Nonconvex $L_{1/2}$ Regularization For Sparse Portfolio Selection

Xu Fengmin [1]   Wang Guan [2]   Gao Yuelin[3]

Abstract. Two sparse optimal portfolio selection models with and without short-selling constraints are proposed by introducing $L_{1/2}$ regularization on portfolio weights in the traditional M-V portfolio selection model. A fast and efficient penalty half thresholding algorithm is presented for the solution of proposed sparse portfolio selection models. The algorithm is an extension of the half thresholding algorithm for solving $L_{1/2}$ regularization problems. A strategy to adjust the value of the regularization parameter in proposed models is derived when the sparsity of optimal portfolios is specified. When the strategy is incorporated into the modified algorithm, the efficiency of the algorithm is improved. Empirical analyzes on the proposed portfolio selection models and the proposed algorithm are executed to test the out-of-sample performance of the optimal portfolios generated from the proposed models. The out-of-sample performance is measured using Sharpe ratio. Empirical results show that the out-of-sample performance of the optimal portfolios generated from the proposed models is better than those of the optimal portfolios generated from M-V portfolio selection models with L_1 regularization on portfolio weights.

Keywords: $L_{1/2}$ regularization; Sparse portfolio selection; Half thresholding algorithm.

AMS Subject Classifications: 90C26, 91G10, 90C90

1 Introduction

The mean-variance model (M-V) for portfolio selection proposed by Markowitz (1952) is to determine a portfolio such that the return and the risk of the portfolio have a favorable trade-off. In the M-V model, the return and the risk of a portfolio are measured by the mean and the variance of the portfolio random returns, respectively, and the optimal portfolio is obtained by solving a quadratic programming problem. The M-V model has a profound impact on economic modeling of finance markets and asset pricing.

Unfortunately, when the number of assets is large and the returns of these assets are highly related, the M-V model is an ill-posed problem in the sense that the optimal portfolio obtained from the model is very sensitive to perturbations in the input parameters of the problem, that is, a slight change in the problem input parameters will result a large change in the resulting portfolio weights. Empirical analysis also shows that the performance of the portfolio generated from M-V model is poor in out-of-sample tests [1, 2]. Several techniques have been suggested to reduce the sensitivity of the M-V optimal portfolio to uncertainties

on input parameters, such as multi-factor models [3], Bayesian estimation of means and covariance [4, 5] and covariance shrinkage [6, 7, 8, 9].

Statistical regularization methods have wide applications in constructing large scale M-V portfolio selection models to find sparse optimal portfolios with better out-of-sample performance and to reduce transaction cost (see [1, 10, 11, 12, 13, 14, 15, 16]). Lorenzo et al (2012) propose a concave optimization model for sparse optimal portfolio selection by introducing an either L_1 or L_2 regularization [17] on portfolio weights. Empirical analyses in a M-V framework support the use of the L_1 regularization when short-selling is permitted. However, the L_1 regularization approach is ineffect in promoting sparsity in presence of budget and no-short-selling constraints [1]. DeMiguel and Garlappi (2009) propose a unified framework for optimal portfolio selection to improve performance through constraining portfolio norms [10]. Brodie et al (2008), Fan et al (2009) propose sparse and stable portfolio selection models by imposing upper bounds on the 1-norm of portfolio weights [18]. Introducing regularization in portfolio selection model relaxes the limit of without short-selling for large portfolio selection (see [6]) and the regularization parameter is used to control the sparsity of the resulting optimal portfolio [11, 12]. Yen (2010) proposes a minimum variance portfolio selection model by adding a constraint on portfolio weights and designs a coordinate-wise descent algorithm to solve the resulting model [13]. Carrasco and Noumon (2010) propose four regularization techniques through stabilizing the inverse of covariance matrices and present methods to determine optimal regularization parameter values [14]. Still and Kondor (2010) propose a regularized optimal portfolio selection model in terms of the statistical learning theory [15]. Brandt et al (2009) present a new portfolio selection approach that avoids the difficulties in estimating asset returns moments by directly modeling the portfolio weight in each asset as a function of the asset returns [16].

Based on authors recent researches on $L_{1/2}$ regularization (see [19, 20, 21]), two non-convex sparse portfolio selection models are proposed in this paper. The proposed models can generate optimal portfolios with better sparsity than the portfolio selection models using L_1 or L_2 regularization do. The reasons for using $L_{1/2}$ regularization on portfolio weights are as follows. Firstly, a sparse portfolio means that less assets are selected in an investment and transaction cost may be saved. Secondly, although the portfolio selection problem with $L_{1/2}$ regularization on portfolio weights is a non-convex, non-smooth and non-Lipschitz continuous optimization problem, a penalty half thresholding algorithm is proposed for the solution of resulting portfolio selection models, and the algorithm is fast and efficient, especially for large scale portfolio selection problems. Finally, a strategy to adjust the value of the regularization parameter is derived, when the sparsity of optimal portfolios is specified using a parameter. The efficiency of the penalty half thresholding algorithm is improved when this strategy is incorporated into the algorithm.

