

## SPARSE MINIMAX PORTFOLIO AND SHARPE RATIO MODELS

CHENCHEN ZU\* AND XIAOQI YANG

Department of Applied Mathematics  
The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

CARISA KWOK WAI YU

Department of Mathematics, Statistics and Insurance  
The Hang Seng University of Hong Kong, Shatin, Hong Kong, China

(Communicated by Jie Sun)

**ABSTRACT.** In this paper, we investigate sparse portfolio selection models with a regularized  $l_p$ -norm term ( $0 < p \leq 1$ ) and negatively bounded 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 negatively bounded 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.

**1. Introduction.** In 1952, Markowitz [28] formulated the portfolio selection problem as a quadratic programming problem, 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

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

---

2020 *Mathematics Subject Classification.* Primary: 91G10, 90C90, 90-10.

*Key words and phrases.* Sparse minimax portfolio selection model, Sharpe ratio, short selling, sparse mean-variance model,  $l_p$  regularization.

\* Corresponding author: Chenchen Zu.

$$\sum_{j=1}^N w_j = 1,$$

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 (see [12, 27, 30]). Many competent algorithms are developed to solve the mean-variance model with parametric formulation, typically the critical line algorithm (see [29]), which has been efficiently used to investigate large-scale portfolio problems (see [33, 34]). The active set method for multi-criteria convex quadratic programming problems (see [18]) 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 [36] viewed the market responsiveness as the risk measure and built a linear approximation of the mean-variance model. After that, mean absolute deviation [24], minimum return [43], and  $l_\infty$  function [6] 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 [7, 41]). 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

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

which modifies the model (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  $\tau$  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 [4, 11, 14]). The  $l_1$  regularizer has been widely adopted to seek sparsity for industrial problems, such as image reconstruction, data analysis, and machine learning (see [9, 13, 39]). When  $0 < p < 1$ , Chen et al. [10] and Fastrich et al. [17] claimed that sparse optimal portfolios can be obtained by solving the  $l_p$ -sparse mean-variance model. In addition, it was shown by Chartrand [8], Xu et al. [42], and Hu et al. [20] 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 [21] obtained an optimal portfolio including only 24.1 stocks out of 500 stocks. Qi et al. [35] applied their portfolio selection model to 1800-stock problems, and the minimal number of selected stocks can be 62.11. Woodside-Oriakhi et al. [41] 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 paper, we take the minimax rule [43] (see more details in Section 2) for instance to discover the linear portfolio models regularized by the  $l_p$  ( $0 < p \leq 1$ ) 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. [42] and Hu et al. [20] 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 the 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 that, 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 consider the  $l_1$ -sparse Sharpe ratio model based on the minimax rule. In the area of performance assessment, the Treynor index [40], Sharpe index [37], and Jensen index [22] are the three most popular performance measures to rank performances of portfolios or mutual fund managers. The *Sharpe index*, also known as the *Sharpe ratio*, was first put forward by Sharpe [37] 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  $\sigma(R)$  represent the expected value of return and the standard deviation of return, which exactly are the return and risk in the mean-variance model. And  $R_f$  stands for the risk-free return. On the basis of the 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), e.g., Calmar ratio [44], Sortino ratio [38], Burke ratio [5], and Sterling ratio [32]. Apart from ranking performances, the Sharpe ratio is also employed as an objective function of the portfolio optimization model (see [2, 16]). Likewise, we generalize the Sharpe ratio maximization model based on the risk measure of the minimax model [43] (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 model, we modify the minimax risk measure by adding a constant to keep the denominator positive. To find a global solution of the (generalized)  $l_1$ -sparse minimax Sharpe ratio model, we develop a parametric algorithm, which extends the algorithm proposed by Konno and Kuno [23].

It's worth noting that the nonnegative return is an underlying assumption for the classical Sharpe ratio. Bacon [1] pointed out that the negative return makes the Sharpe ratio difficult to interpret. In fact, a larger 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 preferable when the return is negative. The nonnegativity of 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 adopt the minimax risk measure as the denominator directly. To overcome this difficulty, we propose a revised minimax risk measure  $\lambda - M_p$  to replace the original  $-M_p$  (see (6)), where  $\lambda$  is a parameter such that  $\lambda - M_p > 0$ . The function of 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 paper are summarized as follows.

- (i) We construct the  $l_p$ -sparse ( $0 < p \leq 1$ ) 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 optimization algorithm to solve the (generalized) non-convex  $l_1$ -sparse minimax Sharpe ratio model.

