A New Regularization Method for High-dimensional Portfolio Selection

GOU ZIKANG

Institute of Statistics and Big Data

Renmin University of China

1 Introduction

In this paper, we conduct exploratory data analysis on stocks in $S\&P$ 500 index that have complete records from 2014 to 2016 and propose a new model to tackle the portfolio allocation problem.

2 Markowitz Mean-Variance Model

Portfolio allocation attracts great interest in the research of financial econometrics and quantitative finance. Markowitz (1952) proposed the mean-variance portfolio, which considers the portfolio allocation with excess returns as an optimization problem.

Suppose that we have a pool of p risky assets. Denote their (random excess) returns by $\boldsymbol{r} = (r_1, r_2, \ldots, r_p)^T$, and let $\boldsymbol{w} = (w_1, w_2, \ldots, w_p)$ be a vector of portfolio weights on the assets. Let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be the mean vector and covariance matrix of $\boldsymbol{r}$.

Then the expected return of the portfolio is $E(\boldsymbol{w^T r}) = \boldsymbol{w^T \mu}$, and the variance of the portfolio is $\sum\limits_{k=1}^{p} w_i^2 Var(r_i) + \sum\limits_{i \neq j} w_i w_j Cov(r_i, r_j) = \boldsymbol{w^T \Sigma w}$.

We have the following Mean-Variance Model:

$$\arg\min_{\boldsymbol{w}} \frac{1}{2}\boldsymbol{w^T \Sigma w} - r\boldsymbol{w^T \mu} + \lambda \|\boldsymbol{w}\|_1, \tag{1}$$

Where r is the inverse of the risk aversion parameter, λ is an ℓ_1-norm constraint tuning parameter.

3 Methodology

3.1 Constrained ℓ_1 Regularization Model

Inspired by Dantzig selector (Candes and Tao, 2007), we consider the following optimization problem for high-dimensional portfolio selection model by adding constraints to the ℓ_1 norm of $\boldsymbol{w}$.

$$\min_{\boldsymbol{w}\in\mathbb{R}^p} \|\boldsymbol{w}\|_1 \quad \text{s.t. } \|\hat{\boldsymbol{\Sigma}}\boldsymbol{w} - r\hat{\boldsymbol{\mu}}\|_\infty \leq \lambda, \quad \boldsymbol{w}^T\mathbf{1} = 1, \tag{2}$$

where $\hat{\boldsymbol{\Sigma}}$ is the sample covariance matrix of the assets' excess return and $\hat{\boldsymbol{\mu}}$ is the sample mean of the assets' excess return.

3.2 ADMM Algorithm

The minimization problem (2) is a convex minimization problem, and we try to solve it using ADMM algorithm when the number of assets in the portfolio is moderately large.
We introduce auxiliary variable $\boldsymbol{z} = \hat{\boldsymbol{\Sigma}}\boldsymbol{w} - r\hat{\boldsymbol{\mu}}$ and move the condition $\|\hat{\boldsymbol{\Sigma}}\boldsymbol{w} - r\hat{\boldsymbol{\mu}}\|_\infty \leq \lambda$ into the objective function which leads to

$$\min_{\boldsymbol{w}\in\mathbb{R}^p} \|\boldsymbol{w}\|_1 + \delta_{\boldsymbol{Z_0}}(\boldsymbol{z}) \quad \text{s.t. } \hat{\boldsymbol{\Sigma}}\boldsymbol{w} - r\hat{\boldsymbol{\mu}} = \boldsymbol{z}, \quad \boldsymbol{w}^T\mathbf{1} = 1, \tag{3}$$

Where $\boldsymbol{Z_0} = \{\boldsymbol{z} : |z_s| \leq \lambda, s = 1, \ldots, p\}$,

$$\delta_{\boldsymbol{Z_0}}(\boldsymbol{z}) = \begin{cases} 0 & if\ \boldsymbol{z} \in \boldsymbol{Z_0}, \\ +\infty & otherwise. \end{cases} \tag{4}$$