The main contributions of the paper are as follows.

• Two sparse optimal portfolio selection models with and without short-selling constraints are proposed by introducing $L_{1/2}$ regularization on portfolio weights, and a fast and efficient penalty half thresholding algorithm is presented to solve the resulting portfolio models by extending the half thresholding algorithm in [19].

• A strategy to adjust the value of regularization parameter in the portfolio selection model is derived when the sparsity of portfolios is specified.

The rest of the paper is organized as follows. Section 2 reviews the sparse M-V portfolio selection model with L_1 regularization on portfolio weights, and then presents the two sparse

portfolio selection models by introducing $L_{1/2}$ regularization on portfolio weights. The penalty half thresholding algorithm and the strategy to adjust the value of the regularization parameter are presented in Section 3. The steps of the algorithm is also described in Section 3. Section 4 reports empirical tests and comparisons on the proposed portfolio selection models and some existing portfolio selection models. Conclusions are given in Section 5.

2 New sparse portfolio selection models

In this section we recall, at first, the sparse M-V model with L_1 regularization presented in[11]. Then the new sparse portfolio selection models that uses $L_{1/2}$ regularization and the analysis of the models are presented.

2.1 The Sparse mean-variance model

It is assumed that there are N assets. Let $r_t = (r_{1t}, r_{2t}, \cdots, r_{Nt})^T \in R^N$ be the vector of asset returns at time t, $t = 1, \cdots, T$, $E(r_t) = \mu$ and $C = E[(r_t - \mu)(r_t - \mu)^T]$ be the mean return vector and the covariance matrix of asset returns, where r_{it} is the return of asset i at time t. Then the traditional M-V portfolio selection model can be expressed as follows

$$
\begin{aligned}
\textbf{min} \quad & w^T C w \\
\textbf{s.t.} \quad & \mu^T w = \rho \\
& e^T w = 1,
\end{aligned}
\tag{2.1}
$$

where $w = (w_1, w_2, \cdots, w_N)^T \in R^N$ is the vector of asset weights, $e \in R^N$ is the vector of all ones and ρ is the minimum expected return from the portfolio that is expected by an investor.

Since $C = E[(r_t - \mu)(r_t - \mu)^T]$, we have

$$
w^T C w = E[|\rho - w^T r_t|^2] = \frac{1}{T} \|\rho e - Rw\|_2^2,
$$

where $R = (r_1, \cdots, r_T)^T \in R^{T \times N}$. Then the M-V model can be expressed, in the statistical regression view, as follows

$$
\begin{aligned}
\textbf{min} \quad & \frac{1}{T} \|\rho e - Rw\|_2^2 \\
\textbf{s.t.} \quad & \hat{\mu}^T w = \rho \\
& e^T w = 1,
\end{aligned}
\tag{2.2}
$$

where $\hat{\mu} = \frac{1}{T} \sum\limits_{t=1}^{T} r_t$ and $\| \cdot \|_2$ is the L_2 vector norm.

Problem (2.2) is unstable when the matrix $R^T R$ is ill-conditioned, Brodie et al.(2008) present a modification to problem (2.2) by introducing a penalty with L_1 regularization on the portfolio weights in the objective function to combat the unstability [11]

$$
\begin{aligned}
\textbf{min} \quad & \frac{1}{T} \|\rho e - Rw\|_2^2 + \lambda \|w\|_1 \\
\textbf{s.t.} \quad & \hat{\mu}^T w = \rho \\
& e^T w = 1,
\end{aligned}
\tag{2.3}
$$

where λ is a regularization parameter. Problem (2.3) is a sparse and stable M-V portfolio selection model, since the introduction of the L_1 penalty term in the objective function stabilize the problem, promotes sparsity in some degree, and regulates the amount of shorting

in the resulting portfolio. However, sparsity of the resulting portfolio is not guaranteed from problem (2.3), since the L_1-norm of the asset weights will result in a constant value of one when asset weights is nonnegative. Moreover, it is difficult to choice a proper value for the penalty parameter λ, when a penalty function method is employed to solve the problem (2.3). An improvement to problem (2.3) is given in the next subsection by using $L_{1/2}$ regularization proposed in [19, 20, 21] to replace the L_1 regularization in (2.3).

