A New Regularization Method for High-dimensional Portfolio Selection

Zikang Gou

Institute of Statistics and Big Data
Renmin University of China

April 19, 2023

Presentation Overview

Markowitz Mean-Variance Model

Suppose that we have a pool of N risky assets. Denote their (random excess) returns by $\boldsymbol{r} = (r_1, r_2, ..., r_N)^T$, and let $\boldsymbol{w} = (w_1, w_2, , w_N)$ 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 $\boldsymbol{E(w^T r) = w^T \mu}$, and the variance of the portfolio is $\sum_{k=1}^{N} w_i^2 Var(r_i) + \sum_{i \neq j} w_i w_j Cov(r_i, r_j) = \boldsymbol{w^T \Sigma w}$.

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.

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 $\boldsymbol{w}$.

Constrained ℓ_1 Regularization Model

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

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.

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 $\boldsymbol{z} = \hat{\boldsymbol{\Sigma}}\boldsymbol{w} - r\hat{\mu}$ and move the condition $\|\hat{\boldsymbol{\Sigma}}\boldsymbol{w} - r\hat{\mu}\|_\infty \leq \lambda$ into the objective function which leads to

Model

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

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

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

ADMM Algorithm

The augmented Lagrangian function for (3) is:

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

We have the update scheme as follows:

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

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:

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

where $\boldsymbol{v}^t = \hat{\boldsymbol{\Sigma}}^T\left(\hat{\boldsymbol{\Sigma}}\boldsymbol{w}^t - r\hat{\boldsymbol{\mu}} - \boldsymbol{z}^t + \frac{\gamma^t}{\rho}\right) + \frac{\phi}{\rho}\mathbf{1}$.

We can derive the solutions of $\boldsymbol{w}$-subproblem and $\boldsymbol{z}$-subproblem as follows:

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

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

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

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

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

ADMM Algorithm
Stopping criteria

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.,

$$\|r^t\|_2 \le \epsilon^{pri}, \ \|s^t\|_2 \le \epsilon^{dual},$$

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

ADMM Algorithm

$$\epsilon^{pri} = \sqrt{p+1}\epsilon^{abs} + \epsilon^{rel} max\{\|\hat{\mathbf{\Sigma}}'\mathbf{w}^t\|_2, \|r\hat{\boldsymbol{\mu}}'\|_2, \|\mathbf{z}'^t\|_2\},$$

$$\epsilon^{dual} = \sqrt{p}\epsilon^{abs} + \epsilon^{rel}\|\hat{\mathbf{\Sigma}}'\gamma'^t\|_2,$$

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

Algorithm 1 ADMM for Model

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

Result: optimal $\boldsymbol{w}$

```
1 while the stopping citerion is not satisfied, do
2     compute w^{t+1} by (8),
3     compute z^{t+1} by (9),
4     compute γ^{t+1} by (10),
5     compute φ^{t+1} by (11).
6 end
```

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

Three Factor Model

$$R_i = b_{i1}f_1 + b_{i2}f_2 + b_{i3}f_3, \quad i = 1, \ldots, p. \tag{12}$$

The matrix form of this model can be presented as

$$\boldsymbol{R} = \boldsymbol{Bf} + \epsilon. \tag{13}$$

- $\boldsymbol{B} \sim N(\mu_b, Cov(b))$
- $\boldsymbol{f} \sim N(\mu_f, Cov(f))$
- $\epsilon \sim N(0, diag(\sigma_1^2, \ldots, \sigma_p^2))$

Simulation Analysis

data

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: $\sigma_1, \ldots, \sigma_p \sim G(3.3586, 0.1876)$

parameters

n=756

- $n_1 = 504$ (in-sample)
- $n_2 = 252$ (out-of-sample)

$p = 500/1000$
$\lambda = 0.01$
$\epsilon^{abs} = 0.001$
$\epsilon^{rel} = 0.001$
$\theta = 1.618$
$\rho = 0.25$
$r = \frac{1}{2}$

We compare ADMM and lpSolve in the simulation study.