The augmented Lagrangian function for (3) is:

$$\begin{aligned} \mathbb{L}_\rho(\boldsymbol{w}, \boldsymbol{z}, \boldsymbol{\gamma}, \phi) = \|\boldsymbol{w}\|_1 + \delta_{\boldsymbol{Z_0}}(\boldsymbol{z}) + \boldsymbol{\gamma}^T(\hat{\boldsymbol{\Sigma}}\boldsymbol{w} - r\hat{\boldsymbol{\mu}} - \boldsymbol{z}) + \phi(\mathbf{1}^T\boldsymbol{w} - 1) \\ + \frac{\rho}{2}\|\hat{\boldsymbol{\Sigma}}\boldsymbol{w} - r\hat{\boldsymbol{\mu}} - \boldsymbol{z}\|_2^2 + \frac{\rho}{2}\|\mathbf{1}^T\boldsymbol{w} - 1\|_2^2. \end{aligned} \tag{5}$$

We have the update scheme as follows:

$$\begin{cases} \boldsymbol{w}^{t+1} = argmin\ \mathbb{L}_\rho(\boldsymbol{w}, \boldsymbol{z}^t; \boldsymbol{\gamma}^t, \phi^t), \\ \boldsymbol{z}^{t+1} = argmin\ \mathbb{L}_\rho(\boldsymbol{w}^{t+1}, \boldsymbol{z}; \boldsymbol{\gamma}^t, \phi^t), \\ \boldsymbol{\gamma}^{t+1} = \boldsymbol{\gamma}^t + \theta\rho(\hat{\boldsymbol{\Sigma}}\boldsymbol{w}^{t+1} - r\hat{\boldsymbol{\mu}} - \boldsymbol{z}^{t+1}), \\ \phi^{t+1} = \phi^t + \theta\rho(\mathbf{1}^T\boldsymbol{w}^{t+1} - 1). \end{cases} \tag{6}$$

Motivated by Bai et al. (2022) and Bai et al. (2019), we add a proximal term to $\boldsymbol{w}$ subproblem, $\frac{1}{2}\|\boldsymbol{w}-\boldsymbol{w}^t\|_{\boldsymbol{\tau}}^2$, where the proximal operators $\boldsymbol{\tau} > \mathbf{0}$ and $\|\boldsymbol{X}\|_{\boldsymbol{P}}^2 = \boldsymbol{X}^T\boldsymbol{P}\boldsymbol{X}$. The positive definiteness of $\boldsymbol{\tau}$ will make $\{\boldsymbol{w}^t\}$ automatically well-defined. One reasonable choice of $\boldsymbol{\tau} = \eta \boldsymbol{I} - \rho\hat{\boldsymbol{\Sigma}}^T\hat{\boldsymbol{\Sigma}}$ with $\eta > \rho\lambda_{max}(\hat{\boldsymbol{\Sigma}}^T\hat{\boldsymbol{\Sigma}})$. The update scheme for $\boldsymbol{w}$ becomes:

$$
\begin{aligned}
\boldsymbol{w}^{t+1} &= \arg\min_{\boldsymbol{w}} \|\boldsymbol{w}\|_1 + \frac{\rho}{2}\|\hat{\boldsymbol{\Sigma}}\boldsymbol{w} - r\hat{\boldsymbol{\mu}} - \boldsymbol{z}\|_2^2 + (\boldsymbol{\gamma}^T\hat{\boldsymbol{\Sigma}} + \phi\mathbf{1}^T)\boldsymbol{w} + \frac{1}{2}\|\boldsymbol{w}-\boldsymbol{w}^t\|_{\boldsymbol{\tau}}^2 \\
&= \arg\min_{\boldsymbol{w}} \|\boldsymbol{w}\|_1 + \rho(\boldsymbol{w}-\boldsymbol{w}^t)^T\boldsymbol{v}^t + \frac{\eta}{2}\|\boldsymbol{w}-\boldsymbol{w}^t\|_2^2,
\end{aligned}
\tag{7}
$$

where $\boldsymbol{v}^t = \hat{\boldsymbol{\Sigma}}^T\left(\hat{\boldsymbol{\Sigma}}\boldsymbol{w}^t - r\hat{\boldsymbol{\mu}} - \boldsymbol{z}^t + \frac{\boldsymbol{\gamma}^t}{\rho}\right) + \frac{\phi}{\rho}\mathbf{1}$. We can derive the solutions of $\boldsymbol{w}$-subproblem and $\boldsymbol{z}$-subproblem as follows:

$$
\boldsymbol{w}^{t+1} = S(\boldsymbol{w}^t - \frac{\rho}{\eta}\boldsymbol{v}^t, \frac{1}{\eta}), \tag{8}
$$

$$
\boldsymbol{z}^{t+1} = min\left(max\left(\hat{\boldsymbol{\Sigma}}\boldsymbol{w}^{t+1} - r\hat{\boldsymbol{\mu}} + \frac{\boldsymbol{\gamma}^t}{\rho}, -\lambda\right), \lambda\right), \tag{9}
$$

where $S(x,t) = sign(x)(|x|-t)I(|x|>t)$ is the soft thresholding function.
The update for $\boldsymbol{\gamma}$ and ϕ are:

$$
\boldsymbol{\gamma}^{t+1} = \boldsymbol{\gamma}^t + \theta\rho(\hat{\boldsymbol{\Sigma}}\boldsymbol{w}^{t+1} - r\hat{\boldsymbol{\mu}} - \boldsymbol{z}^{t+1}), \tag{10}
$$

$$
\phi^{t+1} = \phi^t + \theta\rho(\mathbf{1}^T\boldsymbol{w}^{t+1} - 1). \tag{11}
$$

Motivated by Boyd et al. (2010), we define:

- ◇ Primal Residual: $r^t = \hat{\boldsymbol{\Sigma}}'\boldsymbol{w}^t - r\hat{\boldsymbol{\mu}}' - \boldsymbol{z}'^t$
- ◇ Dual Residual: $s^t = \rho\hat{\boldsymbol{\Sigma}}'^T(\boldsymbol{z}'^t - \boldsymbol{z}'^{(t-1)})$

where $\boldsymbol{z}' = (\boldsymbol{z}^T, 0)^T \in \mathbb{R}^{p+1}$, $\hat{\boldsymbol{\mu}}' = (\hat{\boldsymbol{\mu}}^T, \frac{1}{r})^T \in \mathbb{R}^{p+1}$, $\hat{\boldsymbol{\Sigma}}' = (\hat{\boldsymbol{\Sigma}}^T, \mathbf{1})^T \in \mathbb{R}^{(p+1)\times p}$.
The stopping criterion is that the primal and dual residuals are small, i.e., $\|r^t\|_2 \leq \epsilon^{pri}$, $\|s^t\|_2 \leq \epsilon^{dual}$, where ϵ^{pri} and ϵ^{dual} are feasibility tolerance for primal and dual feasibility conditions. These tolerances can be chosen using an absolute and relative criterion, such as

$$
\begin{aligned}
\epsilon^{pri} &= \sqrt{p+1}\epsilon^{abs} + \epsilon^{rel} max\{\|\hat{\boldsymbol{\Sigma}}'\boldsymbol{w}^t\|_2, \|r\hat{\boldsymbol{\mu}}'\|_2, \|\boldsymbol{z}'^t\|_2\}, \\
\epsilon^{dual} &= \sqrt{p}\epsilon^{abs} + \epsilon^{rel}\|\hat{\boldsymbol{\Sigma}}'\boldsymbol{\gamma}'^t\|_2,
\end{aligned}
\tag{12}
$$

where $\boldsymbol{\gamma}' = (\boldsymbol{\gamma}^T, \phi)^T$. We summarize the approach in Algorithm 1.

Algorithm 1 ADMM Algorithm

Data: $\boldsymbol{w}^0, \boldsymbol{z}^0, \boldsymbol{\gamma}^0, \phi_0$, and $\rho > 0,\ \theta > 0$
Result: optimal $\boldsymbol{w}$

```
while the stopping citerion is not satisfied, do
    compute w^{t+1} by (2.6),
    compute z^{t+1} by (2.7),
    compute γ^{t+1} by (2.8),
    compute φ^{t+1} by (2.9).
end
```

4 Data Exploration

4.1 Simulation Data

Fama and French (1993) proposed the three-factor model for the US stock market.

$$R_i = b_{i1}f_1 + b_{i2}f_2 + b_{i3}f_3, \quad i = 1, \ldots, p. \tag{13}$$

The matrix form of this model can be presented as

$$\boldsymbol{R} = \boldsymbol{B}\boldsymbol{f} + \boldsymbol{\epsilon}, \tag{14}$$

where $\boldsymbol{B} \sim N(\boldsymbol{\mu}_b, Cov(b))$, $\boldsymbol{f} \sim N(\boldsymbol{\mu}_f, Cov(f))$, $\boldsymbol{\epsilon} \sim N(0, diag(\sigma_1^2, \ldots, \sigma_p^2))$. We use the parameters in Fan,et al (2012).