2.2   New sparse portfolio selection models

$L_{1/2}$ regularization is one of statistical regularization methods and introducing the $L_{1/2}$ regularization in a problem variables aims to find a sparse solution of the problem. A simple presentation of an optimization problem with $L_{1/2}$ regularization is as follows

$$\min_{x \in R^N} \|Ax - b\|^2 + \lambda \|x\|_{1/2}^{1/2}, \tag{2.4}$$

where $\lambda > 0$ is the regularization parameter and controls the sparsity of optimal solutions, $\|x\|_{1/2}$ is the $L_{1/2}$ quasi-norm of the vector $x \in R^N$, and is defined by

$$\|x\|_{1/2}^{1/2} = \sum_{i=1}^{N} |x_i|^{1/2}.$$

Problem (2.4) is non-convex, non-smooth and non-Lipschitz continuous. However, an iterative algorithm, called half thresholding algorithm, is proposed for the fast solution of problem (2.4) in [19], and a novel strategy to set the value of the regularization parameter λ is suggested. Numerical experiments show that the sparsity of the solutions generated from problem (2.4) using the half thresholding algorithm is satisfactory (see [19]).

Now, we are ready to describe the new sparse portfolio selection model. The model is a variant of problem (2.3) and is formulated by using $L_{1/2}$ regularization to replace the L_1 regularization in (2.3)), that is,

$$\begin{aligned} \min \quad & \frac{1}{T}\|Rw - \rho e\|_2^2 + \lambda \|w\|_{1/2}^{1/2} \\ \text{s.t.} \quad & \hat{\mu}^T w = \rho \\ & e^T w = 1. \end{aligned} \tag{2.5}$$

If sort-selling is not permitted, the model for portfolio selection has the form

$$\begin{aligned} \min \quad & \frac{1}{T}\|Rw - \rho e\|_2^2 + \lambda \|w\|_{1/2}^{1/2} \\ \text{s.t.} \quad & \hat{\mu}^T w = \rho \\ & e^T w = 1 \\ & w \geq 0. \end{aligned} \tag{2.6}$$

Problems (2.5) and (2.6) can be expressed in the matrix-vector form

$$\begin{aligned} \min \quad & \frac{1}{T}\|Rw - \rho e\|_2^2 + \lambda \|w\|_{1/2}^{1/2} \\ \text{s.t.} \quad & Aw = b \end{aligned} \tag{2.7}$$

and

$$\begin{aligned}
\mathbf{min} \quad & \tfrac{1}{T}\|Rw - \rho e\|_2^2 + \lambda\|w\|_{1/2}^{1/2} \\
\mathbf{s.t.} \quad & Aw = b \\
& w \geq 0,
\end{aligned} \tag{2.8}$$

where $b = (\rho, 1)^T$, $A^T = (\hat{\mu}, \ e)$.

Since $\|w\|_{1/2}^{1/2}$ is non-convex, non-smooth and non-Lipschitz continuous, Problems (2.7) and (2.8) are hard to solve. Thus, the algorithm given in the next section aims to find local solutions of problems (2.7) and (2.8).

3 An efficient algorithm for the solution of problem (2.7)

The algorithm given in this section is an extension of the half thresholding algorithm for solving $L_{1/2}$ minimization problem proposed by Xu et al in [19].

3.1 The penalty function

The proposed method uses the penalty function minimization technique to find a local solution of problem (2.7). The penalty function problem for problem (2.7) is given by

$$\mathrm{min} \quad \frac{1}{T}\|Rw - \rho e\|_2^2 + \lambda\|w\|_{1/2}^{1/2} + \gamma\|Aw - b\|_2^2, \tag{3.1}$$

where $\gamma > 0$ is a penalty parameter. Let $w = (w_1, x_2, \cdots, w_N)^T$ be a local minimizer of problem (3.1) with $w_i \neq 0$ for all $i = 1, 2 \cdots, N$. Then the first order necessary optimality condition

$$\frac{1}{T}R^T(Rw - \rho e) + \gamma A^T(Aw - b) + \frac{\lambda}{2}\partial(\|w\|_{1/2}^{1/2}) = 0 \tag{3.2}$$

is satisfied at w (see Xu [19]). Condition (3.2) can be rearranged into the following form by multiplying a parameter $\mu(0 < \mu < 1)$ at the both sides of equation (3.2)

$$w + \mu\frac{1}{T}R^T(\rho e - Rw) + \mu\gamma A^T(b - Aw) = w + \frac{\lambda\mu}{2}\partial(\|w\|_{1/2}^{1/2}). \tag{3.3}$$

Let

$$P_{\lambda\mu,1/2}(\cdot) = (I + \frac{\lambda\mu}{2}\partial(\|\cdot\|_{1/2}^{1/2}))^{-1},$$

where it is assumed that the inverse of $\partial(\|\cdot\|_{1/2}^{1/2})$ exists. Then it follows from (3.3) that

$$\begin{aligned}
w &= (I + \frac{\lambda\mu}{2}\partial(\|\cdot\|))^{-1}(w + \mu\frac{1}{T}R^T(\rho e - Rw) + \mu\gamma A^T(b - Aw)) \\
&= P_{\lambda\mu,1/2}(w + \mu\frac{1}{T}R^T(\rho e - Rw) + \mu\gamma A^T(b - Aw)) \\
&= P_{\lambda\mu,1/2}B(w),
\end{aligned}$$

where $B(w) = w + \mu\frac{1}{T}R^T(\rho e - Rw) + \mu\gamma A^T(b - Aw)$.