The rest of the paper is organized as follows. In Sections 2 and 3, we construct and analyze the  $l_p$ -sparse ( $0 < p \leq 1$ ) minimax models and the  $l_1$ -sparse minimax Sharpe ratio model, then follow numerical experiments in Section 4. Eventually, the paper 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  $w_j$  ( $j = 1, 2, \dots, N$ ), 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 model 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  $\alpha$  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 [43], maximizing the minimum return of the portfolio over all the time periods, is given as

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

which is equivalent to

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

In general, from the theory of linear programming (see [31]), 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

$$\begin{aligned} \min_{w \in \mathcal{F}, M_p} \quad & -M_p + \tau \|w\|_p^p \\ \text{s.t.} \quad & \sum_{j=1}^N y_{jt} w_j \geq M_p, \quad t = 1, \dots, T, \end{aligned} \quad (3)$$

where  $\|w\|_p := \left( \sum_{j=1}^N |w_j|^p \right)^{\frac{1}{p}}$  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 model may generate an infinite optimal value if  $\alpha = -\infty$ . But for a finite  $\alpha$ , the feasible region  $\mathcal{F}$  is bounded, then the corresponding optimal value is bounded. Therefore, we set  $\alpha > -\infty$  to guarantee the validness of the problem (3). Furthermore, when considering the  $l_1$ -sparse minimax model, we restrict ourselves on the case that  $\alpha < 0$ , which means the limited short selling is allowed in the investment; otherwise, the problem (3) will reduce to the original minimax model in that  $\|w\|_1 = 1$ .

Here are remarks to discuss the parameter  $\tau$  of the  $l_p$ -sparse minimax models.

- (i) Let  $(w_{(\tau)}, M_{p(\tau)})$  denote the solution of the model (3) with a specific  $\tau$ . Following Brodie et al. [4], we observe that an optimal solution of the  $l_p$ -sparse minimax model satisfies the following relation

$$(\tau_1 - \tau_2) \left( \|w_{(\tau_2)}\|_p^p - \|w_{(\tau_1)}\|_p^p \right) \geq 0. \quad (4)$$

Indeed, we have

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

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

- (ii) Let  $w^+$  and  $w^-$  denote 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 (4) that

$$(\tau_1 - \tau_2) \left( \|w_{(\tau_2)}^-\|_1 - \|w_{(\tau_1)}^-\|_1 \right) \geq 0. \quad (5)$$

Noting  $\sum_{j=1}^N w_j = \|w^+\|_1 - \|w^-\|_1$  and  $\|w\|_1 = \|w^+\|_1 + \|w^-\|_1$ , we obtain the relation  $\|w_{(\tau_2)}\|_1 - \|w_{(\tau_1)}\|_1 = 2(\|w_{(\tau_2)}^-\|_1 - \|w_{(\tau_1)}^-\|_1)$ . Hence, (5) holds from (4). The inequality (5) demonstrates that, with a smaller  $\tau$ , the portfolio produced by (3) has more short selling stocks and that a nonnegative portfolio may be obtained when  $\tau$  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  $l_1$ -sparse model, a parametric algorithm is proposed as a generalization of the algorithm introduced in [23].

**3.1. The (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

$$\begin{aligned}
 \min_{w, M_p} \quad & -\frac{(\sum_{j=1}^N \bar{y}_j w_j - r_f)}{\lambda - M_p} + \tau \|w\|_1 \\
 \text{s.t.} \quad & \sum_{j=1}^N y_{jt} w_j \geq M_p, \quad t = 1, \dots, T \\
 & \sum_{j=1}^N w_j = 1 \\
 & w_j \geq \alpha, \quad j = 1, \dots, N,
 \end{aligned} \tag{6}$$

where  $r_f$  is the risk-free rate of return and  $\lambda$  is a parameter such that  $\lambda - M_p > 0$ . Also, we assume that  $\sum_{j=1}^N \bar{y}_j w_j - r_f \geq 0$  subject to the above constraints. Relations (4) and (5) still hold for the  $l_1$ -sparse minimax Sharpe ratio model.

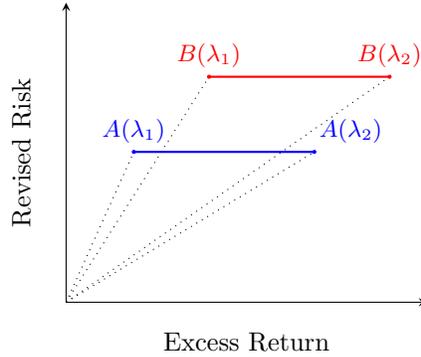


FIGURE 1. Different  $\lambda$  of (generalized) Sharpe ratio

The choice of  $\lambda$  may influence the final result. Figure 1 states a risk-return space. Point  $A(\lambda_1)$  represents the excess return of portfolio  $A$  and its corresponding risk revised by  $\lambda_1$ , and other points are denoted similarly. In this space, the value of the Sharpe ratio is expressed by the slope of the point. As we can see from Figure 1, under the selection of  $\lambda_1$ , portfolio  $A$  is better than portfolio  $B$  in that the gradient of  $A(\lambda_1)$  is steeper than that of  $B(\lambda_1)$ ; while in the situation with  $\lambda_2$ , the result is opposite. But we have to emphasize that no matter how  $\lambda$  is chosen, the revised minimax risk measure remains the essence of the original, 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 (6). To this end, we study a generalization of the parametric algorithm proposed by Konno and Kuno [23], 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

$$\begin{aligned} \min_x \quad & g(x) - \frac{c_1^T x + c_{10}}{c_2^T x + c_{20}} \\ \text{s.t.} \quad & x \in X =: \{x \in \mathbb{R}^n : A_1 x \geq b_1, A_2 x = b_2\}, \end{aligned} \quad (7)$$

where  $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  is a convex but not necessarily differentiable function. Moreover,  $c_1, c_2 \in \mathbb{R}^n$ ,  $c_{10}, c_{20} \in \mathbb{R}$ ,  $A_1 \in \mathbb{R}^{p \times n}$ ,  $A_2 \in \mathbb{R}^{q \times n}$ ,  $b_1 \in \mathbb{R}^p$ , and  $b_2 \in \mathbb{R}^q$  are given parameters of the problem (7). We assume that the feasible region  $X$  is non-empty and bounded; for any  $x \in X$ , it holds that