- ◇ Factor loadings and factor returns:

Table 1. Parameters used in the simulation

Parameters for factor loadings				Parameters for factor returns			
$\boldsymbol{\mu}_b$		cov_b		$\boldsymbol{\mu}_f$		cov_f	
0.7828	0.02914	0.02387	0.010184	0.02355	1.2507	−0.0350	−0.2042
0.5180	0.02387	0.05395	−0.006967	0.01298	−0.0350	0.3156	−0.0023
0.4100	0.01018	−0.00696	0.086856	0.02071	−0.2042	−0.0023	0.1930

- ◇ Standard deviations of the errors: $\sigma_1, \ldots, \sigma_p \sim G(3.3586, 0.1876)$.

We set $n = 756$, where $n_1 = 504$ (in-sample) and $n_2 = 252$ (out-of-sample), $p = 500$ and 1000 respectively, $\lambda = 0.01$, $\epsilon^{abs} = 0.001$, $\epsilon^{rel} = 0.001$, $\theta = 1.618$, $\rho = 0.25$, and $r = \frac{1}{2}$.

We compare our new model and primal model in the simulation study, the ressults are given in Table 2, Table 3, Table 4 and Table 5.

Table 2. Numerical comparions of two models when p=500 (in-sample)

Model	Mean (%)	SD (%)	Sharpe ratio	Max. Weight	Min. weight	No. of long positions	No. of short positions
L1 Regularization Model	5.43	2.69	1.991	0.127	-0.0751	108.8	62
Primal M-V Model	7.32	6.36	1.151	0.0272	-0.0153	263.6	153

Table 3. Numerical comparions (out-of-sample)

Model	Mean (%)	SD (%)	Sharpe ratio
L1 Regularization Model	5.98	3.15	1.897
Primal M-V Model	6.76	6.44	1.049

Table 4. Numerical comparions of two algorithms when p=1000 (in-sample)

Model	Mean (%)	SD (%)	Sharpe ratio	Max. Weight	Min. weight	No. of long positions	No. of short positions
L1 Regularization Model	11.61	2.55	4.533	0.031	-0.0346	336.2	199.4
Primal M-V Model	7.23	8.68	0.847	0.00633	-0.00363	666.4	318.8

As can be seen, our method has better performance in both in sample and out of sample, which has a better sharpe ratio, and selects fewer assets in the portfolio.

4.2 Real Data

- ◇ Data: stocks in $S\&P$ 500 index that have complete records from 2014 to 2016.

Table 5. Numerical comparions (out-of-sample)			
Model	**Mean (%)**	**SD (%)**	**Sharpe ratio**
L1 Regularization Model	10.53	2.63	4.003
Primal M-V Model	9.31	8.82	1.055

◇ Source: https://www.kaggle.com/datasets/gauravmehta13/sp-500-stock-prices

	symbol	date	open	high	low	close	volume
1	AAL	2014-01-02	25.0700	25.8200	25.0600	25.3600	8998943
2	AAPL	2014-01-02	79.3828	79.5756	78.8601	79.0185	58791957
3	AAP	2014-01-02	110.3600	111.8800	109.2900	109.7400	542711
4	ABBV	2014-01-02	52.1200	52.3300	51.5200	51.9800	4569061
5	ABC	2014-01-02	70.1100	70.2300	69.4800	69.8900	1148391
6	ABT	2014-01-02	38.0900	38.4000	38.0000	38.2300	4967472
7	ACN	2014-01-02	81.5000	81.9200	81.0900	81.1300	2405384
8	ADBE	2014-01-02	59.0600	59.5300	58.9400	59.2900	2746370
9	ADI	2014-01-02	49.5200	49.7500	49.0400	49.2800	2799092
10	ADM	2014-01-02	43.2200	43.2900	42.7900	42.9900	2753765
11	ADP	2014-01-02	80.1700	80.4500	79.3800	79.8600	1965869
12	ADSK	2014-01-02	49.3300	49.7400	48.8800	49.2500	2488043
13	ADS	2014-01-02	262.4400	262.6800	258.7800	262.3400	547808

Figure 1: stocks in $S\&P$ 500 index

In this section, we illustrate the proposed procedure by an empirical analysis of real world data set, which consists of daily returns for stocks in $S\&P$ 500 index that have complete records from 2014 to 2016. The raw data of stocks in $S\&P$ 500 index has 497,472 rows and 7 total columns.

