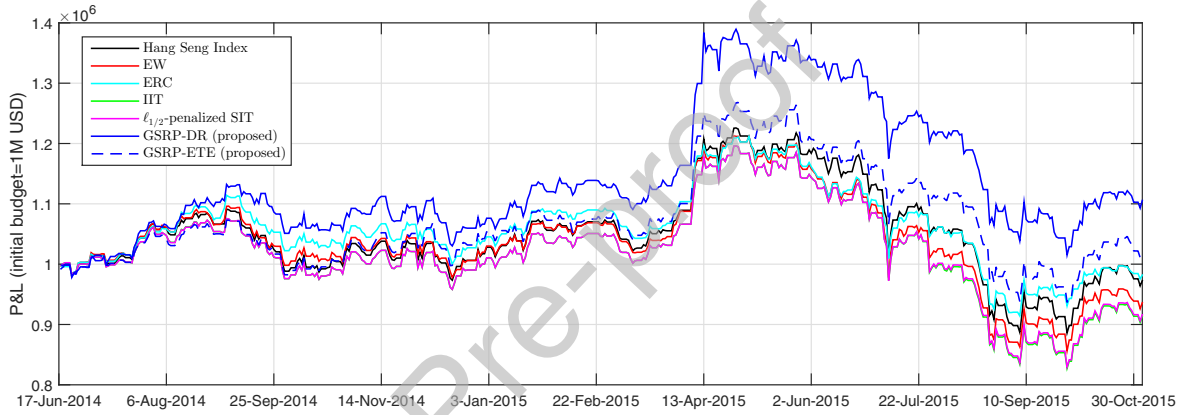


Figure 11. Comparison of tracking curves driven by different portfolios.

curve is always above the $\ell_{1/2}$ -penalized SIT curve. Recall that both portfolios incorporate sparsity and use the empirical tracking error. The only difference is that the GSRP-ETE controls the risk parity level. Last but not least, it is surprising to note that even the EW and ERC portfolios are better than the IIT and $\ell_{1/2}$ -penalized SIT portfolios. In the case of the HSI, the risk diversification is much more important than pure tracking based on the result. A portfolio with good risk diversification can probably guarantee an acceptable or even better tracking performance.

TABLE V
TRANSACTION COST, NET PROFIT, ROT AND PROFIT/COST
(INITIAL BUDGET=1M USD)

Portfolio	Cost	Net profit	ROT (bps)	Profit/Cost
EW	10,890	-63,895 (-6.39%)	-939	-5.8673
ERC	11,122	-18,035 (-1.80%)	-76	-1.6216
IIT	11,080	-86,514 (-8.66%)	-224	-7.808
SIT	11,118	-83,914 (-8.39%)	-214	-7.5475
GSRP-DR	16,464	100,560 (10.56%)	116	6.4141
GSRP-ETE	15,185	20,218 (2.02%)	25	1.3315

In addition, the numerical results for the transaction cost, net profit, Profit/Cost and ROT are presented in Table V. Our portfolios has the largest transaction cost, but this is the

necessary price for small cardinality and a low Gini index. Small cardinality and a low Gini index make our portfolios more robust, which is rewarded by the highest net profit, Profit/Costs and ROTs.

VI. CONCLUSION

In this paper, we have proposed a flexible and general portfolio formulation, which not only considers risk parity and asset selection jointly, but can also be applied to various investment scenarios. We have also derived an efficient algorithm based on the successive convex approximation method, and the derived algorithm can the guarantee the convergence under some mild assumptions. The numerical results show that the proposed formulation not only achieves a good balance between asset selection and risk diversification for different investment scenarios, but also outperforms the benchmarks in terms of many out-of-sample evaluations.

DECLARATION OF COMPETING INTEREST

No conflict of interest.

REFERENCES

- [1] H. M. Markowitz, "Portfolio selection," *Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.

