Sparse minimax portfolio and Sharpe ratio models

ZU CHENCHEN[1], YANG XIAOQI[1], AND CARISA KWOK WAI YU[2]

Abstract

In this paper we investigate sparse portfolio selection models with a regularized l_p-norm term $(0 < p \leq 1)$ and bounded negative shorting constraints. We obtain some basic properties of several linear l_p-sparse minimax portfolio models in terms of the regularization parameter. In particular, we introduce an l_1-sparse minimax Sharpe ratio model by guaranteeing a positive denominator with a pre-selected parameter and design a parametric algorithm for finding its global solution. We carry out numerical experiments of linear l_p-sparse minimax portfolio models with 1200 stocks from Hang Seng Index, Shanghai Securities Composite Index, and NASDAQ Index and compare their performance with l_p-sparse mean-variance models. We test the effect of the regularization parameter and the bounded negative shorting parameter on the level of sparsity, risk, and rate of return respectively and find that portfolios including fewer stocks of the linear l_p-sparse minimax models tend to have lower risks and lower rates of return. However, for the l_p-sparse mean-variance models, the corresponding changes are not so significant.

Key words: sparse minimax portfolio selection model; Sharpe ratio, shorting, sparse mean-variance model, l_q regularization.

[1]Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong

[2]Department of Mathematics, Statistics and Insurance, The Hang Seng University of Hong Kong, Shatin, Hong Kong

1 Introduction

In 1952, Harry Markowitz formulated the portfolio selection problem as a quadratic programming problem (Markowitz, 1952), which is known as the mean-variance model. This quantitative framework has since then become the milestone in the field of portfolio selection and remains the dominant technique in use today. In this framework, two critical elements, return and risk, are expressed by the expected return and the variance of the portfolio. Given N securities with return vector $(\bar{y}_j)_N$ and covariance matrix $(Q_{ij})_{N\times N}$, the optimal portfolio of the mean-variance model is the solution of the following linear constrained quadratic optimization problem

$$\begin{aligned} \min_{w_1,\ldots,w_N} \quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} Q_{ij}w_iw_j \\ \text{s.t.} \quad & \sum_{j=1}^{N} \bar{y}_j w_j \geq G \\ & \sum_{j=1}^{N} w_j = 1, \end{aligned} \tag{1.1}$$

where G is the required return in the investment. In addition, the constraint $w_j \geq 0$ is included if the short selling is prohibited in the investment.

The mean-variance model provides a quantitative way to seek the balance between return and risk and has been demonstrated to be effective in empirical studies (Markowitz and Van Dijk, 2003; Das et al., 2011; Markowitz, 2013). Many competent algorithms are developed to solve the mean-variance model with parametric formulation, typically the critical line algorithm (Markowitz and Todd, 2000), which has been efficiently used to investigate large-scale portfolio problems (Niedermayer and Niedermayer, 2010; Qi, 2020). The active set method for multi-criteria convex quadratic programming problems (Goh and Yang, 1996) also substantially contributes to finding the efficient frontier of the parametric mean-variance model. On the other hand, various risk measures were proposed to replace the portfolio variance in the mean-variance model and establish alternative portfolio selection rules. For example, the ones with linear structure are representative. Sharpe (1967) viewed the market responsiveness as the risk measure and built a linear approximation of the mean-variance model. After that, mean absolute deviation (Konno and Yamazaki, 1991), minimum return (Young, 1998), and l_∞ function (Cai et al., 2000) were introduced as new types of linear risk measure and were also proved to be competitive.

In modern society, portfolios including many securities are not desirable, especially for large-scale investments or retail investors. Therefore, finding sparse optimal portfolios becomes an essential issue for portfolio selection. The terminology *cardinality* is also universally used in literature when discussing the sparse portfolios (see Chang et al. (2000) and Woodside-Oriakhi et al. (2011)). The regularization method is a promising method for pursuing the sparse portfolios under the mean-variance model. In particular, the target portfolio is generated by the following *l_p-sparse ($0 < p \leq 1$) mean-variance* model

$$\begin{aligned} \min_{w_1,\dots,w_N} \quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} Q_{ij} w_i w_j + \tau\|w\|_p \\ \text{s.t.} \quad & \sum_{j=1}^{N} \bar{y}_j w_j \geq G \\ & \sum_{j=1}^{N} w_j = 1, \end{aligned} \tag{1.2}$$

which modifies the model (1.1) by adding $\tau\|w\|_p$. The term $\|w\|_p := (\sum_{j=1}^{N} |w_j|^p)^{\frac{1}{p}}$ is known as the *l_p norm* or *l_p regularizer*, and the regularization parameter τ provides a tradeoff between sparsity and accuracy. The l_1-sparse mean-variance model has been demonstrated practical for promoting the sparsity of portfolios (see Brodie et al. (2009), DeMiguel et al. (2009), and Dai and Wen (2018)). The l_1 regularizer has been widely adopted to seek sparsity for industrial problems, such as image reconstruction, data analysis, and machine learning (see Tibshirani (1996), Chen et al. (2001), and Daubechies et al. (2004)). When $0 < p < 1$, Chen et al. (2013) and Fastrich et al. (2015) claimed that sparse optimal portfolios can be obtained by solving the l_p-sparse mean-variance model. In addition, it was shown by Chartrand (2007), Xu et al. (2010), and Hu et al. (2017) that the use of l_p norm rather than l_1 norm produces more sparse solutions for industrial problems, although the computation of the l_p-sparse formulation is relatively complicated. Apart from the regularization technique, other selection criteria can be applied to achieve sparse optimal portfolios based on the mean-variance model. By considering nonnegativity constrained portfolios, Jagannathan and Ma (2003) obtained an optimal portfolio including only 24.1 stocks out of 500 stocks. Qi et al. (2019) applied their portfolio selection model to 1800-stock problems, and the minimal number of selected stocks can be 62.11. Woodside-Oriakhi et al. (2011) presented a series of heuristic algorithms for the cardinality constrained mean-variance model, which in essence is a quadratic mixed-integer problem. As indicated in their experiments, the proposed algorithms effectively generate the efficient frontier under the condition that the number of stocks included is fixed.

To the best of our knowledge, research on the sparsity of portfolios mainly centers on the mean-variance model but absents from linear portfolio models. In this article, we take the minimax rule (Young, 1998) (see more details in Section 2) for instance to discover the linear portfolio models regularized by the l_1 and l_p $(0 < p < 1)$ norms. For the regularized models, the tunable regularization parameter can be viewed as a controller to adjust the level of sparsity and the space for short selling. Related features or properties are analyzed for the better use of the sparse models. Correspondingly, we also study the l_p-sparse $(0 < p \leq 1)$ minimax models numerically. Xu et al. (2010) and Hu et al. (2017) justified that, among l_p $(0 < p \leq 1)$ regularizers, the $l_{\frac{1}{2}}$ regularizer performs best. Thus, we take the $l_{\frac{1}{2}}$-sparse portfolio model as representative of the l_p-sparse (when $0 < p < 1$) portfolio model. The benchmarks we select in the numerical experiments are the equal-weighted rule, the l_1-sparse mean-variance model, and the $l_{\frac{1}{2}}$-sparse mean-variance model. In order to compare different sparse models, we observe their out-of-sample performances at the same level of sparsity. We find, compared with the l_1-sparse minimax model, the $l_{\frac{1}{2}}$-sparse minimax model is more competitive when the level of sparsity is extremely high. As resulting portfolios become less sparse, the performances of both models are comparable.