After data processing, there are total 477 stocks that have complete records, 504 records for each stock from the Year 2014 to 2015 and 252 records for each stock in the Year 2016.

We partition data into training and testing data sets. The training data set consists of data from the Year 2014 to 2015, while the testing data set consists of the data in the Year 2016.

We calculate the daily return of each stock by $r_i = (Close_i - Open_i)/Open_i$, and the results are given as follows:

symbol date	A	AAL	AAP	AAPL	ABBV	ABC	ABT	ACN	ADBE
2014-01-02	-0.015587	0.011568	-0.005618	-0.004589	-0.002686	-0.003138	0.003676	-0.004540	0.003894
2014-01-03	0.009399	0.030680	0.014834	-0.021488	0.000957	0.000715	0.007037	0.002340	-0.000507
2014-01-06	-0.013240	0.015402	-0.011494	0.012057	-0.043833	-0.009241	-0.001021	-0.009470	0.001033
2014-01-07	0.008780	-0.010846	0.010716	-0.007868	-0.006884	0.009023	-0.009686	0.008412	0.012187
2014-01-08	0.018489	0.047782	-0.001512	0.008630	-0.005333	0.011517	0.009009	0.010579	-0.003721
...	...	...	...	...	...	...	...	...	...
2015-12-24	0.008375	0.012480	0.005732	-0.008899	0.001370	0.002314	0.003560	0.000096	-0.001588
2015-12-28	-0.005712	-0.008941	0.014214	-0.007157	0.008238	0.003274	0.002449	0.004245	0.004371
2015-12-29	0.006654	0.005780	0.003098	0.016642	0.007286	0.000764	0.011926	0.007847	0.007717
2015-12-30	-0.004485	-0.019248	-0.005256	-0.011604	0.006734	0.000762	-0.012650	0.005318	0.000945
2015-12-31	-0.002148	-0.004466	-0.004958	-0.016354	-0.004370	-0.007275	-0.004434	-0.008539	-0.011886

Figure 2: return of stocks from 2014 to 2015

symbol date	A	AAL	AAP	AAPL	ABBV	ABC	ABT	ACN	ADBE
2016-01-04	-0.009011	-0.009203	0.020034	0.026703	-0.007751	-0.004301	-0.022986	-0.007698	0.002179
2016-01-05	-0.004419	-0.016027	-0.007548	-0.028747	-0.010691	0.011350	-0.000931	0.003825	0.003587
2016-01-06	0.012177	0.029463	-0.016569	0.001392	0.017917	0.003154	0.005909	0.013392	-0.000878
2016-01-07	-0.028401	-0.012692	0.027831	-0.022598	0.012208	-0.023198	-0.003359	-0.005915	-0.003244
2016-01-08	-0.016063	-0.013923	-0.023284	-0.016134	-0.033015	-0.012102	-0.028428	-0.012867	-0.020406
...	...	...	...	...	...	...	...	...	...
2016-12-23	0.002604	-0.002674	0.001172	0.008046	0.009064	0.006663	0.002348	0.000681	0.003919
2016-12-27	0.003019	0.006627	0.006560	0.006351	-0.000320	0.005101	0.003119	0.001619	-0.000190
2016-12-28	-0.016986	-0.021753	-0.009244	-0.006467	-0.003839	-0.006467	-0.010614	-0.011780	-0.010112
2016-12-29	-0.000876	-0.011506	0.001176	0.002404	0.005450	0.009286	0.000261	0.000256	0.000482
2016-12-30	-0.004371	-0.015394	-0.012841	-0.007115	-0.001738	-0.014619	0.002087	-0.003658	-0.010762

Figure 3: return of stocks in 2016

We compare our new model and primal model in the simulation study, the ressults are given in Table 6 and Table 7.