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

Consider the following auxiliary problem of the parametric algorithm

$$\begin{aligned} \min_{x, \xi} \quad & g(x) - 2\xi \sqrt{c_1^T x + c_{10}} + \xi^2 (c_2^T x + c_{20}) \\ \text{s.t.} \quad & x \in X, \quad \xi \geq 0, \end{aligned} \quad (8)$$

where  $\xi \in \mathbb{R}$  is an auxiliary variable. For a fixed  $x = \bar{x}$ , (8) is a convex quadratic problem with respect to the single variable  $\xi$ , and the optimal value is attained at  $\xi = \frac{\sqrt{c_1^T \bar{x} + c_{10}}}{c_2^T \bar{x} + c_{20}}$ . The relation between (7) and (8) is given by Proposition 1.

**Proposition 1.** (Theorem 4.3, [23]) *Let  $(x^*, \xi^*)$  be an optimal solution of (8). Then  $x^*$  is an optimal solution of (7).*

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

$$\begin{aligned} \min_x \quad & F(x, \xi) =: g(x) - 2\xi \sqrt{c_1^T x + c_{10}} + \xi^2 (c_2^T x + c_{20}) \\ \text{s.t.} \quad & x \in X. \end{aligned} \quad (9)$$

It is evident that  $F(\cdot, \xi)$  is a convex function due to the concavity of  $\sqrt{c_1^T x + c_{10}}$ . According to Proposition 1, an optimal solution of (7) can be obtained by solving (9) with  $\xi = \xi^*$ , where  $\xi^*$  is a nonnegative number such that  $P(\xi^*) \leq P(\xi)$  holds for all  $\xi \geq 0$ . Therefore, the main idea of the parametric algorithm is to solve (9) over all  $\xi \geq 0$ , then the solution corresponding to the smallest optimal value is an optimal solution of (7). However, it is impossible to compute (9) when  $\xi \rightarrow \infty$ . As a matter of fact, from the point of parametric programming, all the problems with a sufficiently large  $\xi$ , say  $\xi \geq \xi_{max}$ , share the same optimal solutions (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  $\xi_{max}$ , and the first stage is to find  $x_{max}^*$  using Proposition 2.

**Proposition 2.** *There exists  $\xi_{max} \in \mathbb{R}$  such that  $x_{max}^*$  is an optimal solution of the problem (8) for any  $\xi \geq \xi_{max}$ , where  $x_{max}^* \in S^* := \arg \min\{g(x) : x \in S_1^*\}$  and*

$$S_1^* := \arg \max\{c_1^T x : x \in S_2^*\}, \quad S_2^* := \arg \min\{c_2^T x : x \in X\}.$$

*If  $S_2^*$  is a singleton, then  $x_{max}^* = \arg \min\{c_2^T x : x \in X\}$ .*

*Proof.* We only need to prove the following statement: there exists  $\xi_{max} \in \mathbb{R}$ , for any  $\xi \geq \xi_{max}$  and  $x \in X \setminus S^*$ , we have  $F(x, \xi) \geq F(x_{max}^*, \xi)$ . To this end, we consider three cases: **(1)**  $x \in S_1^* \setminus S^*$ ; **(2)**  $x \in S_2^* \setminus S_1^*$ ; **(3)**  $x \in X \setminus S_2^*$ . Note

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

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

For case **(1)**, we have  $\gamma_1(x) = \gamma_2(x) = 0$  and  $\gamma_3(x) > 0$ , then  $F(x, \xi) \geq F(x_{max}^*, \xi)$  holds for all  $\xi_{max} \in \mathbb{R}$ . For case **(2)**, we have  $\gamma_1(x) = 0$  and  $\gamma_2(x) < 0$ , then  $\xi_{max}$  satisfying  $F(x, \xi) \geq F(x_{max}^*, \xi)$  exists in that  $\gamma_3(x)$  is bounded. For case **(3)**, we have  $\gamma_1(x) > 0$  and

$$\frac{F(x, \xi) - F(x_{max}^*, \xi)}{\gamma_1(x)} = \left( \xi - \frac{\gamma_2(x)}{\gamma_1(x)} \right)^2 + \frac{\gamma_3(x)}{\gamma_1(x)} - \left( \frac{\gamma_2(x)}{\gamma_1(x)} \right)^2.$$

Then, by the boundedness of  $\frac{\gamma_2(x)}{\gamma_1(x)}$  and  $\frac{\gamma_3(x)}{\gamma_1(x)}$ , there exists  $\xi_{max}$  such that  $F(x, \xi) \geq F(x_{max}^*, \xi)$ , for all  $x \in X \setminus S_2^*$  and  $\xi \geq \xi_{max}$ . Therefore, the proof is completed.  $\square$

We have to point out that, although Konno and Kuno [23] gave an approach to finding  $x_{max}^*$  of a generalized linear multiplicative programming problem, their criteria may fail when the related problem has more than one solution.