Note that the above conclusion is obtained based on the assumption that w is a local minimizer of problem (3.1) with $w_i \neq 0$ for all $i = 1, 2, \cdots, N$. Now we consider a local minimizer of problem (3.1) without the assumption. Let w^* be such a local minimizer. The following theorem directly comes from Theorem 2 in [19].

5

Theorem 3.1 *Let w^* be a local minimizer of problem (3.1), If either $w_i^* = 0$ or $|B(w^*)_i| > \frac{3}{4}(\lambda)^{2/3}$ holds for any $i = 1, 2, \cdots, N$, then*

$$|B(w^*)_i| \leq \frac{3}{4}(\lambda\mu)^{2/3} \Leftrightarrow w_i^* = 0.$$

Theorem 3.1 indicates that the elements of the local minimizer w^* either takes the value $w_i^* = 0$ or satisfies the condition $|B(w^*)_i| > \frac{3}{4}(\lambda\mu)^{2/3}$. In fact, if the elements of a solution are all positive and satisfy the condition $|B(w)_i| > \frac{3}{4}(\lambda\mu)^{2/3}$, then the operator $P_{\lambda,1/2}$ exists and is given by (see [19])

$$P_{\lambda\mu,1/2}(B(w)) = (h_{\lambda\mu,1/2}(B(w)_1), h_{\lambda\mu,1/2}(B(w)_2), ..., h_{\lambda\mu,1/2}(B(w)_N)), \tag{3.4}$$

where

$$h_{\lambda\mu,1/2}(B(w)_i) = \begin{cases} \frac{2}{3}B(w)_i \left(1 + \cos(\frac{2\pi}{3} - \frac{2}{3}\varphi_{\lambda\mu}(B(w)_i))\right), & |B(w)_i| \geq \frac{3}{4}(\lambda\mu)^{\frac{2}{3}} \\ 0, & otherwise \end{cases} \tag{3.5}$$

with

$$\varphi_{\lambda\mu}(B(w)_i) = \arccos\left(\frac{\lambda\mu}{8}\left(\frac{|B(w)_i|}{3}\right)^{-\frac{3}{2}}\right). \tag{3.6}$$

Based on the above analysis, given an initial point w^1, the penalty half thresholding algorithm uses the following iteration

$$w^{k+1} = P_{\lambda\mu,1/2}(B(w^k)), \quad k = 1, 2, \cdots \tag{3.7}$$

to generate iterates, where w^k is the k-th iterate.

3.2   Adjusting values for the regularization parameter

In this subsection we suggest a way to automatically adjust the value for the regularization parameter λ. Various methods (see Akaike (1973)[22], Schwarz(1978) [23]) have been proposed for estimation of regularization parameters through available information of the problem under consideration. As for a portfolio selection problem, the prior information on assets is generally available and can be used to estimate the value of regularization parameter.

Suppose that the portfolio is required to be k-sparsity, that is, the portfolio should consist of k assets. Let w^* be a local minimizer of the penalty problem (3.1), and $|[B(w^*)]_k|$ be the $k-$th largest value among the absolute elements of $B(w^*)$. Then it follows from Theorem 3.1 that we have

$$|[B(w^*)]_1| \geq |[B(w^*)]_2| \geq \cdots \geq |[B(w^*)]_N|,$$

$$i \in \{1, 2, \cdots, k\} \Leftrightarrow |[B(w^*)]_i| > \frac{3}{4}(\lambda\mu)^{\frac{2}{3}}$$

and

$$i \in \{k+1, \cdots, N\} \Leftrightarrow |[B(w^*)]_i| \leq \frac{3}{4}(\lambda\mu)^{\frac{2}{3}}.$$

It follows that

$$\{|[\frac{4}{3}B(w^*)]_{k+1}|\}^{3/2} \le \lambda\mu < \{|[\frac{4}{3}B(w^*)]_k|\}^{3/2}. \tag{3.8}$$

From (3.8) we obtain

$$\lambda^* \in [\frac{1}{\mu}\{|[\frac{4}{3}B(w^*)]_{k+1}|\}^{3/2}, \quad \frac{1}{\mu}\{|[\frac{4}{3}B_\mu(w^*)]_k|\}^{3/2}], \tag{3.9}$$

(3.9) gives a suggestion on how to adjust the value of the regularization parameter λ. Since w^* is unknown, when w^n is the best available approximation to w^*, a proper choice for the value of λ^n at n-th iteration is given by

$$\lambda^n = \frac{1}{\mu}|\frac{4}{3}[B(w^n)]_{k+1}|^{\frac{3}{2}}. \tag{3.10}$$

That is, (3.10) can be used to adjust the value of the regularization parameter λ during iteration.