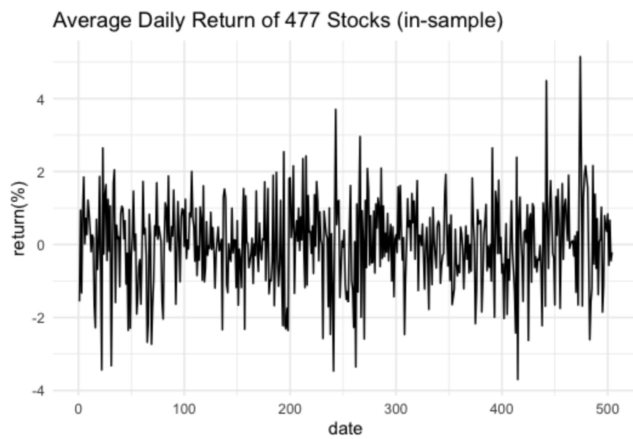

Figure 4: average daily return of 477 stocks from 2014 to 2015

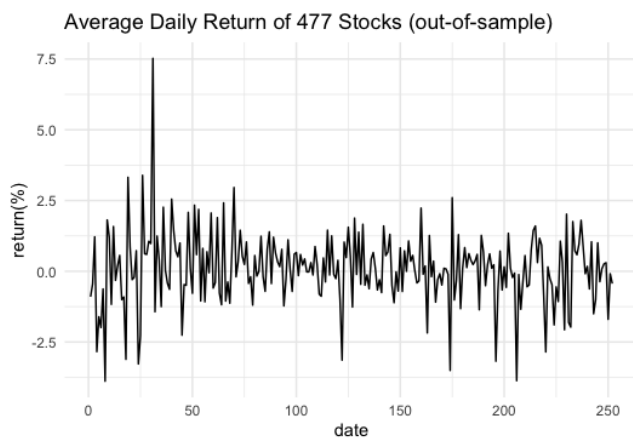

Figure 5: average daily return of 477 stocks in 2016

As can be seen, our method has better performance in both in sample and out of sample, which has a better sharpe ratio, and selects fewer assets in the portfolio.

5 Discussion

Our method has better performance in both simulation and empirical analysis, it has a better sharpe ratio, and selects fewer assets in the portfolio.
For our future work, one is the choice of the parameter, especially the ℓ_1-norm constraint tuning parameter λ; the other is to find a better algorithm for our model besides ADMM, in order to tackle the problem more efficiently when the dimension of the data

Table 6. Numerical comparions of two models (in-sample)							
Algorithm	**Mean (%)**	**SD (%)**	**Sharpe ratio**	**Max. Weight**	**Min. weight**	**No. of long positions**	**No. of short positions**
L1 Regularization Model	17.84	6.58	2.71	0.0972	-0.096	70	39
Primal M-V Model	12.79	8.53	1.5	0.0132	-0.00742	252	10

Table 7. Numerical comparions (out-of-sample)			
Algorithm	**Mean (%)**	**SD (%)**	**Sharpe ratio**
L1 Regularization Model	24.69	10.48	2.36
Primal M-V Model	17.52	8.49	2.06

is extremely large.

References

Candes, E. and Tao, T. (2007) The Dantzig selector: Statistical estimation when p is much larger than n *Annals of Statistics* 35(6), 2313 – 2351.

Bai, J., Hager, W. W., and Zhang, H. (2022) An inexact accelerated stochastic admm for separable convex optimization *Computational Optimization and Applications* 81(2), 479 — 518.

Bai, J., Liang, J., Guo, K., and Jing, Y. (2019) ccelerated symmetric admm and its applications in signal processing *arXiv preprint arXiv* 1906.12015.

Boyd, S. and Parikh, N. and Chu, E. and Peleato, B. and Eckstein, J. (2010) Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers *Foundations & Trends in Machine Learning* 3(1), 1 – 122.

Fan, J., Zhang, J., and Yu, K. (2012) ast portfolio selection with gross-exposure constraints Journal of the American Statistical Association *FJournal of the American Statistical Association* 107(498), 592 – 606.