Now, we introduce how to locate  $\xi_{max}$  by the use of  $x_{max}^*$ . When  $\xi \geq \xi_{max}$ , since  $x_{max}^*$  is an optimal solution of (9) and the linearity constraint qualification (LCQ) is satisfied, then there exist  $\lambda := (\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p$  and  $\mu := (\mu_1, \dots, \mu_q) \in \mathbb{R}^q$  such that

$$\begin{cases} 0 \in \partial g(x_{max}^*) - \frac{c_1}{\sqrt{c_1^T x_{max}^* + c_{10}}} \xi + c_2 \xi^2 - A_1^T \lambda - A_2^T \mu \\ \lambda_i (A_1 x_{max}^* - b_1)_i = 0, \lambda_i \geq 0, i = 1, \dots, p, \end{cases} \quad (10)$$

where  $(A_1 x_{max}^* - b_1)_i$  represents the  $i^{th}$  entry of the vector  $A_1 x_{max}^* - b_1$ . Let  $\bar{\lambda}$  and  $\bar{A}_1$  be the sub-vector and sub-matrix of  $\lambda$  and  $A_1$  corresponding to the active constraints in  $A_1 x_{max}^* \geq b_1$  (i.e., the inequality is strict). Then (10) becomes

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

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

$$\begin{cases} A_0^T \nu \in \partial g(x_{max}^*) - \frac{c_1}{\sqrt{c_1^T x_{max}^* + c_{10}}} \xi + c_2 \xi^2 \\ \bar{\lambda} \geq 0, \end{cases} \quad (11)$$

Solving (11) 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, (11) can be rearranged as

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

where  $Q_0 := \{(A_0^T)^{-1}\} \times \partial g(x_{max}^*)$ ,  $q_1 := \frac{(A_0^T)^{-1} c_1}{\sqrt{c_1^T x_{max}^* + c_{10}}}$ , and  $q_2 := (A_0^T)^{-1} c_2$ . 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  $\lambda$ . Then, the existence of  $\lambda$  implies

$$q_0^{(\lambda)} - q_1^{(\lambda)}\xi + q_2^{(\lambda)}\xi^2 \geq 0 \quad \text{and} \quad q_0^{(\lambda)} = \max\{Q_0^{(\lambda)}\}, \quad (12)$$

where  $q_0^{(\lambda)} = \max\{Q_0^{(\lambda)}\}$  means  $q_0^{(\lambda)}$  is a vector consisting of the maximums of all sets in  $Q_0^{(\lambda)}$ . Noting that the existence of  $\lambda$  is certain, we have  $q_2^{(\lambda)} \geq 0$ . That is, the solution of (12) can be derived explicitly.

Here are remarks about the parametric algorithm.

- (i) Konno and Kuno [23] directly utilized the generalized inverse matrix to solve the system (11). However, this method may lose information and lead to a wrong  $\xi_{max}$  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 very strict. For instance, it is satisfied if  $S_2^*$  (c.f. Proposition 2) is a singleton and the Linear independence constraint qualification (LICQ) holds at  $x_{max}^*$ .
- (iii) Every step of deriving  $\xi_{max}$  is sufficient and necessary; thus,  $\xi_{max}$  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 the optimization problems in Proposition 2 (see  $S^*$ ,  $S_1^*$ , and  $S_2^*$ ) and  $\xi_{max}$  through solving the system (11) or (12).

**Step 2.** If  $\xi_{max} \leq 0$ , then  $x_{max}^*$  is the global solution of the problem (7); otherwise, solve (9) over  $\xi \in [0, \xi_{max}]$ , then the solution  $x^*$  corresponding to the smallest optimal value is the global solution of the problem (7).

A basic method to search the minimal optimal value of the problem (9) over  $[0, \xi_{max}]$  is to discretize the interval. More precisely, we divide the interval  $[0, \xi_{max}]$  into many subdivisions and compute  $P(\xi)$  at every endpoint. When the subdivisions are narrow enough, the resulting solution would be sufficiently close to the global 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  $y_{jt} = \frac{p_{j,t+1} - p_{jt}}{p_{jt}}$ , where  $p_{jt}$  represents the price of stock  $j$  in week  $t$ . The benchmarks are the  $l_1$ -sparse and  $l_{\frac{1}{2}}$ -sparse mean-variance models (see (2)) and the equal-weighted rule (see [15]).