On the other hand, we also consider the l_1-sparse Sharpe ratio model based on the minimax rule. In the area of performance assessment, Treynor index (Treynor, 1965), Sharpe index (Sharpe, 1966), and Jensen index (Jensen, 1968) are three most popular performance measures used to rank the performance of portfolios or mutual fund managers. The Sharpe index, also known as the Sharpe ratio, was first put forward by Sharpe (1966) as an extension of the Treynor index. The classical Sharpe ratio was originally defined as $\frac{E(R)-R_f}{\sigma(R)}$, where $E(R)$ and $\sigma(R)$ stand for the expected value of return and the standard deviation of return, which are exactly the return and risk in mean-variance framework; R_f represents the risk-free return. On the basis of Sharpe ratio, a series of Sharpe-type ratios were constructed where alternative risk measures were used to substitute the variance (i.e., the denominator of the classical Sharpe ratio), say the Calmar ratio (Young, 1991), the Sortino ratio (Sortino and van de Meer, 1991), the Burke ratio (Burke, 1994), and the Sterling ratio (McCafferty, 2003). Apart from ranking the performance of portfolios, the Sharpe ratio is also employed as an objective function of the portfolio optimization model (Benninga, 2014; Elton et al., 2014). Likewise, we generalize the Sharpe ratio maximization model based on the risk measure of the minimax model (we call it the minimax risk measure for simplicity) and study its l_1-sparse formulation. For the proper use of the generalized Sharpe ratio in the optimization model, we introduce a modified minimax risk measure by adding a constant to keep the denominator positive. To find a global solution of the l_1-sparse generalized minimax Sharpe ratio model, we develop a parametric algorithm, which extends the algorithm in Konno and Kuno (1990).

It's worth noting that a non-negative return is an underlying assumption for the original Sharpe ratio. Bacon (2008) pointed out that the negative return makes the Sharpe ratio difficult to interpret. In fact, a higher value of the Sharpe ratio represents a higher rank of the portfolio. However, a negative return generates a negative Sharpe ratio, then results in a perverse ranking. Specifically, a larger not less variance is more preferable when the return is negative. The nonnegativity of the return is also assumed for the generalized Sharpe ratio based on the minimax risk measure. For a similar reason, the denominator of the generalized Sharpe ratio should be positive. But unfortunately, the value of the minimax risk measure is possibly negative. As a result, it is infeasible to directly adopt the minimax risk measure as the denominator. To overcome this difficulty, we propose a revised minimax risk measure $\lambda - M_p$ to replace $-M_p$ (see (3.1)), where λ is a parameter such that $\lambda - M_p > 0$. The function of the risk measure is to rank the risk; in this sense, $\lambda - M_p$ is consistent with $-M_p$. Therefore, the revision is valid.

The main contributions of this article are summarized as follows.

(i) We construct the l_1-sparse and l_p-sparse ($0 < p < 1$) minimax models and investigate their mathematical properties.

(ii) We formulate the l_1-sparse minimax Sharpe ratio model by guaranteeing a positive denominator with a pre-selected parameter.

(iii) We develop a global algorithm to solve the (generalized) nonconvex l_1-sparse minimax Sharpe ratio model.

The rest of the article is organized as follows. In Sections 2 and 3, we construct and analyze the l_p-sparse ($0 < p \leq 1$) minimax model and the l_1-sparse minimax Sharpe ratio model, then follow numerical experiments in Section 4. Eventually, the article is concluded in Section 5.

2 The sparse minimax models

We observe N securities over T time periods and let y_{jt} represent the rate of return of security j in time period t. A portfolio is denoted by a vector of weights w_j ($j = 1, 2, \ldots, N$), which stands for the percentage of the budget invested in security j. Let $\bar{y}_j$ be the average rate of return of security j, i.e. $\bar{y}_j = \frac{1}{T}\sum_{t=1}^{T} y_{jt}$, then the feasible region of the portfolio is given by

$$\mathcal{F} := \left\{ w = (w_1, w_2, \dots, w_N) \ : | \ \sum_{j=1}^{N} \bar{y}_j w_j \geq G; \ \sum_{j=1}^{N} w_j = 1; \ w_j \geq \alpha, \ j = 1, \dots, N \right\},$$

where G is the minimum level of rate of return and α is the lower bound of the portfolio. A security j is called an active security if $w_j \neq 0$.

The minimax portfolio selection model, proposed by Young (1998), by maximizing the minimum return of the portfolio over all the time periods, is given as follows

$$\max_{w \in \mathcal{F}} \min_{t=1,\dots,T} \quad \sum_{j=1}^{N} y_{jt} w_j,$$

which is equivalent to

$$\min_{w \in \mathcal{F}, M_p} -M_p \qquad \text{s.t.} \ \sum_{j=1}^{N} y_{jt} w_j \geq M_p, \ t = 1, \dots, T.$$

In general, from the theory of linear programming (Matousek and Gartner, 2007), we know that the sparsity level of optimal solutions to this model tends to be very low if $\alpha \neq 0$, even the number of non-active securities can be zero in many situations. Therefore, we add an l_p $(0 < p \leq 1)$ norm in the objective function to seek a sparse optimal portfolio.

We consider the following l_p-sparse $(0 < p \leq 1)$ minimax model

$$\begin{aligned} \min_{w \in \mathcal{F}, M_p} \quad & -M_p + \tau \|w\|_p^p \\ \text{s.t.} \quad & \sum_{j=1}^{N} y_{jt} w_j \geq M_p, \ t = 1, \dots, T, \end{aligned} \tag{2.1}$$

where $\|w\|_p := (\sum_{j=1}^{N} |w_j|^p)^{\frac{1}{p}}$ and $\tau(\geq 0)$ is a tunable parameter of the l_p norm. Unlike the mean-variance model, the minimax model requires a finite lower bound condition, since the problem may lead to an infinite optimal value if $\alpha = -\infty$. But for a finite α, the feasible region $\mathcal{F}$ is bounded, then the corresponding optimal value is bounded. Thus, we set $\alpha > -\infty$ to guarantee the validness of the problem (2.1). Furthermore, when considering the l_1-sparse minimax model, we restrict ourselves on the case $\alpha < 0$, which means the limited short selling is allowed in the portfolio; otherwise, the problem will reduce to the original minimax model due to $\|w\|_1 = 1$.

Here are remarks to discuss the parameter τ of l_p-sparse minimax models.

(i) Let the solution of (2.1) be denoted by $(w_{(\tau)}, M_{p(\tau)})$. Following Brodie et al. (2009), we observe that an optimal solution of the l_p-sparse minimax model satisfies the following relation

$$(\tau_1 - \tau_2)(\|w_{(\tau_2)}\|_p^p - \|w_{(\tau_1)}\|_p^p) \geq 0. \tag{2.2}$$

Indeed, we have

$$\begin{aligned} M_{p(\tau_1)} + \tau_1 \|w_{(\tau_1)}\|_p^p \leq & M_{p(\tau_2)} + \tau_1 \|w_{(\tau_2)}\|_p^p = M_{p(\tau_2)} + \tau_2 \|w_{(\tau_2)}\|_p^p + (\tau_1 - \tau_2)\|w_{(\tau_2)}\|_p^p \\ \leq & M_{p(\tau_1)} + \tau_2 \|w_{(\tau_1)}\|_p^p + (\tau_1 - \tau_2)\|w_{(\tau_2)}\|_p^p \\ = & M_{p(\tau_1)} + \tau_1 \|w_{(\tau_1)}\|_p^p + (\tau_1 - \tau_2)(\|w_{(\tau_2)}\|_p^p - \|w_{(\tau_1)}\|_p^p). \end{aligned}$$

Notice that two inequalities are obtained by the minimization properties of respective optimal solutions. If we use the l_p-norm as an indicator of sparsity, inequality (2.2) indicates that a larger τ leads to a portfolio with a higher level of sparsity (see Figures 3(a) and 4(a)).

(ii) Let w^+ and w^- stand for the componentwise positive and negative parts of w, respectively. When $p = 1$, by making use of the constraint $\sum_{j=1}^N w_j = 1$, it follows from the inequality (2.2) that

$$(\tau_1 - \tau_2)(\|w^-_{(\tau_2)}\|_1 - \|w^-_{(\tau_1)}\|_1) \geq 0. \tag{2.3}$$

Noting $\sum_{j=1}^N w_j = \|w^+\|_1 - \|w^-\|_1$ and $\|w\|_1 = \|w^+\|_1 + \|w^-\|_1$, we obtain the relation $\|w_{(\tau_2)}\|_1 - \|w_{(\tau_1)}\|_1 = 2(\|w^-_{(\tau_2)}\|_1 - \|w^-_{(\tau_1)}\|_1)$. Thus (2.3) holds from (2.2). Inequality (2.3) implies that the portfolio with a smaller τ has more short selling stocks and that a non-negative portfolio may be obtained when τ is sufficiently large (see Figures 3(b) and 4(b)).