3.3   The penalty half thresholding algorithm

Based on the analyses given in subsections 3.1 and 3.2, the proposed penalty half thresholding algorithm for a local solution of problem (2.7) can be described as follows.

Algorithm 1   The penalty half thresholding algorithm

Step 1.   *Generate an initial point w^1 and μ. Give values to parameters γ and δ and a tolerance $\epsilon > 0$, and Set $n = 1$.*

Step 2.   *Compute*

$$\lambda^n = \frac{1}{\mu}[\frac{4}{3}|(B(w^n)_{k+1})|^{3/2}],$$

$$B(w^n) = w^n + \mu\frac{1}{T}R^T(\rho e - Rw^n) + \mu\gamma A^T(b - Aw^n);$$

Step 3.   *Compute w^{n+1}.*
 If $B(w^n)_i > \frac{3}{4}(\lambda^n\mu)^{2/3}$, then $w_i^{n+1} = h_{\lambda^n,1/2}(w_i^n)$, else $w_i^{n+1} = 0$, $i = 1, 2, \cdots, N$.

Step 4. *If $\|w^{n+1} - w^n\| < \epsilon$, then go to step 5; Else $n = n + 1$, go to step 2.*

Step 5. *If $\|Aw^{n+1} - b\| < \epsilon$, then stop; Else $\gamma = \delta\gamma$, go to step 1.*

   Remark 1. It follows from [19] that when

$$\mu = \min\{\frac{1}{\gamma\|A\|}, \quad \frac{T}{\|R\|}\} \tag{3.11}$$

the convergence of the algorithm to a local minimizer of problem (2.7) is ensured. Thus, (3.11) is used to adjust the value of the penalty factor γ in the iteration of the algorithm 1.
   Remark 2. The algorithm can also be effectively used to solve the problem (2.8) in which short-shelling is not permitted.

4 Numerical results

Experiments on real market data are performed to test the out-of-sample performance of the sparse portfolio selection models (2.7) and (2.8)) and the efficiency of the modified penalty half thresholding algorithm in solving the problems (2.7) and (2.8). The out-of-sample performance of a portfolio is evaluated using Sharpe ratio. The performance of portfolios generated from the sparse portfolio selection models (2.7) and (2.8) is compared with those generated from the L_1 regularization portfolio model in [11] (solved using Lars algorithm proposed by Efron and Hastie (2004) [25]), the equally weighted strategy (denoted by $1/N$), and the traditional Markowitz M-V portfolio selection model [26]. The tests and comparisons are performed on the benchmark problems FF48 and FF100 in [24]. The problem FF48 forms portfolios from 48 industrial assets and the problem FF100 forms portfolios from 100 assets. These portfolios are conducted at the end of every June. The parameter values $\mu = \min\{\frac{1}{\|A\|}, \frac{T}{\|R\|}\}$ and $\lambda^n = \frac{1}{\mu}|\frac{4}{3}[B(w^n)]_{k+1}|^{\frac{3}{2}}$ are used in applying algorithm 1 to perform these tests, where T is the length of the time period.

4.1 Tests on problem FF48

In this subsection, we report the test results on the problem FF48. Tests use the real market data of these 48 industrial assets during a period of 30 years from July 1976 to June 2006. The time period is divided into 6 equal sub-periods. The tests are performed as follows. The historical daily data of these assets during a sub-period is used as training data, that is, used to estimate the means and covariance matrices of these assets returns, and then these means and covariance matrices are used as the input parameters in different testing portfolio selection models. Note that the average return of all assets during the period is used as the input value of ρ in models. These models are then solved to generate testing portfolios. The out-of-sample performance of these resulting portfolios are tested using the first years data of the following sub-period. That is, the data during July 1976 and June 1981 is used to estimate input parameters in portfolio selection models, and the performance of portfolios generated from portfolio selection models are tested on the data during July 1981 and June 1982. Repeating the process for the 6 sub-periods generates 6 test results for each portfolio selection model. Note that the results in the last row of the following tables are generated by using monthly data in the whole test period. In order to understand the effect of sparsity on the performance of resulting portfolios, the value of k in the following tables specifies the number of assets in a portfolio.

    A. **Test results on model (2.7)**

    Table 1 gives the test results on out-of-sample performance of portfolios generated from models (2.7), the L_1 regularization portfolio selection model, the equal weighting portfolio strategy (1/48) and the M-V model with different sparsity, where the results of the traditional Makowitz model, the L_1 model and the $1/N$ strategy are quoted form [11]. CPU is the computing time of algorithms to generate the solutions that measured by seconds, $Iter$ is the number of algorithm iterations. It can be observed from the table that the out-of-sample performance of the portfolio generated from model (2.7) is the best, the one of the portfolio generated from the L_1 regularization portfolio selection model is the second best, while the out-of-sample performance of the portfolio generated from the M-V model is the worst. It also has been found through further observation that the penalty half thresholding

algorithm is more suitable and free from the choice for the value of regularization parameter in solving model (2.7)