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 the slackness variable  $u_j (= |w_j|)$ , the problem (3) can be equivalently transformed as a linear programming problem

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

Notably, this is a parametric linear programming problem with respect to the parameter  $\tau$ . According to [3], there exists a finite set of breakpoints  $0 \leq \tau_0 < \tau_1 < \dots < \tau_K < \infty$  such that the optimal solution set keeps unchanged on any (open) interval between two successive breakpoints, which is consistent with the figures in Example 2. The computation of the  $l_1$ -sparse mean-variance model is completed by the CVX toolbox [19]. The iterative reweighted minimization method (see [25]) is utilized to compute the  $l_{\frac{1}{2}}$ -sparse minimax model and the  $l_{\frac{1}{2}}$ -sparse mean-variance model. 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  $\tau$  fixed. For reliability, we target  $\tau$  corresponding to 12-14 active stocks for each model. The results are listed in Table 1.

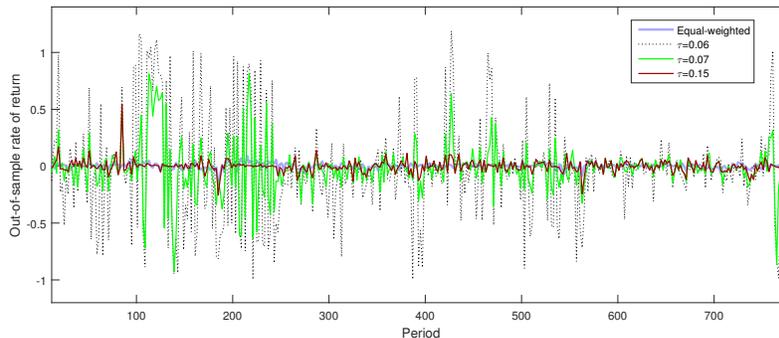
TABLE 1. Computational time

$l_1$ -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  $\alpha$  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 rate of return of the  $l_1$ -sparse minimax rule with different values of  $\tau$  and compare it with that of the equal-weighted rule, which has been shown to outperform many portfolio selection rules (see [15]). As mentioned above, the number of time periods is 11. That is, 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 (3). 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 in 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

FIGURE 2. Rates of return with different  $\tau$

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  $\tau$  leads to evident fluctuations while a larger  $\tau$  produces fewer variations. This tendency is partly due to the relationship between short selling and the value of  $\tau$  (see (5)). According to [26], short selling is considered quite risky, thus causes fluctuations.

**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 and the  $l_{\frac{1}{2}}$ -sparse mean-variance models. For this purpose, we observe their sparsity and short selling with  $\tau$  going through 0 – 0.05. To see the influence of the parameter  $\alpha$ , we conduct experiments with  $\alpha = -0.2$  and  $\alpha = -0.5$ , respectively. Figures 3(a) and 4(a) show that, for all the five models, the number of nonzero stocks in the optimal portfolio (i.e., the level of sparsity) decreases as the value of  $\tau$  increases, which can be explained by (4). 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  $\tau$  goes up, the curve representing the  $l_{\frac{1}{2}}$ -sparse minimax (resp. mean-variance)

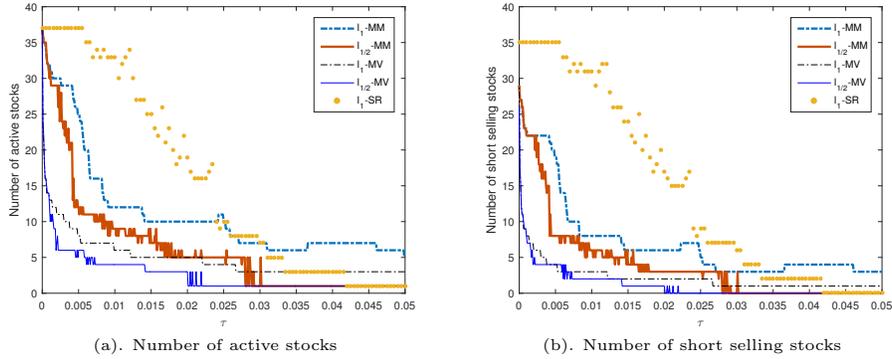


FIGURE 3. Performances with  $\alpha = -0.2$

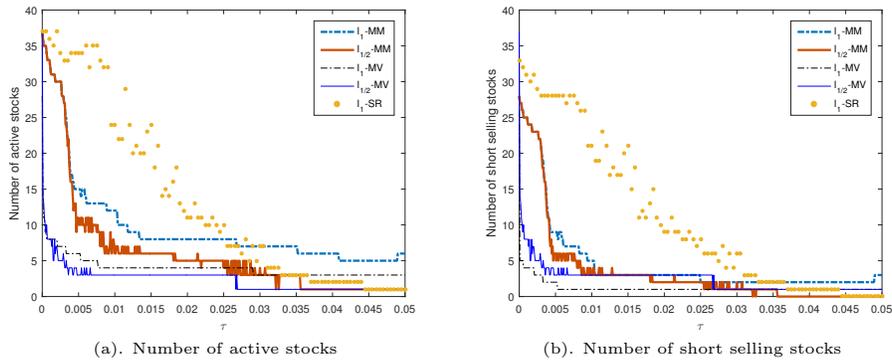


FIGURE 4. Performances with  $\alpha = -0.5$

model reduces more dramatically than that of the  $l_1$ -sparse minimax (resp. mean-variance) model. The monotonic property is quite applicable and plays a critical part in the following examples. 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  $\tau$  over a finite and smaller interval.