3 The sparse minimax Sharpe ratio model

In this section, we extend the classical Sharpe ratio to a generalized version based on the minimax risk measure and consider the minimization of this modified Sharpe ratio plus an l_1 norm. To solve this model, a parametric algorithm is proposed as a generalization of the algorithm introduced by Konno and Kuno (1990).

3.1 (Generalized) l_1-sparse minimax Sharpe ratio model

As the original minimax risk measure $-M_p$ can be negative, in this subsection, we consider the following generalized l_1-sparse minimax Sharpe ratio model

$$\begin{aligned}
\min_{w\in\mathcal{F},M_p\in\mathbb{R}} \quad & -\frac{(\sum_{j=1}^N \bar{y}_j w_j - r_f)}{\lambda - M_p} + \tau\|w\|_1 \\
\text{s.t.} \quad & \sum_{j=1}^N y_{jt} w_j \geq M_p, t = 1, \ldots, T \\
& \sum_{j=1}^N w_j = 1 \\
& w_j \geq \alpha, j = 1, \ldots, N,
\end{aligned} \tag{3.1}$$

where r_f represents the risk-free rate of return and λ is a constant such that $\lambda - M_p > 0$. Also, relations (2.2) and (2.3) still hold for this l_1-sparse minimax Sharpe ratio model.

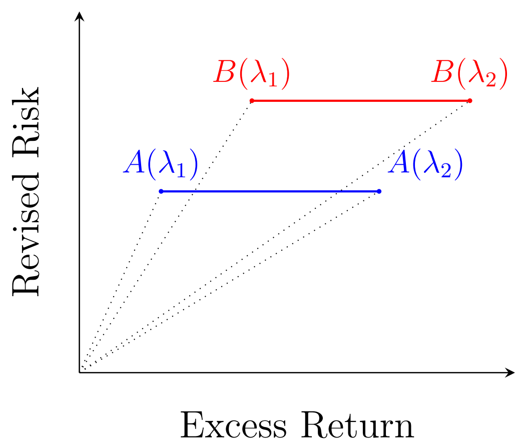

Figure 1. Different λ of (generalized) Sharpe ratio

It is remarkable that the choice of λ may influence the final result. Figure 1 states a risk-return space. Point A(λ_1) represents the excess return of portfolio A and its corresponding risk revised by λ_1, and other points are denoted similarly. Clearly, in this space, the value of the Sharpe ratio is expressed by the slope of the point. As we can see from this figure, under the selection of λ_1, portfolio A has a better performance than B since the gradient of A(λ_1) is steeper than that of B(λ_1), while in the situation with λ_2, the rank is opposite. But we have to emphasize that no matter how λ is chosen, the method still keeps the essence of the minimax risk and the corresponding optimal portfolio is reasonable on logic.

3.2 The parametric algorithm

In this subsection, we introduce an algorithm to find a global solution of the l_1-sparse minimax Sharpe ratio model (3.1). To this end, we study a generalization of the parametric algorithm proposed by Konno and Kuno (1990), which is to minimize the sum of a differentiable convex function and a linear fractional function subject to linear inequality constraints. More specifically, we aim to develop a parametric algorithm for the following nondifferentiable generalized linear fractional programming problem

$$\begin{aligned} \min \quad & g(x) - \frac{c_1^{\mathrm{T}}x + c_{10}}{c_2^{\mathrm{T}}x + c_{20}} \\ \text{s.t.} \quad & x \in X =: \{x \in \mathbb{R}^n : A_1 x \geq b_1,\ A_2 x = b_2\}, \end{aligned} \tag{3.2}$$

where function $g(x)$ is convex but not necessarily differentiable, and c_1, $c_2 \in \mathbb{R}^n$, c_{10}, $c_{20} \in \mathbb{R}$, $A_1 \in \mathbb{R}^{p\times n}$, $A_2 \in \mathbb{R}^{q\times n}$, $b_1 \in \mathbb{R}^p$, and $b_2 \in \mathbb{R}^q$ are given parameters of the problem. We assume that the feasible region X is non-empty and bounded; for any $x \in X$,

$$c_1^{\mathrm{T}} + c_{10} \geq 0,\ c_2^{\mathrm{T}} + c_{20} > 0.$$

The auxiliary problem for the parametric algorithm is as follows.

$$\begin{aligned} \min_{x,\xi} \quad & g(x) - 2\xi\sqrt{c_1^{\mathrm{T}}x + c_{10}} + \xi^2(c_2^{\mathrm{T}}x + c_{20}) \\ \text{s.t.} \quad & x \in X,\ \xi \geq 0, \end{aligned} \tag{3.3}$$

where ξ is an auxiliary variable. For fixed $x = \bar{x}$, (3.3) is a convex quadratic problem with respect to the single variable ξ and the optimal value is attained at the point $\xi = \frac{\sqrt{c_1^{\mathrm{T}}\bar{x}+c_{10}}}{c_2^{\mathrm{T}}\bar{x}+c_{20}}$. The relation between (3.2) and (3.3) is given by Proposition 1.

Proposition 1. *[Theorem 4.3, Konno and Kuno, 1990] Let (x^*, ξ^*) be an optimal solution of (3.3). Then x^* is an optimal solution of (3.2).*

Next, with a fixed $\xi \geq 0$, we define $P(\xi)$ as the optimal value of the following problem

$$\begin{aligned} \min_{x} \quad & F(x,\xi) =: g(x) - 2\xi\sqrt{c_1^{\mathrm{T}}x + c_{10}} + \xi^2(c_2^{\mathrm{T}}x + c_{20}) \\ \text{s.t.} \quad & x \in X. \end{aligned} \tag{3.4}$$

It is evident that $F(\cdot, \xi)$ is a convex function due to the concavity of $\sqrt{c_1^{\mathrm{T}}x + c_{10}}$. According to Proposition 1, the optimal solution of (3.2) can be obtained by solving (3.4) with $\xi = \xi^*$, where ξ^* is a non-negative value such that $\mathrm{P}(\xi^*) \leq \mathrm{P}(\xi)$ holds for all $\xi \geq 0$. Thus, the main idea of the parametric algorithm is to solve (3.4) over all $\xi \geq 0$ and the one with the smallest optimal value produces an optimal solution of (3.2). However, it is impossible to compute (3.4) when $\xi \to \infty$. As a matter of fact, from the point of parametric programming, all the problems with a sufficiently large ξ, say $\xi \geq \xi_{max}$, share the same optimal solution (we prove it in Proposition 2), say x^*_{max}. Therefore, we only need to focus on $[0, \xi_{max}]$ instead of $[0, \infty)$. In the rest of this subsection, we introduce a method to locate ξ_{max} and the first stage is to find the x^*_{max}.

Proposition 2. *There exists $\xi_{max} \in \mathbb{R}$, such that x^*_{max} is an optimal solution of $P(\xi)$ for any $\xi \geq \xi_{max}$, where $x^*_{max} \in S^* := \arg\min\{g(x) : x \in S^*_1\}$ and*

$$S^*_1 := \arg\max\{c_1^{\mathrm{T}} x : x \in S^*_2\},\ S^*_2 := \arg\min\{c_2^{\mathrm{T}} x : x \in X\}.$$

If the optimal solution set S^*_2 is a singleton, then $x^*_{max} = \arg\min\{c_2^{\mathrm{T}} x : x \in X\}$.