Table 1: Comparison results with no short-selling model (FF48)

Period	$L_{1/2}$ model				L_1 model	$1/N$ strategy	Markowitz
	k	S(%)	*CPU*	*Iter*	S(%)	S(%)	S(%)
76.07-81.06	24	**50**	0.015	32	49	44	39
81.07-86.06	10	**61**	0.016	32	57	31	44
86.07-91.06	21	**29**	0.018	33	20	7	22
91.07-96.06	9	**64**	0.017	32	62	44	46
96.07-01.06	10	**41**	0.016	31	40	17	19
01.07-06.06	15	**36**	0.017	32	30	30	16
76.07-06.06	12	**46**	0.12	54	41	27	30

Table 2 gives the test results for optimal portfolios generated from L_1 and $L_{1/2}$ models with sparsity k being in different intervals $[8, 16], [17, 24], [25, 32], [33, 40], [41, 48]$, where the results for L_1 model are quoted from [11]. It can be observed from Table 2 that the performance of the portfolios generated from model (2.7) is better than the performance of the portfolios generated from the L_1 regularization portfolio selection model in all periods except in the period 07/76-06/81.

Table 2: Comparison results of Sharpe ratio

Period	(k = 8 − 16)		(k = 17 − 24)		(k = 25 − 32)		(k = 33 − 40)		(k = 41 − 48)	
	$L_{1/2}$	L_1	$L_{1/2}$	L_1	$L_{1/2}$	L_1	$L_{1/2}$	L_1	$L_{1/2}$	L_1
76.07-81.06	44	50	**50**	46	47	49	45	50	43	46
81.07-86.06	**61**	58	61	52	56	52	56	46	50	39
86.07-91.06	25	18	**29**	15	26	9	27	7	24	8
91.07-96.06	64	57	**67**	47	61	36	60	30	48	21
96.07-01.06	**41**	38	28	22	28	6	26	4	23	0
01.07-06.06	37	30	**39**	29	33	28	30	27	28	27
76.07-06.06	43	40	**47**	34	47	28	47	26	43	22

B. **Test results on model (2.8)**

Table 3 gives the performance results of the portfolios generated from model (2.8) with different sparsity. The second column in the table gives the number of assets in portfolios that has the best performance in each period. It can be observed from the column that portfolios with 5-8 assets generally have better performance for the problem FF48, and taking sparsity into account in a portfolio selection model is significant.

Figure 1 and Figure 2 below present the relationship between the performance of a portfolio and the sparsity of the portfolio for both models (2.7) and (2.8). It can be observed from the two figures that the performance of portfolios gradually reduces when the number of assets in portfolios increases, especially for model (2.8). This fact further indicates that

9

Table 3: Comparison results of FF48 with short-selling model

Period	optimal k	S(%)	$k = 8 - 16$ S(%)	$k = 17 - 24$ S(%)	$k = 25 - 32$ S(%)	$k = 33 - 40$ S(%)	$k = 41 - 48$ S(%)
76.07-81.06	6	47	44	41	39	43	38
81.07-86.06	23	74	70	74	71	63	54
86.07-91.06	15	34	34	27	25	25	24
91.07-96.06	5	70	62	53	53	47	37
96.07-01.06	5	41	31	30	27	21	14
01.07-06.06	5	37	31	30	22	21	18
76.07-06.06	8	45	45	38	33	34	37

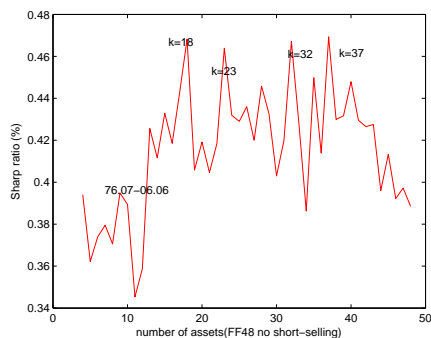the consideration of sparsity is necessary in constructing optimal portfolios.
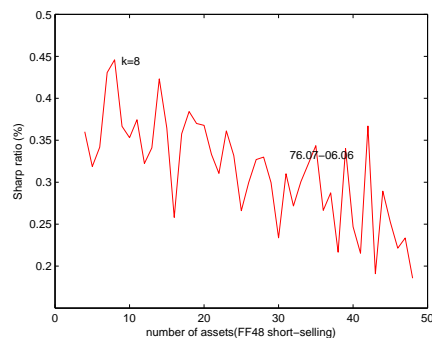

Figure 1: FF48 (no short-selling)

Figure 2: FF48 (short-selling)

4.2   Tests on problem FF100

This subsection reports the test result of portfolios generated from different portfolio selection models on problem FF100. Test results are given in Tables 4-6 and Figures 3 and 4.