Figures 3(b) and 4(b) demonstrate a similar monotonic trend for short selling, which coincides with (5). Although an analogous inequality to (5) 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 Figures 3(a) and 3(b) (or Figures 4(a) and 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, as we mentioned in remark (ii) in Section 2. Comparing Figure 3 and Figure 4, we find that the selection of  $\alpha$  does not influence the descent tendencies; and an extremely sparse optimal portfolio can always be attained with different  $\alpha$ . The only difference is that the extremely sparse optimal portfolio is obtained at a smaller  $\tau$  for the model with a larger  $\alpha$ . 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.

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 range, say  $0 - 0.07$ , while the sparse mean-variance models vary in a much smaller range, say  $0 - 10^{-8}$ . This is attributed to the different orders of magnitude of the objective functions. For the 37-stock problem, the orders of the 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  $\alpha$  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 713 – 723 are used to determine the optimal weights, and then we use them to compute the out-of-sample rate of return of period 724 – 734.

Since the same regularization parameter  $\tau$  in different sparse models generally corresponds to different levels of sparsity, there is little comparability between different sparse models with the same  $\tau$ . A more practical method is to compare them at the same level of sparsity. The comparison in the following example is 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 (see the last five columns of Tables 2(a) to 2(e)), from period 699 – 709 to period 703 – 713. The  $l_1$ -sparse and  $l_{\frac{1}{2}}$ -sparse mean-variance models (see (2)) 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, which can be achieved by adjusting the value of  $\tau$  (see Example 2). However, in general, more than one portfolios fall into the target level of sparsity. For this situation, the smallest risk of these portfolios and its corresponding rate of return, Sharpe ratio, and number of short selling stocks are considered. If the

portfolio with the minimal risk is still not unique, we select one with the maximal rate of return.

TABLE 2. Performances of different sparse models

(a).  $l_1$ -sparse minimax model

$l_1$ -MM	Equal Weight				11-20				21-30				31-40				41-50				51-60			
	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S
period 699-709	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 700-710	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 701-711	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 702-712	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 703-713	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

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

$l_{\frac{1}{2}}$ -MM	Equal Weight				11-20				21-30				31-40				41-50				51-60			
	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S
period 699-709	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 700-710	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 701-711	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 702-712	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 703-713	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

(c).  $l_1$ -sparse mean-variance model

$l_1$ -MV	Equal Weight				11-20				21-30				31-40				41-50				51-60			
	R	Risk <sub>MV</sub>	SR <sub>MV</sub>	S	R	Risk <sub>MV</sub>	SR <sub>MV</sub>	S	R	Risk <sub>MV</sub>	SR <sub>MV</sub>	S	R	Risk <sub>MV</sub>	SR <sub>MV</sub>	S	R	Risk <sub>MV</sub>	SR <sub>MV</sub>	S	R	Risk <sub>MV</sub>	SR <sub>MV</sub>	S
period 699-709	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 700-710	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 701-711	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 702-712	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 703-713	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

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

$l_{\frac{1}{2}}$ -MV	Equal Weight				11-20				21-30				31-40				41-50				51-60			
	R	Risk <sub>MV</sub>	SR <sub>MV</sub>	S	R	Risk <sub>MV</sub>	SR <sub>MV</sub>	S	R	Risk <sub>MV</sub>	SR <sub>MV</sub>	S	R	Risk <sub>MV</sub>	SR <sub>MV</sub>	S	R	Risk <sub>MV</sub>	SR <sub>MV</sub>	S	R	Risk <sub>MV</sub>	SR <sub>MV</sub>	S
period 699-709	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 700-710	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 701-711	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 702-712	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 703-713	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

(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 <sub>MM</sub>	SR <sub>MM</sub>	S	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S	R	Risk <sub>MM</sub>	SR <sub>MM</sub>	S
period 699-709	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 700-710	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 701-711	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 702-712	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 703-713	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

In Tables 2(a) to 2(e), R, Risk<sub>MM</sub>/Risk<sub>MV</sub>, SR<sub>MM</sub>/SR<sub>MV</sub>, 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 results of the equal-weighted rule is also presented for reference. The equal-weighted rule outperforms all the sparse models in terms of the Sharpe ratio due to its extremely small risk. 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 sparse minimax models, a more sparse optimal portfolio tends to have a lower risk and a lower rate of return. However, the changes of risk and rate of return are not so significant for the  $l_1$ -sparse and  $l_{\frac{1}{2}}$ -sparse mean-variance models.