Proof. We only need to prove: there exists $\xi_{max} \in \mathbb{R}$, for any $\xi \geq \xi_{max}$ and $x \in X \backslash S^*$, we have $F(x,\xi) \geq F(x^*_{max},\xi)$. To this end, $x \in X\backslash S^*$ is clarified into three cases: **1**). $x \in S^*_1 \backslash S^*$; **2**). $x \in S^*_2 \backslash S^*_1$; **3**). $x \in X \backslash S^*_2$. Note

$$F(x,\xi) - F(x^*_{max},\xi) = \gamma_1 \xi^2 - 2\gamma_2 \xi + \gamma_3,$$

where $\gamma_1(x) := c_2^{\mathrm{T}} x - c_2^{\mathrm{T}} x^*_{max}$, $\gamma_2(x) := \sqrt{c_1^{\mathrm{T}} x + c_{10}} - \sqrt{c_1^{\mathrm{T}} x^*_{max} + c_{10}}$, and $\gamma_3(x) := g(x) - g(x^*_{max})$.

In the first case, we have $\gamma_1(x) = \gamma_2(x) = 0$ and $\gamma_3(x) > 0$. Then $F(x,\xi) \geq F(x^*_{max},\xi)$ holds naturally for all $\xi_{max} \in \mathbb{R}$. In the second case, we have $\gamma_1(x) = 0$ and $\gamma_2(x) < 0$. Then the existence of ξ_{max} satisfying $F(x,\xi) \geq F(x^*_{max},\xi)$ holds again in that $\gamma_3(x)$ is bounded. In the third case, we have $\gamma_1(x) > 0$ and $\frac{F(x,\xi) - F(x^*_{max},\xi)}{\gamma_1(x)} = (\xi - \frac{\gamma_2(x)}{\gamma_1(x)})^2 + \frac{\gamma_3(x)}{\gamma_1(x)} - \frac{\gamma_2(x)^2}{\gamma_1(x)^2}$. Thus by the boundedness of $\frac{\gamma_2(x)}{\gamma_1(x)}$ and $\frac{\gamma_3(x)}{\gamma_1(x)} - \frac{\gamma_2(x)^2}{\gamma_1(x)^2}$ and $\gamma_1(x) > 0$, there exists ξ_{max} such that $F(x,\xi) \geq F(x^*_{max},\xi)$, for all $x \in X \backslash S^*_2$ and $\xi \geq \xi_{max}$. In combining all the three cases, we have that

$$\exists\, \xi_{max} \in \mathbb{R} \ \text{ s.t. } \ F(x,\xi) \geq F(x^*_{max},\xi),\ \forall x \in X \text{ and } \xi \geq \xi_{max}.$$

The proof is completed. □

We have to point out that, although Konno and Kuno (1990) gave an approach to finding x^*_{max} of a generalized linear multiplicative programming, their criteria may fail when the related problem has more than one solution.

Now, let's introduce how to locate ξ_{max} by the use of x^*_{max}. Since x^*_{max} is a global optimal solution to (3.4) with $\xi \geq \xi^*_{max}$ and linearity constraint qualification (LCQ) holds, then there exist multipliers $\lambda \geq 0$ and μ, such that

$$\begin{cases} 0 \in \partial g(x^*_{max}) - \dfrac{c_1}{\sqrt{c_1^{\mathrm{T}} x^*_{max} + c_{10}}} \xi + c_2 \xi^2 - A_1^{\mathrm{T}} \lambda - A_2^{\mathrm{T}} \mu \\ \lambda^{\mathrm{T}} (A_1 x^*_{max} - b_1) = 0. \end{cases} \tag{3.5}$$

Let $\bar{\lambda}$ and $\bar{A}_1$ be the sub-vector and the sub-matrix of λ and A_1 corresponding to the active constraints of $A_1 x^*_{max} \geq b_1$. Then (3.5) becomes

$$A_0^{\mathrm{T}} \nu \in \partial g(x^*_{max}) - \frac{c_1}{\sqrt{c_1^{\mathrm{T}} x^*_{max} + c_{10}}} \xi + c_2 \xi^2,$$

where $A_0 = \begin{bmatrix} \bar{A}_1 \\ A_2 \end{bmatrix}$, $\nu = \begin{bmatrix} \bar{\lambda} \\ \mu \end{bmatrix}$. Naturally, ξ_{max} can be estimated via the following system

$$\begin{cases} A_0^{\mathrm{T}} \nu \in \partial g(x^*_{max}) - \dfrac{c_1}{\sqrt{c_1^{\mathrm{T}} x^*_{max} + c_{10}}} \xi + c_2 \xi^2 \\ \bar{\lambda} \geq 0, \end{cases} \tag{3.6}$$

Solving (3.6) can be very expensive and complicated, especially for large-scale problems. But fortunately, when A_0 is a matrix of full-rank square, the process can be much simplified. Under this assumption, (3.6) can be rewritten as

$$\begin{cases} \nu \in Q_0 - q_1 \xi + q_2 \xi^2 \\ \bar{\lambda} \geq 0, \end{cases}$$

where $Q_0 := \{(A_0^{\mathrm{T}})^{-1}\} \times \partial g(x^*_{max})$, $q_1 := \frac{(A_0^{\mathrm{T}})^{-1} c_1}{\sqrt{c_1^{\mathrm{T}} x^*_{max} + c_{10}}}$, and $q_2 := (A_0^{\mathrm{T}})^{-1} c_2$. Note that Q_0 can be viewed as a vector, whose elements are sets rather than numbers. Let $Q_0^{(\lambda)}, q_1^{(\lambda)}$, and $q_2^{(\lambda)}$ be the sub-vectors of Q_0, q_1, and q_2 corresponding to λ. Then, the existence of λ implies that

$$q_0^{(\lambda)} - q_1^{(\lambda)} \xi + q_2^{(\lambda)} \xi^2 \geq 0 \quad \text{and} \quad q_0^{(\lambda)} = \max\{Q_0^{(\lambda)}\}, \tag{3.7}$$

where $q_0^{(\lambda)} = \max\{Q_0^{(\lambda)}\}$ means $q_0^{(\lambda)}$ is a vector consisting of the maximum of each set in $Q_0^{(\lambda)}$. Since the existence of λ is certain, we have $q_2^{(\lambda)} \geq 0$. That is, the solution of (3.7) can be derived explicitly. Hence, the computation of (3.6) becomes convenient. Here are remarks for this algorithm.

(i) Konno and Kuno's (1990) directly utilized generalized inverse to solve (3.6). However, this method may lose information and lead to a wrong ξ_{max} when A_0 is not a matrix of full-rank square.

(ii) The assumption that A_0 is a matrix of full-rank square is not strict. For instance, considering x^*_{max} is obtained by solving linear programming problems (c.f. Proposition 2), then the assumption is satisfied when S_2^* is a singleton and Linear independence constraint qualification (LICQ) holds at x^*_{max}.

(iii) Every step of deriving ξ_{max} is sufficient and necessary, thus ξ_{max} located by the above process is exact for the problem.

Eventually, we conclude the parametric algorithm in 2 steps.

Step 1. Find x^*_{max} through solving optimization problems in Proposition 2 (see S^*, S^*_1, and S^*_2) and ξ_{max} through solving system (3.6) or (3.7);

Step 2. If $\xi_{max} \leq 0$, then x^*_{max} is the global solution of the model (3.2); otherwise, solve (3.4) over $\xi \in [0, \xi_{max}]$, then the solution x^* corresponding to the smallest optimal value is the global solution of the problem (3.2).

A basic method to search for the minimal optimal value of (3.4) over $[0, \xi_{max}]$ is the discretization of the interval. More precisely, we divide the interval $[0, \xi_{max}]$ into many subdivisions and compute $P(\xi)$ at every endpoint. Then, when the subdivisions are narrow enough, the resulting solution would be sufficiently close to the solution x^*.

4 Numerical Experiments

In this section, we examine performances of the l_1-sparse minimax model, the $l_{\frac{1}{2}}$-sparse minimax model and the l_1-sparse minimax Sharpe ratio model by using the weekly historical data of 1200 stocks from Hang Seng Index, Shanghai Securities Composite Index, and NASDAQ Index (400 stocks from each), during the period from January 1, 2005 to December 31, 2019. The rate of return y_{jt} is derived by $y_{jt} = (p_{j,t+1} - p_{jt})/p_{jt}$, where p_{jt} represents the price of stock j in week t.