Result

p=500

Table 1. Numerical comparions of two algorithms when p=500 (in-sample)							
Algorithm	**Mean (%)**	**SD (%)**	**Sharpe ratio**	**Max. Weight**	**Min. weight**	**No. of long positions**	**No. of short positions**
ADMM	5.43	2.69	1.991	0.127	-0.0751	108.8	62
lpSolve	7.32	6.36	1.151	0.0272	-0.0153	263.6	153

Table 2. Numerical comparions (out-of-sample)			
Algorithm	**Mean (%)**	**SD (%)**	**Sharpe ratio**
ADMM	5.98	3.15	1.897
lpSolve	6.76	6.44	1.049

Result

p=1000

Table 3. Numerical comparions of two algorithms when p=1000 (in-sample)							
Algorithm	**Mean (%)**	**SD (%)**	**Sharpe ratio**	**Max. Weight**	**Min. weight**	**No. of long positions**	**No. of short positions**
ADMM	11.61	2.55	4.533	0.031	-0.0346	336.2	199.4
lpSolve	7.23	8.68	0.847	0.00633	-0.00363	666.4	318.8

Table 4. Numerical comparions (out-of-sample)			
Algorithm	**Mean (%)**	**SD (%)**	**Sharpe ratio**
ADMM	10.53	2.63	4.003
lpSolve	9.31	8.82	1.055

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

497,472 rows, 7 total columns

Empirical Analysis
data

- in-sample: return of stocks from 2014 to 2015

symbol	A	AAL	AAP	AAPL	ABBV	ABC	ABT	ACN
date								
2014-01-02	-0.015587	0.011568	-0.005618	-0.004589	-0.002686	-0.003138	0.003676	-0.004540
2014-01-03	0.009399	0.030680	0.014834	-0.021488	0.000957	0.000715	0.007037	0.002340
2014-01-06	-0.013240	0.015402	-0.011494	0.012057	-0.043833	-0.009241	-0.001021	-0.009470
2014-01-07	0.008780	-0.010846	0.010716	-0.007868	-0.006884	0.009023	-0.009686	0.008412
2014-01-08	0.018489	0.047782	-0.001512	0.008630	-0.005333	0.011517	0.009009	0.010579
...	...	...	...	...	...	...	...	...
2014-12-24	-0.007001	0.015199	-0.007495	-0.005063	0.022548	-0.000546	-0.000875	-0.001749

n=504, p=477

Empirical Analysis

data

- out-of-sample: return of stocks from 2016

symbol	A	AAL	AAP	AAPL	ABBV	ABC	ABT	ACN
date								
2016-01-04	-0.009011	-0.009203	0.020034	0.026703	-0.007751	-0.004301	-0.022986	-0.007698
2016-01-05	-0.004419	-0.016027	-0.007548	-0.028747	-0.010691	0.011350	-0.000931	0.003825
2016-01-06	0.012177	0.029463	-0.016569	0.001392	0.017917	0.003154	0.005909	0.013392
2016-01-07	-0.028401	-0.012692	0.027831	-0.022598	0.012208	-0.023198	-0.003359	-0.005915
2016-01-08	-0.016063	-0.013923	-0.023284	-0.016134	-0.033015	-0.012102	-0.028428	-0.012867
...	...	...	...	...	...	...	...	...
2016-12-23	0.002604	-0.002674	0.001172	0.008046	0.009064	0.006663	0.002348	0.000681

n=252, p=477

Result

Table 5. Numerical comparions of two algorithms (in-sample)							
Algorithm	**Mean (%)**	**SD (%)**	**Sharpe ratio**	**Max. Weight**	**Min. weight**	**No. of long positions**	**No. of short positions**
ADMM	17.84	6.58	2.71	0.0972	-0.096	70	39
lpSolve	12.79	8.53	1.5	0.0132	-0.00742	252	10

Table 6. Numerical comparions (out-of-sample)			
Algorithm	**Mean (%)**	**SD (%)**	**Sharpe ratio**
ADMM	24.69	10.48	2.36
lpSolve	17.52	8.49	2.06

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.

Future work:

- The choice of λ
- Better algorithm of our model: AS-ADMM

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.