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 better than the  $l_1$ -sparse minimax rule, both in the aspect of the rate of return and Sharpe ratio. When 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 seek extremely sparse portfolios, while the  $l_1$  formulation is more beneficial to those who prefer less sparse portfolios due to its computational simplicity (see Table 1). For the  $l_1$ -sparse and  $l_{\frac{1}{2}}$ -sparse mean-variance models, we do not observe superiorities of the  $l_{\frac{1}{2}}$ -sparse model. As a whole, their out-of-sample performances appear to be commensurate for all levels of sparsity.

Remarkably, the Sharpe ratios of the minimax model and mean-variance model are not comparable in that they are based on their own, but the values of distinct risk measures are not comparable. Therefore, the only performance measure for comparing the sparse minimax models and sparse mean-variance models is the out-of-sample rate of return. From Tables 2(a) and 2(c) (resp. Tables 2(b) and 2(d)), we find that the optimal portfolios of the  $l_1$ -sparse (resp.  $l_{\frac{1}{2}}$ -sparse) minimax model tend to achieve higher rates of return than those of the  $l_1$ -sparse (resp.  $l_{\frac{1}{2}}$ -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 still would be a good choice for investors who do not have the desired return in advance.

TABLE 3. Performances with different  $\alpha$

period 702-712 (level 11-20)	$\alpha = -0.02$				$\alpha = -0.05$				$\alpha = -0.2$				$\alpha = -0.5$			
	R	Risk <sub>out</sub>	SR <sub>out</sub>	S	R	Risk <sub>out</sub>	SR <sub>out</sub>	S	R	Risk <sub>out</sub>	SR <sub>out</sub>	S	R	Risk <sub>out</sub>	SR <sub>out</sub>	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

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  $\alpha$ . The results of period 702 – 712 with  $11 - 20$  active stocks are listed in Table 3 as the representative.

**5. Conclusion.** In this paper, we considered the  $l_p$ -sparse ( $0 < p \leq 1$ ) linear portfolio models and took the minimax selection rule (see [43]) as the 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 [23] 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 investment with a less strict requirement for sparsity. For the  $l_1$ -sparse minimax Sharpe ratio model, it is preferred when the desired return is not given in advance.

**Acknowledgments.** The authors would like to thank two reviewers for their constructive and detailed suggestions and comments which have improved the presentation of the paper. Xiaoqi Yang is supported in part by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (RGC Ref No. 15212817) and Carisa Kwok Wai Yu is supported in part by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/ FDS14/P02/17).

## REFERENCES

- [1] C. R. Bacon, *Practical Portfolio Performance Measurement and Attribution*, 2<sup>nd</sup> edition, John Wiley & Sons, New York, 2008.
- [2] S. Benninga, *Financial Modeling*, 4<sup>th</sup> edition, The MIT Press, London, 2014.
- [3] A. B. Berkelaar, K. Roos and T. Terlaky, The optimal set and optimal partition approach to linear and quadratic programming, in *Advances in Sensitivity Analysis and Parametric Programming* (eds. T. Gal and H.J. Greenberg), Springer, (1997), 159–202.
- [4] J. Brodie, I. Daubechies, C. De Mol, D. Giannone and I. Loris, [Sparse and stable Markowitz portfolios](#), *PNAS*, **106** (2009), 12267–12272.
- [5] G. Burke, A sharper Sharpe ratio, *Futures*, **23** (1994), 56.
- [6] X. Cai, K.-L. Teo, X. Yang and X. Zhou, [Portfolio optimization under a minimax rule](#), *Management Science*, **46** (2000), 957–972.
- [7] T.-J. Chang, N. Meade, J. E. Beasley and Y. M. Sharaiha, [Heuristics for cardinality constrained portfolio optimisation](#), *Computers & Operations Research*, **27** (2000), 1271–1302.
- [8] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, *IEEE Signal Processing Letters*, **14** (2007), 707–710.
- [9] S. S. Chen, D. L. Donoho and M. A. Saunders, [Atomic decomposition by basis pursuit](#), *SIAM Rev.*, **43** (2001), 129–159.
- [10] C. Chen, X. Li, C. Tolman, S. Wang and Y. Ye, Sparse portfolio selection via quasi-norm regularization, preprint, [arXiv:1312.6350](#).
- [11] Z. Dai and F. Wen, [A generalized approach to sparse and stable portfolio optimization problem](#), *J. Ind. Manag. Optim.*, **14** (2018), 1651–1666.
- [12] S. R. Das, H. M. Markowitz, J. Scheid and M. Statman, Portfolios for investors who want to reach their goals while staying on the mean-variance efficient frontier, *The Journal of Wealth Management*, **14** (2011), 25–31.
- [13] I. Daubechies, M. Defrise and C. De Mol, [An iterative thresholding algorithm for linear inverse problems with a sparsity constraint](#), *Comm. Pure Appl. Math.*, **57** (2004), 1413–1457.
- [14] V. DeMiguel, L. Garlappi, F. J. Nogales and R. Uppal, A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, *Management Science*, **55** (2009), 798–812.
- [15] V. DeMiguel, L. Garlappi and R. Uppal, [Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy?](#), *Review of Financial Studies*, **22** (2009), 1915–1953.
- [16] E. J. Elton, M. J. Gruber, S. J. Brown and W. N. Goetzmann, *Modern Portfolio Theory and Investment Analysis*, 9<sup>th</sup> edition, John Wiley & Sons, New York, 2014.
- [17] B. Fastrich, S. Paterlini and P. Winker, [Constructing optimal sparse portfolios using regularization methods](#), *Comput. Manag. Sci.*, **12** (2015), 417–434.
- [18] C. J. Goh and X. Q. Yang, [Analytic efficient solution set for multi-criteria quadratic programs](#), *Nonlinear Anal.*, **30** (1997), 4309–4316.
- [19] M. Grant and S. Boyd, *CVX: Matlab Software for Disciplined Convex Programming, Version 2.2*, 2020. Available from: <http://cvxr.com/cvx>.
- [20] Y. Hu, C. Li, K. Meng, J. Qin and X. Yang, Group sparse optimization via  $l_{p,q}$  regularization, *J. Mach. Learn. Res.*, **18** (2017), 960–1011.
- [21] R. Jagannathan and T. Ma, [Risk reduction in large portfolios: why imposing the wrong constraints helps?](#), *The Journal of Finance*, **58** (2003), 1651–1683.