We translate the l_1-sparse minimax model into a smooth formulation and use the optimization toolbox (function 'linprog') in Matlab to solve the equivalent problem. In virtue of slackness variables $u_j(= |w|_j)$, the problem (2.1) can be equivalently transformed as a linear programming problem

$$\begin{aligned}
\min_{w\in\mathcal{F}, M_p\in\mathbb{R}} \quad & -M_p + \tau\sum_{j=1}^{N} u_j \\
\text{s.t.} \quad & -u \leq w \leq u \\
& \sum_{j=1}^{N} y_{jt}w_j \geq M_p, t = 1, \ldots, T.
\end{aligned}$$

Notably, this is a parametric linear programming problem concerning parameter τ. Referring to Berkelaar et al. (1997), there exists a finite set of breakpoints $0 \leq \tau_0 < \tau_1 <$

$\ldots < \tau_K < \infty$, such that the optimal solution set keeps unchanged on any (open) interval between two successive breakpoints, which is consistent with figures in Example 2. The computation of the l_1-sparse mean-variance model is completed by CVX toolbox (Grant and Boyd, 2020). Iterative reweighted minimization method (Lu, 2014) is utilized to compute the $l_{\frac{1}{2}}$-sparse minimax model and the $l_{\frac{1}{2}}$-sparse mean-variance model. And the l_1-sparse minimax Sharpe ratio model is solved by the parametric algorithm proposed in section 3.2. Initially, we test the computational time of five models with fixed τ. For reliability, we target τ corresponding to 12-14 active stocks for each model. The results are listed in Table 1.

Table 1. Computational time

l_1-MM	$l_{\frac{1}{2}}$-MM	l_1-MV	$l_{\frac{1}{2}}$-MV	l_1-SR
0.56s	2.90s	24.35s	85.91s	225.33s

In Examples 1 and 2, we set the required rate of return, G, to be the average rate of return of all the stocks. Each time period is taken as 1 week and the number of periods is set as $T = 11$. The lower bound α of the portfolio is fixed at -0.2, which means that the amount of short selling for each stock is limited under 20%.

Example 1. In the first experiment, we test the out-of-sample return of the l_1-sparse minimax rule under different values of τ and compare it with that of the equal-weighted portfolio, which has been shown to outperform many portfolio selection rules (DeMiguel et al., 2007). As mentioned above, the number of time periods is $T = 11$. That means, for the current period, data from the previous 11 periods are utilized to determine parameters y_{jt}, $\bar{y}_j$, and G. An optimal portfolio is then obtained by solving the l_1-sparse minimax model (2.1). The out-of-sample rate of return is computed using the obtained optimal portfolio and the rate of return of the following period. For example, the first out-of-sample rate of return is computed using the rate of return of period 12, and the same procedure is repeated for the sequential periods.

Figure 2 plots the out-of-sample rates of return of the equal-weighted portfolio and the l_1-sparse minimax model with $\tau = 0.06$, 0.07, and 0.15, respectively. We observe that there are many similarities between the four curves in terms of the trend. Specifically, their rates of return increase or decrease at the same time in most periods. For the rates of return of the l_1-sparse minimax model, we observe that a small τ leads to evident fluctuations while a larger τ produces fewer variations. This tendency is partly due to the relationship between short selling and the value of τ (see (2.3)). And according to Luenberger (1997), short selling is considered to be quite risky, and thus causes fluctuations.

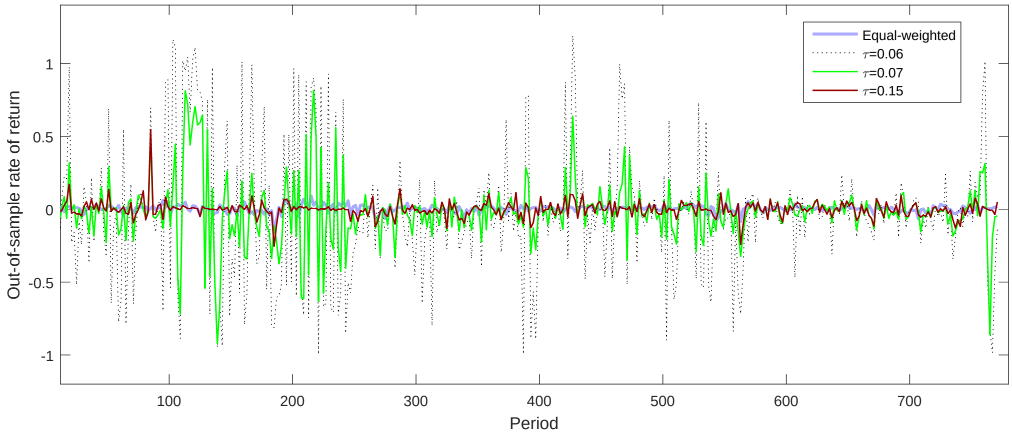

Figure 2. Rates of return with different τ

Example 2. In this experiment, we select 37 blue chips from Hang Seng Index to illustrate the variational tendency of sparsity and short selling of the l_1-sparse minimax model, the $l_{\frac{1}{2}}$-sparse minimax model, and the l_1-sparse minimax Sharpe ratio model together with two benchmark models, the l_1-sparse mean-variance model and the $l_{\frac{1}{2}}$-sparse mean-variance model. For this purpose, we obverse their sparsity and short selling with τ going through $0 - 0.05$. To see the influence of the parameter α, we conduct experiments with $\alpha = -0.2$ and $\alpha = -0.5$, respectively. Figure 3(a) and Figure 4(a) show that, for all the five models, the number of nonzero stocks in the optimal portfolio (i.e. the sparsity) decreases as the value of τ increases, which can be explained by (2.2). The monotonicity of the l_1-sparse minimax Sharpe ratio model is less exact compared with those of the other two l_1-sparse models. As the value of τ increases, the curve representing $l_{\frac{1}{2}}$-sparse minimax (or mean-variance) model reduces more dramatically than that of the l_1-sparse minimax (or mean-variance) model. The monotonic property is quite applicable and plays a critical part in the subsequent example. More precisely, we can target optimal portfolios in which the number of selected stocks is required within a specific range by taking the value of τ over a finite and smaller interval.

Figure 3(b) and Figure 4(b) demonstrate a similar monotonic trend for short selling, which coincides with (2.3). Although an analogous inequality to (2.3) is not obtained for the $l_{\frac{1}{2}}$ model, the relation $\|x\|_p^p = \|x^+\|_p^p + \|x^-\|_p^p$ $(0 < p \leq 1)$ also partly explains the coincident tendency between the level of sparsity and short selling. According to Figure 3(a) and 3(b) (or Figure 4(a) and Figure 4(b)), a more sparse portfolio, at the same time, is a portfolio with a smaller number of negative-weighted stocks and a quite sparse portfolio may not include any short positions, just as what we mentioned in remark (ii) in Section 2. Comparing Figure 3 and Figure 4, we get to know that the selection of α does not influence the decreasing tendencies and an extremely sparse optimal portfolio always can be attained with different α. The only difference is that the extremely sparse optimal portfolio is obtained at a smaller τ for the model with a larger α. It is also noteworthy that in all the figures, graphs are piecewise constant due to the parametric construction of sparse models; see related analysis at the beginning of this section.