Tables 4, 5 and figure 3 give the test results of optimal portfolios from portfolio selection models with on-short-selling constraints. It can be observed from Tables 4 and 5 that the portfolio generated from $L_{1/2}$ model has higher out-of-sample performance than the portfolios generated from the $1/N$ strategy and the L_1 model have, and that the portfolios generated from the $L_{1/2}$ model display higher out-of-sample performance than portfolios generated from the L_1 model do in most cases of $k = 11 - 20$, $k = 21 - 30$, $k = 31 - 40$, $k = 41 - 50$, and $k = 51 - 60$. While figure 3 also clearly shows the feature that out-of-sample performance of a portfolio decreases when the number of assets in the portfolio increases.

Table 6 and figure 4 give the test results of optimal portfolios generated from the $L_{1/2}$ model with the short-selling constraint on problem FF100 for different values of k. It can be observed from the table that the number of assets in a portfolio with higher out-of-sample performance is from 17 to 37 in all the periods. This feature can also be observed from

10

Table 4: Comparison results of Sharpe ration with no short-selling model(FF100)

Period	$L_{1/2}$ model				L_1 model	$1/N$ strategy
	k	S(%)	CPU	$Iter$	S(%)	S(%)
76.07-81.06	14	**48**	0.021	42	21	38
81.07-86.06	22	**58**	0.019	41	49	38
86.07-91.06	29	**72**	0.023	43	15	13
91.07-96.06	37	**68**	0.021	43	61	53
96.07-01.06	39	**44**	0.022	42	35	26
01.07-06.06	34	**32**	0.024	43	21	19
76.07-06.06	8	**46**	0.15	67	30	28

Table 5: Comparison results of Sharpe ratio(FF100 no short-selling)

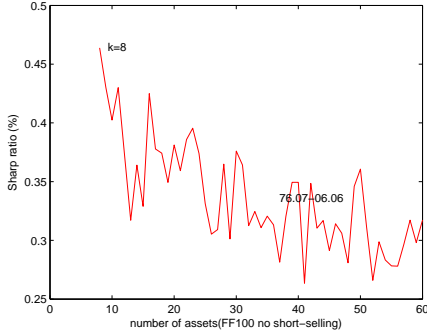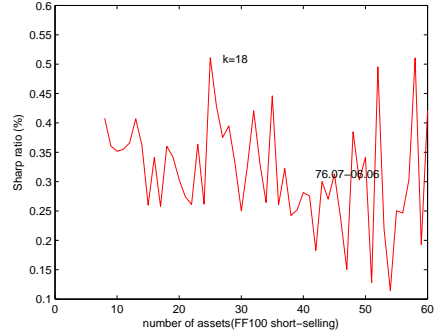| Period	$(k = 11 - 20)$		$(k = 21 - 30)$		$(k = 31 - 40)$		$(k = 41 - 50)$		$(k = 51 - 60)$	
	$L_{1/2}$	L_1	$L_{1/2}$	L_1	$L_{1/2}$	L_1	$L_{1/2}$	L_1	$L_{1/2}$	L_1
76.07-81.06	48	22	34	24	37	22	36	20	33	10
81.07-86.06	56	64	58	81	56	77	54	72	55	67
86.07-91.06	62	14	72	16	68	18	65	19	67	17
91.07-96.06	67	70	66	86	68	73	68	70	66	67
96.07-01.06	45	35	42	44	44	61	42	62	43	54
01.07-06.06	33	22	29	29	30	26	27	23	29	22
76.07-06.06	43	33	40	39	36	40	36	39	32	34

Figure 3: FF100 (no short-selling)

Figure 4: FF100 (short-selling)

figure 4.

Therefore, the following results can be obtained from the test results of these models on problems FF48 and FF100.

- The proposed sparse portfolio selection models outperform the L_1 model, the Markowitz model and the $1/N$ strategy significantly and over most of the evaluation periods in both with short-selling and without short-selling cases.

Table 6: Comparison results of FF100 with short-selling model

Period	optimal k	S(%)	$k = 11 - 20$ S(%)	$k = 21 - 30$ S(%)	$k = 31 - 40$ S(%)	$k = 41 - 50$ S(%)	$k = 51 - 60$ S(%)
76.07-81.06	31	59	52	46	59	50	42
81.07-86.06	19	86	86	72	74	55	50
86.07-91.06	27	51	35	51	44	44	43
91.07-96.06	34	80	72	66	80	71	64
96.07-01.06	37	83	68	72	83	63	69
01.07-06.06	17	48	48	44	47	42	37
76.07-06.06	25	51	41	51	45	34	51

- Test results support the conclusion that the modified penalty half thresholding algorithm is fast, effective and powerful for the solution of the proposed sparse portfolio selection models. The algorithm is suitable, and free from the choice of regularization parameters when the sparsity of a required portfolio is specified.