- [2] —, “The optimization of a quadratic function subject to linear constraints,” *Naval Research Logistics Quarterly*, vol. 3, no. 1-2, pp. 111–133, 1956.
- [3] —, *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press, 1968.
- [4] R. O. Michaud, “The Markowitz optimization enigma: Is optimized optimal?” *ICFA Continuing Education Series*, vol. 1989, no. 4, pp. 43–54, 1989.
- [5] T. Roncalli, *Introduction to Risk Parity and Budgeting*. CRC Press, 2013.
- [6] R. Jagannathan and T. Ma, “Risk reduction in large portfolios: Why imposing the wrong constraints helps,” *Journal of Finance*, vol. 58, no. 4, pp. 1651–1684, 2003.
- [7] V. DeMiguel, L. Garlappi, F. J. Nogales, and R. Uppal, “A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms,” *Management Science*, vol. 55, no. 5, pp. 798–812, 2009.
- [8] V. DeMiguel, L. Garlappi, and R. Uppal, “Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?” *Review of Financial Studies*, vol. 22, no. 5, pp. 1915–1953, 2009.
- [9] P. Jorion, “Bayes-Stein estimation for portfolio analysis,” *Journal of Financial and Quantitative Analysis*, vol. 21, no. 03, pp. 279–292, 1986.
- [10] O. Ledoit and M. Wolf, “Improved estimation of the covariance matrix of stock returns with an application to portfolio selection,” *Journal of Empirical Finance*, vol. 10, no. 5, pp. 603–621, 2003.
- [11] J. Brodie, I. Daubechies, C. De Mol, D. Giannone, and I. Loris, “Sparse and stable Markowitz portfolios,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 30, pp. 12 267–12 272, 2009.
- [12] Y.-M. Yen, “Sparse weighted norm minimum variance portfolio,” Ph.D. dissertation, London School of Economics and Political Science, 2012.
- [13] B. Fastrich, S. Paterlini, and P. Winker, “Constructing optimal sparse portfolios using regularization methods,” *Computational Management Science*, pp. 1–18, 2013.
- [14] E. Qian, “Risk parity portfolios: Efficient portfolios through true diversification,” *Panagora Asset Management*, Sept. 2005.
- [15] —, “On the financial interpretation of risk contribution: Risk budgets do add up,” *Journal of Investment Management*, vol. 4, no. 4, p. 41, 2006.
- [16] S. Maillard, T. Roncalli, and J. Teiletche, “The properties of equally weighted risk contribution portfolios,” *Journal of Portfolio Management*, vol. 36, no. 4, pp. 60–70, 2010.
- [17] X. Bai, K. Scheinberg, and R. Tutuncu, “Least-squares approach to risk parity in portfolio selection,” *Available at SSRN 2343406*, 2013.
- [18] T. Roncalli and G. Weisang, “Risk parity portfolios with risk factors,” *Available at SSRN 2155159*, 2012.
- [19] B. Bruder and T. Roncalli, “Managing risk exposures using the risk budgeting approach,” University Library of Munich, Germany, Tech. Rep., 2012.
- [20] K. Boudt, P. Carl, and B. G. Peterson, “Asset allocation with conditional value-at-risk budgets,” *J. Risk*, vol. 15, no. 3, pp. 39–68, 2013.
- [21] M. Haugh, G. Iyengar, and I. Song, “A generalized risk budgeting approach to portfolio construction,” *Available at SSRN 2462145*, 2014.
- [22] D. B. Chaves, J. C. Hsu, F. Li, and O. Shakernia, “Efficient algorithms for computing risk parity portfolio weights,” *Journal of Investing*, vol. 21, pp. 150–163, 2012.
- [23] F. Spinu, “An algorithm for computing risk parity weights,” *Available at SSRN 2297383*, 2013.
- [24] T. Griveau-Billion, J.-C. Richard, and T. Roncalli, “A fast algorithm for computing high-dimensional risk parity portfolios,” *arXiv preprint arXiv:1311.4057*, 2013.
- [25] Y. Feng and D. P. Palomar, “SCRIP: Successive convex optimization methods for risk parity portfolio design,” *IEEE Transactions on Signal Processing*, vol. 63, no. 19, pp. 5285–5300, 2015.
- [26] P. Demey, S. Maillard, and T. Roncalli, “Risk-based indexation,” *Available at SSRN 1582998*, 2010.
- [27] Z. Cazalet, P. Grison, and T. Roncalli, “The smart beta indexing puzzle,” *Available at SSRN 2294395*, 2013.