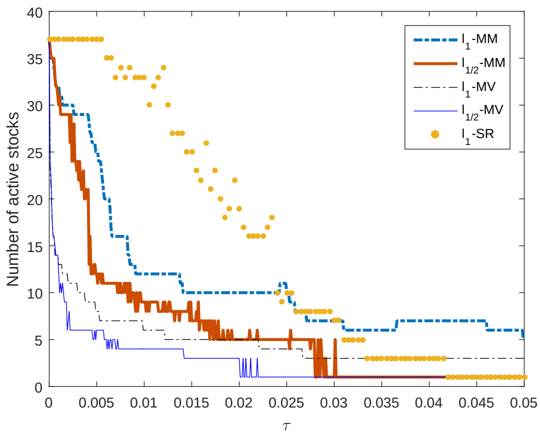

Figure 3(a). Number of active stocks, $\alpha = -0.2$

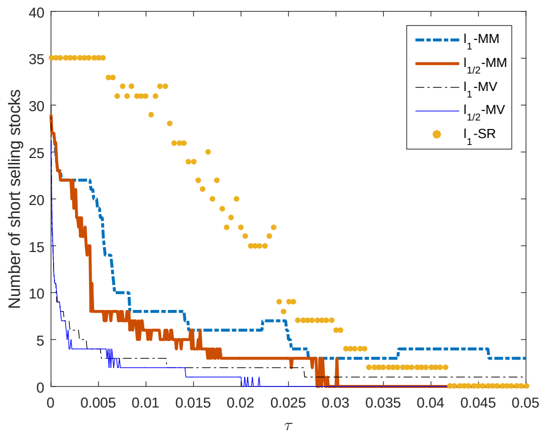

Figure 3(b). Number of short selling stocks, $\alpha = -0.2$

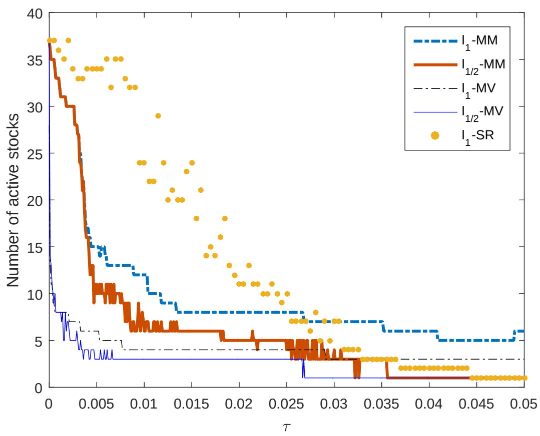

Figure 4(a). Number of active stocks, $\alpha = -0.5$

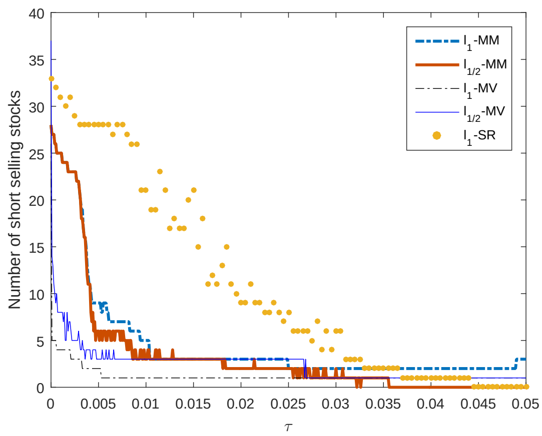

Figure 4(b). Number of short selling stocks, $\alpha = -0.5$

In fact, the 1200-stock problem shares the same monotonic trend. The only thing that is changed is that, for the 1200-stock problem, the sparse minimax models vary in a larger interval, say $0 - 0.07$; while sparse mean-variance models vary in a much smaller interval, say $0 - 10^{-8}$. This is attributed to the different orders of magnitude of the objective functions. In fact, for the 37-stock problem, the orders of minimax and mean-variance models are both -1. However, for the 1200-stock problem, the orders of magnitude are 1 and -11, respectively.

In the next example, we repeat the process mentioned in Example 1 over five observation periods and compare five different sparse models. The required rate of return G is taken to be the maximal rate of return of stocks. The level of short selling α and the number of periods T remain unchanged while the out-of-sample observation period is reset as 11 weeks. For example, data from period 688 to period 698 are used to determine the optimal weights, and then we use them to compute the out-of-sample rate of return of period $699 - 709$.

In fact, there is no comparability among different sparse models with the same regularization parameter τ, since the same τ in different sparse models generally corresponds to different levels of sparsity. Apparently, a more practical method is comparing them at the same level of sparsity. The comparisons in this example are all completed under this consideration.

Example 3. This experiment examines the out-of-sample performances of the l_1-sparse minimax model, the $l_{\frac{1}{2}}$-sparse minimax model, and the l_1-sparse minimax Sharpe ratio model under five levels of sparsity, from period $699 - 709$ to period $703 - 713$ (see the last five columns of Tables 2(a) to 2(e)). The l_1-sparse mean-variance model (Brodie et al., 2009) and the $l_{\frac{1}{2}}$-sparse mean-variance model are considered to be benchmarks in this example. The level of sparsity, say $11 - 20$, means the number of active stocks is between 11 and 20 over all the five periods. This can be done by adjusting the value of τ (see Example 2). However, in general, more than one portfolios fall into this level of sparsity. In this situation, the smallest risk of these portfolios and its corresponding rate of return, Sharpe ratio, and the number of short selling stocks are considered. If the portfolio with minimal risk is still not unique, we select one with a maximal rate of return.

In Tables 2(a) to 2(e), R, $Risk_{MM}/Risk_{MV}$, SR, and S represent the out-of-sample rate of return, the out-of-sample risk, the out-of-sample Sharpe ratio, and the number of short selling stocks. Related result of the equal-weighted rule is also presented in all tables for reference. It is apparent that the equal-weighted rule outperforms all the sparse models in terms of the Sharpe ratio due to its extremely small risk. However, on the contrary, the rates of return of five sparse models are more favorable than those of the equal-weighted strategy. From Tables 2(a), 2(b), and 2(e), we observe that for all the three sparse minimax models, a more sparse optimal portfolio tends to have a lower risk and a lower rate of return. However, the changes of the risk and the rate of return are not so significant for the l_1-sparse and $l_{\frac{1}{2}}$-sparse mean-variance models.

Table 2(a). l_1-sparse minimax model

l_1-MM	Equal Weight				11-20				21-30				31-40				41-50				51-60			
	R	Risk$_{MM}$	SR$_{MM}$	S	R	Risk$_{MM}$	SR$_{MM}$	S	R	Risk$_{MM}$	SR$_{MM}$	S	R	Risk$_{MM}$	SR$_{MM}$	S	R	Risk$_{MM}$	SR$_{MM}$	S	R	Risk$_{MM}$	SR$_{MM}$	S
period 724-734	0.007	0.032	0.218	0	0.012	0.112	0.109	10	0.033	0.248	0.133	21	0.034	0.443	0.076	31	0.046	0.511	0.090	40	0.072	0.616	0.118	50
period 725-735	0.009	0.032	0.270	0	0.019	0.271	0.070	10	0.042	0.383	0.111	20	0.033	0.453	0.073	29	0.058	0.502	0.116	39	0.077	0.668	0.116	50
period 726-736	0.008	0.032	0.237	0	0.097	0.325	0.298	12	0.150	0.424	0.353	21	0.169	0.612	0.277	31	0.238	0.713	0.334	40	0.286	0.949	0.301	51
period 727-737	0.006	0.032	0.199	0	-0.004	0.307	-0.014	11	0.003	0.633	0.005	22	-0.002	0.974	-0.002	31	0.010	1.086	0.010	42	0.024	1.237	0.019	51
period 728-738	0.004	0.029	0.150	0	0.022	0.179	0.123	9	0.099	0.252	0.393	21	0.127	0.387	0.329	31	0.155	0.525	0.296	40	0.196	0.677	0.289	51