5 Conclusions

Two new portfolio selection models are proposed to take the sparsity of the resulting optimal portfolios into account. These models are variants of the traditional M-V portfolio selection model by introducing a non-convex and non-smooth $L_{1/2}$ regularization on the weights of portfolios. Since the proposed portfolio selection models are NP-hard, a penalty half thresholding algorithm is then presented for the local solutions of the proposed portfolio selection models. A strategy to adjust the value of the regularization parameter in the algorithm is given through analyzing the property of optimal sparse portfolios. When the strategy is incorporated into the algorithm, the algorithm adaptively adjusts the value of the regularization parameter and the efficiency of the algorithm is improved. Empirical tests on benchmark test problems FF48 and FF100 with real market data are performed to test the out-of-sample performance of the portfolios generated from the solution of proposed portfolio selection models. Comparisons with the L_1 regularization portfolio selection model, the traditional M-V portfolio selection model and the equal weighting strategy show that the proposed portfolio selection models generate portfolios with better out-of-sample performance and that it is necessary to take sparsity of resulting optimal portfolios into account in constructing portfolio selection models for large scale portfolio optimization.

References

[1] D. Lorenzo, G. Liuzzi, F. Rinaldi, F. Schoen and M. Sciandrone, A concave optimization-based approach for sparse portfolio selection, Optimization Method and Software, 27 (2012), 983-10000.

[2] M. Broadie, Computing efficient frontiers using estimated parameters, Annals of Operation Research, 45 (1993), 21-58.

[3] E. J. Elton and M. J. Gruber, Modern portfolio theory and investment analysis, Wiley, New York, 1995.

[4] A. Frost and E. Savarino, An empirical bayes approach to efficient portfolio selection, Journal of Financial and Quantitative, 21 (1986), 293-305.

[5] F. Black and R. Litterman, Asset allocation: Combining investor views with market equilibrium, Journal of Fixed Income, 1 (1991), 7-18.

[6] R. Jagannathan and T. Ma, Risk reduction in large portfolios: Why imposing the wrong constraints helps, Journal of Finance, 58 (2003), 1651-1684.

[7] O. Ledoit and M. Wolf, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, Journal of Empirical Finance, 10 (2003), 603-621.

[8] O. Ledoit and M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, Journal of Multivariate Analysis, 88 (2004), 365-411.

[9] O. Ledoit and M. Wolf, Honey, I shrunk the sample covariance matrix, The Journal of Portfolio Management, 30 (2004), 110-119.

[10] V. DeMiguel, L. Garlappi, J. Nogales and R. Uppal, A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, Management Science, 55 (2009), 798-812.

[11] J. Brodie, I. Daubechies, C. DeMol, D. Giannone and D. Loris, Sparse and stable markowitz portfolios, ECB Working Paper Series, 2008.

[12] J. Fan, J. Zhang and Y. Ke, Asset allocation and risk assessment with gross exposure constraints for vast portfolios, Working paper, Princeton University, 2009.

[13] Y. Yen, A note on sparse minimum variance portfolios and coordinate-wise descent algorithms, Available at SSRN 1604093, 2010.

[14] M. Carrasco and N. Noumon, Optimal portfolio selection using regulaization, http://www.cireq.umontreal.ca/activites/100514/carrasco.pdf, 2010.

[15] S. Still and I. Kondor, Regularizing portfolio optimization, New Journal of Physics, 12 (2010), 075034.

[16] M. W. Brandt, S. C. Pedro and V. Rossen, Parametric portfolio policies: exploiting characteristics in the cross section of equity returns, Review of Financial Studies, 22 (2009), 3411-3447.

[17] A. E. Hoerl and R. W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems, Technometrics, 8 (1970), 27-51.

[18] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society, 46 (1996), 431-439.

[19] Z. Xu, X. Chang, F. Xu and H. Zhang, L1/2 regularization: an iterative half thresholding algorithm, IEEE Transactions on neural networks and learning systems, 23 (2012), 1013-1027.

[20] Z. Xu, H. Guo, Y. Wang, and H. Zhang, The representation of L1/2 regularizer among Lq regularizer: an experimental study based on phase diagram, 2010, Submitted.

[21] Z. Xu, H. Zhang, Y. Wang, and X. Chang, L1/2 regularization, Science in China (series F), 40 (2010), 1-11.

[22] H. Akaike, 1973. Information theory and an extension of the maximum likelihood principle, In: B N Petrov, F Caki, eds. Second International Symposium on Information Theory, Budapest: Akademiai Kiado, 1973, pp. 267-281.

[23] G. Schwarz, Estimating the dimension of a model, Annals of Statistical, 6 (1978), 461-464.

[24] Fama and K. R. French, Data library, download by $http$: $//mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html$, 1986.

[25] B. Efron, T. Hastie, T. Johnstone and R. Tibshirani, Least angle regression, Annals of Statistics, 23 (2004), 407-499.

[26] H. M. Markowitz, Portfolio selection, The Journal of Finance, 7 (1952), 77-91.