- [22] M. C. Jensen, The performance of mutual funds in the period 1945-1964, *The Journal of Finance*, **23** (1968), 389–416.
- [23] H. Konno and T. Kuno, Generalized linear multiplicative and fractional programming, *Ann. Oper. Res.*, **25** (1990), 147–161.
- [24] H. Konno and H. Yamazaki, Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market, *Management Science*, **37** (1991), 519–531.
- [25] Z. Lu, Iterative reweighted minimization methods for  $l_p$  regularized unconstrained nonlinear programming, *Math. Program.*, **147** (2014), 277–307.
- [26] D. G. Luenberger, *Investment Science*, Oxford University Press, New York, 1997.
- [27] H. M. Markowitz, How to represent mark-to-market possibilities with the general portfolio selection model, *Journal of Portfolio Management*, **39** (2013), 1–3.
- [28] H. M. Markowitz, Portfolio selection, *The Journal of Finance*, **7** (1952), 77–91.
- [29] H. M. Markowitz and G. P. Todd, *Mean-Variance Analysis in Portfolio Choice and Capital Markets*, John Wiley & Sons, New York, 2000.
- [30] H. M. Markowitz and E. L. van Dijk, Single-period mean-variance analysis in a changing world, *Financ. Anal. J.*, **59** (2003), 30–44.
- [31] J. Matousek and B. Gartner, *Understanding and Using Linear Programming*, Springer-Verlag, Heidelberg, 2007.
- [32] T. A. McCafferty, *The Market is Always Right*. McGraw Hill, New York, 2003.
- [33] A. Niedermayer and D. Niedermayer, Applying Markowitz’s critical line algorithm, in *Handbook of Portfolio Construction* (eds. J. B. Guerard, Jr.), Springer, (2010), 383–400.
- [34] Y. Qi, Parametrically computing efficient frontiers of portfolio selection and reporting and utilizing the piecewise-segment structure, *Journal of the Operational Research Society*, **71** (2020), 1675–1690.
- [35] Y. Qi, Y. Zhang and S. Ma, Parametrically computing efficient frontiers and reanalyzing efficiency-diversification discrepancies and naive diversification, *INFOR Inf. Syst. Oper. Res.*, **57** (2019), 430–453.
- [36] W. F. Sharpe, A linear programming algorithm for mutual fund portfolio selection, *Management Science*, **13** (1967), 499–510.
- [37] W. F. Sharpe, Mutual fund performance, *Journal of Business*, **39** (1966), 119–138.
- [38] F. A. Sortino and R. van der Meer, Downside risk, *The Journal of Portfolio Management*, **17** (1991), 27–31.
- [39] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. Roy. Statist. Soc. Ser. B*, **58** (1996), 267–288.
- [40] J. L. Treynor, How to rate management of investment funds, *Harvard Business Review*, **43** (1965), 63–75.
- [41] M. Woodside-Oriakhi, C. Lucas and J. E. Beasley, Heuristic algorithms for the cardinality constrained efficient frontier, *European J. Oper. Res.*, **213** (2011), 538–550.
- [42] Z. Xu, H. Zhang, Y. Wang, X. Chang and Y. Liang,  $L_{1/2}$  regularization, *Sci. China Inf. Sci.*, **53** (2010), 1159–1169.
- [43] M. R. Young, A minimax portfolio selection rule with linear programming solution, *Management Science*, **44** (1998), 673–683.
- [44] T. W. Young, Calmar ratio: A smoother tool, *Futures*, **20** (1991), 40–40.

Received October 2020; 1st revision January 2021; 2nd revision March 2021.

*E-mail address:* [chen-chen.zu@connect.polyu.hk](mailto:chen-chen.zu@connect.polyu.hk)

*E-mail address:* [mayangxq@polyu.edu.hk](mailto:mayangxq@polyu.edu.hk)

*E-mail address:* [carisayu@hsu.edu.hk](mailto:carisayu@hsu.edu.hk)