Table 2(b). $l_{\frac{1}{2}}$-sparse minimax model

$l_{1/2}$-MM	Equal Weight				11-20				21-30				31-40				41-50				51-60			
	R	Risk$_{MM}$	SR$_{MM}$	S	R	Risk$_{MM}$	SR$_{MM}$	S	R	Risk$_{MM}$	SR$_{MM}$	S	R	Risk$_{MM}$	SR$_{MM}$	S	R	Risk$_{MM}$	SR$_{MM}$	S	R	Risk$_{MM}$	SR$_{MM}$	S
period 724-734	0.007	0.032	0.218	0	0.030	0.235	0.128	13	0.049	0.332	0.148	23	0.057	0.424	0.134	36	0.065	0.598	0.108	42	0.072	0.687	0.105	48
period 725-735	0.009	0.032	0.270	0	0.050	0.476	0.105	14	0.008	0.625	0.013	24	0.033	0.762	0.043	37	0.033	0.835	0.039	42	0.044	0.967	0.046	53
period 726-736	0.008	0.032	0.237	0	0.067	0.337	0.198	11	0.142	0.506	0.280	24	0.201	0.736	0.273	34	0.278	0.942	0.295	44	0.299	1.131	0.264	51
period 727-737	0.006	0.032	0.199	0	0.051	3.182	0.016	11	0.758	6.499	0.117	16	-0.009	0.418	-0.022	31	-0.010	1.259	-0.008	44	0.016	1.817	0.009	50
period 728-738	0.004	0.029	0.150	0	0.037	0.098	0.384	5	0.104	0.381	0.273	23	0.155	0.411	0.377	32	0.194	0.603	0.321	43	0.235	0.710	0.331	53

Table 2(c). l_1-sparse mean-variance model

l_1-MV	Equal Weight				11-20				21-30				31-40				41-50				51-60			
	R	Risk$_{MV}$	SR$_{MV}$	S	R	Risk$_{MV}$	SR$_{MV}$	S	R	Risk$_{MV}$	SR$_{MV}$	S	R	Risk$_{MV}$	SR$_{MV}$	S	R	Risk$_{MV}$	SR$_{MV}$	S	R	Risk$_{MV}$	SR$_{MV}$	S
period 724-734	0.007	0.000	31.384	0	0.014	0.004	3.177	7	0.013	0.004	3.156	11	0.014	0.004	3.259	14	0.014	0.004	3.393	18	0.015	0.004	3.653	27
period 725-735	0.009	0.000	32.612	0	0.012	0.009	1.442	6	0.011	0.008	1.362	7	0.010	0.008	1.288	16	0.009	0.007	1.193	24	0.007	0.007	1.040	35
period 726-736	0.008	0.000	25.308	0	0.021	0.027	0.809	5	0.025	0.026	0.963	7	0.032	0.027	1.215	15	0.029	0.023	1.249	23	0.027	0.022	1.226	27
period 727-737	0.006	0.000	21.137	0	0.011	0.008	1.337	7	0.012	0.008	1.404	9	0.013	0.008	1.531	12	0.010	0.008	1.288	14	0.019	0.010	1.909	23
period 728-738	0.004	0.000	18.594	0	0.006	0.006	0.995	6	0.005	0.005	0.916	9	0.004	0.006	0.737	10	0.005	0.006	0.843	14	0.005	0.006	0.843	21

Table 2(d). $l_{\frac{1}{2}}$-sparse mean-variance model

$l_{1/2}$-MV	Equal Weight				11-20				21-30				31-40				41-50				51-60			
	R	Risk$_{MV}$	SR$_{MV}$	S	R	Risk$_{MV}$	SR$_{MV}$	S	R	Risk$_{MV}$	SR$_{MV}$	S	R	Risk$_{MV}$	SR$_{MV}$	S	R	Risk$_{MV}$	SR$_{MV}$	S	R	Risk$_{MV}$	SR$_{MV}$	S
period 724-734	0.007	0.000	31.384	0	0.033	0.016	2.060	8	0.015	0.006	2.592	12	0.009	0.007	1.341	19	0.030	0.015	1.940	24	0.008	0.007	1.231	36
period 725-735	0.009	0.000	32.612	0	0.006	0.009	0.687	6	0.040	0.034	1.197	17	0.006	0.009	0.692	20	0.006	0.009	0.718	25	0.007	0.009	0.769	32
period 726-736	0.008	0.000	25.308	0	0.055	0.043	1.263	7	0.054	0.043	1.247	13	0.053	0.045	1.162	18	0.052	0.045	1.145	30	0.052	0.045	1.155	32
period 727-737	0.006	0.000	21.137	0	0.024	0.012	2.078	5	0.021	0.011	1.851	12	0.021	0.011	1.867	14	0.021	0.011	1.865	19	0.024	0.012	2.081	23
period 728-738	0.004	0.000	18.594	0	0.018	0.006	2.911	4	0.015	0.006	2.484	10	0.015	0.006	2.441	15	0.008	0.006	1.534	17	0.013	0.006	2.424	26

Table 2(e). l_1-sparse minimax Sharpe ratio model

l_1-SR	Equal Weight				11-20				21-30				31-40				41-50				51-60			
	R	Risk$_{MM}$	SR$_{MM}$	S	R	Risk$_{MM}$	SR$_{MM}$	S	R	Risk$_{MM}$	SR$_{MM}$	S	R	Risk$_{MM}$	SR$_{MM}$	S	R	Risk$_{MM}$	SR$_{MM}$	S	R	Risk$_{MM}$	SR$_{MM}$	S
period 724-734	0.007	0.032	0.218	0	0.026	0.274	0.095	11	0.069	0.533	0.130	20	0.090	0.804	0.112	31	0.113	1.042	0.108	42	0.130	1.199	0.108	50
period 725-735	0.009	0.032	0.270	0	0.023	0.431	0.054	11	0.028	0.563	0.050	20	0.063	0.893	0.071	32	0.084	1.148	-0.073	42	0.097	1.312	0.074	50
period 726-736	0.008	0.032	0.237	0	0.028	0.244	0.113	11	0.032	0.517	0.062	20	0.064	0.771	0.084	30	0.107	1.184	0.091	40	0.151	1.462	0.103	52
period 727-737	0.006	0.032	0.199	0	0.025	0.295	0.083	10	0.046	0.556	0.082	21	0.065	0.791	0.082	30	0.082	1.056	0.077	40	0.082	1.316	0.063	50
period 728-738	0.004	0.029	0.150	0	-0.001	0.306	-0.005	10	0.006	0.552	0.011	21	0.000	0.820	0.000	30	0.000	1.168	0.000	40	0.008	1.563	0.005	52

Next, we compare the l_1-sparse minimax model and the $l_{\frac{1}{2}}$-sparse minimax model. When the level of sparsity is extremely high, i.e., with $11-20$ active stocks, the $l_{\frac{1}{2}}$-sparse minimax rule is more preferable than the l_1-sparse minimax rule, both in the aspect of the rate of return and the Sharpe ratio. The optimal portfolios are less sparse, i.e., with $41-50$ or $51-60$ active stocks, both models perform identically. That is, the $l_{\frac{1}{2}}$-sparse minimax model would be a desirable choice for investors who require extremely sparse portfolios, while the l_1 formulation is more beneficial to those who prefer less sparse portfolio due to its computational simplicity (see Table 1). By comparing the l_1-sparse and the $l_{\frac{1}{2}}$-sparse mean-variance models, we do not observe any superiority of the $l_{\frac{1}{2}}$-sparse mean-variance model. As a whole, their out-of-sample performances appear to be commensurate for all levels of sparsity.

It is remarkable that the Sharpe ratios of the minimax rule and the mean-variance rule are not comparable in that they are based on their own risk, but the values of distinct risk measures are not comparable. Therefore, the only performance measure to compare the sparse minimax rule and the sparse mean-variance rule is the out-of-sample rate of return. From Tables 2(a) and 2(c) (resp. Tables 2(b) and 2(d)), we observe that the optimal portfolios of the l_1-sparse (resp. the $l_{\frac{1}{2}}$-sparse) minimax model tend to achieve higher rates of return than those of the l_1-sparse (resp. the $l_{\frac{1}{2}}$-sparse) mean-variance model. For the l_1-sparse minimax model and the l_1-sparse minimax Sharpe ratio model, Tables 2(a) and 2(e) show that two models perform similarly. Although the computation of the l_1-sparse minimax model is much easier, the l_1-sparse minimax Sharpe ratio model would be a good choice for investors who do not have a desired expected return in advance.