- [28] H. Lohre, U. Neugebauer, and C. Zimmer, “Diversified risk parity strategies for equity portfolio selection,” *Journal of Investing*, vol. 21, no. 3, pp. 111–128, 2012.
- [29] R. Leote de Carvalho, X. Lu, and P. Moulin, “Demystifying equity risk-based strategies: A simple alpha plus beta description,” *Journal of Portfolio Management*, vol. 38, no. 3, pp. 56–70, 2012.
- [30] A. Kohler and H. Wittig, “Rethinking portfolio rebalancing: introducing risk contribution rebalancing as an alternative approach to traditional value-based rebalancing strategies,” *Journal of Portfolio Management*, vol. 40, no. 3, pp. 34–46, 2014.
- [31] C. Asness, A. Frazzini, and L. H. Pedersen, “Leverage aversion and risk parity,” *Financial Analysts Journal*, vol. 68, no. 1, pp. 47–59, 2012.
- [32] D. Chaves, J. Hsu, F. Li, and O. Shakernia, “Risk parity portfolio vs. other asset allocation heuristic portfolios,” *Journal of Investing*, vol. 20, no. 1, pp. 108–118, 2011.
- [33] R. Deguest, L. Martellini, and A. Meucci, “Risk parity and beyond—from asset allocation to risk allocation decisions,” *Available at SSRN 2355778*, 2013.
- [34] R. M. Anderson, S. W. Bianchi, and L. R. Goldberg, “Will my risk parity strategy outperform?” *Financial Analysts Journal*, vol. 68, no. 6, pp. 75–93, 2012.
- [35] H. Konno and A. Wijayanayake, “Portfolio optimization problem under concave transaction costs and minimal transaction unit constraints,” *Mathematical Programming*, vol. 89, no. 2, pp. 233–250, 2001.
- [36] M. S. Lobo, M. Fazel, and S. Boyd, “Portfolio optimization with linear and fixed transaction costs,” *Annals of Operations Research*, vol. 152, no. 1, pp. 341–365, 2007.
- [37] Y. Feng and D. P. Palomar, “Portfolio optimization with asset selection and risk parity control,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 6585–6589.
- [38] W. Kim, J. Kim, and F. Fabozzi, *Robust Equity Portfolio Management: Formulations, Implementations, and Properties Using MATLAB*. John Wiley & Sons, 2015.
- [39] A. Takeda, M. Niranjan, J.-y. Gotoh, and Y. Kawahara, “Simultaneous pursuit of out-of-sample performance and sparsity in index tracking portfolios,” *Computational Management Science*, vol. 10, no. 1, pp. 21–49, 2013.
- [40] J. Song, P. Babu, and D. P. Palomar, “Sparse generalized eigenvalue problem via smooth optimization,” *IEEE Transactions on Signal Processing*, vol. 63, no. 7, pp. 1627–1642, 2015.
- [41] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, “Decomposition by partial linearization: Parallel optimization of multi-agent systems,” *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 641–656, Feb. 2014.
- [42] F. Facchinei, G. Scutari, and S. Sagratella, “Parallel selective algorithms for nonconvex big data optimization,” *IEEE Transactions on Signal Processing*, vol. 63, no. 7, pp. 1874–1889, April 2015.
- [43] A. Daneshmand, F. Facchinei, V. Kungurtsev, and G. Scutari, “Hybrid random/deterministic parallel algorithms for convex and nonconvex big data optimization,” *IEEE Transactions on Signal Processing*, vol. 63, no. 15, pp. 3914–3929, Aug. 2015.
- [44] MOSEK, “The MOSEK optimization toolbox for MATLAB manual,” Tech. Rep., 2013. [Online]. Available: <http://www.mosek.com>
- [45] Y. Yang, G. Scutari, D. P. Palomar, and M. Pesavento, “A parallel stochastic approximation method for nonconvex multi-agent optimization problems,” *arXiv preprint arXiv:1410.5076*, 2014.
- [46] Y. Sun, P. Babu, and D. P. Palomar, “Regularized Tyler’s scatter estimator: Existence, uniqueness, and algorithms,” *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5143–5156, 2014.
- [47] J. Liu, S. Kumar, and D. P. Palomar, “Parameter estimation of heavy-tailed ar model with missing data via stochastic em,” *IEEE Transactions on Signal Processing*, vol. 67, no. 8, pp. 2159–2172, 2019.
- [48] F. Xu, Z. Xu, and H. Xue, “Sparse index tracking based on $\ell_{1/2}$ model and algorithm,” *arXiv preprint arXiv:1506.05867*, 2015.