We also conduct the above experiment with $\alpha = -0.02, -0.05$ and -0.5, respectively. With a larger level of short selling, we observe that higher rates of return and Sharpe ratios are obtained for all the models and that the risks of three sparse minimax models increase. However, the risks of two sparse mean-variance models are stable with different values of τ. The results of period $702-712$ with 11 to 20 active stocks are listed in Table 3 as a representative.

Table 3. Performances with different α

period 702-712 (level 11-20)	α=-0.02				α=-0.05				α=-0.2				α=-0.5			
	R	$Risk_{MM}$	SR_{MM}	S	R	$Risk_{MM}$	SR_{MM}	S	R	$Risk_{MM}$	SR_{MM}	S	R	$Risk_{MM}$	SR_{MM}	S
l_1-MM	0.003	0.130	0.020	14	0.010	0.130	0.074	11	-0.004	0.307	-0.014	11	0.103	0.505	0.205	13
$l_{1/2}$-MM	-0.004	0.122	-0.036	10	0.016	0.196	0.079	15	0.051	3.182	0.016	11	0.023	0.439	0.052	7
l_1-MV	0.013	0.009	1.401	14	0.020	0.012	1.623	11	0.011	0.008	1.337	7	0.019	0.010	1.971	5
$l_{1/2}$-MV	-0.007	0.013	-0.561	16	0.019	0.011	1.727	13	0.024	0.012	2.078	5	0.035	0.015	2.305	5
l_1-SR	0.004	0.141	0.028	14	0.010	0.204	0.049	17	0.025	0.295	0.083	10	0.079	0.795	0.100	12

5 Conclusion

In this article, we considered the l_1-sparse and l_p-sparse ($0 < p < 1$) linear portfolio models and took the minimax selection rule (Young, 1998) as representative to discover their properties and numerical performances. On the other hand, we constructed the l_1-sparse minimax Sharpe ratio model based on a modified minimax risk measure. To overcome the computational difficulty of the l_1-sparse minimax Sharpe ratio model, we extended the parametric algorithm in Konno and Kuno (1990) to a more general framework. In numerical experiments, we found all sparse minimax models are efficient for promoting the sparsity of the optimal portfolios. The $l_{\frac{1}{2}}$-sparse minimax model is advantageous when the investor requires an extremely sparse portfolio; while the l_1-sparse minimax model is favorable for the investment with less requirement of sparsity. For the l_1-sparse minimax Sharpe ratio model, it is preferred when the desired expected return is not given in advance.

References

Bacon, C. R. (2008). *Practical Portfolio Performance Measurement and Attribution*. New York: John Wiley & Sons.

Benninga, S. (2014). *Financial Modeling*. Cambridge: The MIT Press.

Berkelaar, A. B., Roos, K., & Terlaky, T. (1997). The optimal set and optimal partition approach to linear and quadratic programmin. *In Advances in Sensitivity Analysis and Parametic Programmingg* (pp. 159-202). Boston: Springer.

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., & Loris, I. (2009). Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences of The United States of America*, 106(30), 12267-12272.

Burke, G. (1994). A sharper Sharpe ratio. *Futures*, 23(3), 56.

Cai, X., Teo, K. L., Yang, X., & Zhou, X. Y. (2000). Portfolio optimization under a minimax rule. *Management Science*, 46(7), 957-972.

Chang, T. J., Meade, N., Beasley, J. E., & Sharaiha, Y. M. (2000). Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research*, 27(13), 1271-1302.

Chartrand, R. (2007). Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Processing Letters*, 14(10), 707-710.

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

Chen, S. S., Donoho, D. L., & Saunders, M. A. (2001). Atomic decomposition by basis pursuit. *SIAM Review*, 43(1), 129-159.

Dai, Z., & Wen, F. (2018). A generalized approach to sparse and stable portfolio optimization problem. *Journal of Industrial & Management Optimization*, 14(4), 1651-1666.

Das, S., Markowitz, H., Scheid, J., & Statman, M. (2011). Portfolios for investors who want to reach their goals while staying on the mean-variance efficient frontier. *The Journal of Wealth Management*, 14(2), 25-31.

Daubechies, I., Defrise, M., & De Mol, C. (2004). An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11), 1413-1457.

DeMiguel, V., Garlappi, L., Nogales, F. J., & Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5), 798-812.

DeMiguel, V., Garlappi, L., & Uppal, R. (2007). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22(5), 1915-1953.

Elton, E. J., Gruber, M. J., Brown, S. J., & Goetzmann, W. N. (2014). *Modern Portfolio Theory and Investment Analysis*. New York: John Wiley & Sons.

Fastrich, B., Paterlini, S., & Winker, P. (2015). Constructing optimal sparse portfolios using regularization methods. *Computational Management Science*, 12(3), 417-434.

Grant, M., & Boyd, S. (2020). *CVX: Matlab software for disciplined convex programming, version 2.2*. http://cvxr.com/cvx.

Goh, C. J., & Yang, X. (1996). Analytic efficient solution set for multi-criteria quadratic programs. *European Journal of Operational Research*, 92(1), 166-181.

Hu, Y., Li, C., Meng, K., Qin, J., & Yang, X. (2017). Group sparse optimization via $l_{p,q}$ regularization. *The Journal of Machine Learning Research*, 18(1), 960-1011.

Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4), 1651-1683.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945-1964. *The Journal of Finance*, 23(2), 389-416.

Konno, H., & Kuno, T. (1990). Generalized linear multiplicative and fractional programming. *Annals of Operations Research*, 25(1), 147-161.

Konno, H., & Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. *Management Science*, 37(5), 519-531.

Lu, Z. (2014). Iterative reweighted minimization methods for l_p regularized unconstrained nonlinear programming. *Mathematical Programming*, 147(1), 277-307.

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

Markowitz, H. M. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77-91.

Markowitz, H. M. (2013). How to Represent Mark-to-Market Possibilities with the General Portfolio Selection Model. *Journal of Portfolio Management*, 39(4), 1-3.

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

Markowitz, H. M., & Van Dijk, E. L. (2003). Single-Period MeanVariance Analysis in a Changing World. *Financial Analysts Journal*, 59(2), 30-44.

Matousek, J., & Gartner, B. (2007). *Understanding and Using Linear Programming*. Berlin/Heidelberg: Springer Berlin/Heidelberg.

McCafferty, T. (2003). *The Market is Always Right*. New York: McGraw Hill.

Niedermayer, A., & Niedermayer, D. (2010). Applying Markowitzs critical line algorithm. *Handbook of Portfolio Construction* edited by Guerard, B. J., Jr. (pp. 383-400). Boston: Springer.

Qi, Y. (2020). Parametrically computing efficient frontiers of portfolio selection and reporting and utilizing the piecewise-segment structure. *Journal of the Operational Research Society*, 71(10), 1675-1690.

Qi, Y., Zhang, Y., & Ma, S. (2019). Parametrically computing efficient frontiers and re-analyzing efficiency-diversification discrepancies and naive diversification. *INFOR: Information Systems and Operational Research*, 57(3), 430-453.

Sharpe, W.F. (1966). Mutual fund Performance. *Journal of Business*, 39(1), 119-138.

Sharpe, W. F. (1967). A linear programming algorithm for mutual fund portfolio selection. *Management Science*, 13(7), 499-510.

Sortino, F. A., & Van Der Meer, R. (1991). Downside risk. *The Journal of Portfolio Management*, 17(4), 27-31.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.

Treynor, J. L. (1965). How to rate management of investment funds. *Harvard Business Review*, 43(1), 63-75.

Woodside-Oriakhi, M., Lucas, C., & Beasley, J. E. (2011). Heuristic algorithms for the cardinality constrained efficient frontier. *European Journal of Operational Research*, 213(3), 538-550.

Xu, Z., Zhang, H., Wang, Y., Chang, X., & Liang, Y. (2010). $L_{1/2}$ regularization. *Science China Information Sciences*, 53(6), 1159-1169.

Young, M. R. (1998). A minimax portfolio selection rule with linear programming solution. *Management Science*, 44(5), 673-683.

Young, T.W. (1991). Calmar ratio: A smoother tool. *Futures*, 20(1), 40.