Portfolio Selection via Text Based Network

Cheng Lu

Advisor: Prof. Majeed Simaan

Stevens Institute of Technology

May 2020

Problem Description

To construct a Mean-Variance optimal portfolio (Markowitz (1952)[9]), vector of mean returns μ and covariance matrix Σ are needed.

While the true parameters are unavailable, estimation from sample data would lead to poor out-of-sample performance because of estimation. (Demiguel (2009)[1])

Especially, when it comes to high dimension this estimation error became more significant. (Michaud (1989)[10]) Also, estimation of covariance matrix is challenging due to the curse of dimension.

Solution to estimation error

Large literature dealing with estimation error problem.

- Klein and Bawa (1976)[5] used Bayesian approaches with diffuse priors.
- Goldfarb and Iyengar (2003)[3] adopt robust optimization methods.
- Ledoit and Wolf (2004)[6] proposed linear shrinkage method.
- Ledoit and Wolf (2012)[7] extended linear shrinkage to non-linear shrinkage method.

Our project would focus on the linear shrinkage method.

Linear shrinkage

- Linear shrinkage is firstly **proposed** by Stein (1956)[11]. He argued that under high dimension a better estimator than the sample mean can be constructed by linearly combining sample mean and zero vector.
- Efron and Morris (1973, 1975, 1977)[2] **improved** shrinkage method by providing the suggestion of identity vector as alternative shrinkage targets.
- Ledoit and Wolf (2004)[6] **extend** Stein's shrinkage estimation of the mean vector to the estimation of the covariance matrix.
- In the sense of the mean squared error (MSE), shrinkage is a classic example of a bias-variance tradeoff.[8]

Shrinkage target : TBN

- Proposed by Hoberg and Phillips (2016)[4], Text-Based Network(TBN) is a square correlation matrix describing industries boundaries. This correlation matrix is created by parsing 10-K report of each firm and compute their similarity.

Better designed shrinkage target would lead to better performance. We choose TBN as shrinkage target because of several advantages.

- TBN is updated annually. Also it doesn't change dramatically between each year. This low volatility would hence reduce the shrinkage estimation error.
- TBN utilizes text data and add new information to the estimator making it further to the true covariance matrix.

Problem Formulation : Correlation Shrinkage

In linear shrinkage method, we consider the performance of Global Minimum Variance Portfolio(GMVP) $\mathbf{x}(\alpha)$, which is a function of shrinkage intensity α. To construct shrank GMVP we need to shrink covariance matrix at first.

$$\tilde{\mathbf{H}}_t = \mathbf{D}_t \tilde{\mathbf{R}}_t \mathbf{D}_t \tag{1}$$

$$= \mathbf{D}_t \left[(1-\alpha)\mathbf{R}_t + \alpha \mathring{\mathbf{R}}_t \right] \mathbf{D}_t \tag{2}$$

$$= (1-\alpha)\mathbf{H}_t + \alpha \mathbf{D}_t \mathring{\mathbf{R}}_t \mathbf{D}_t \tag{3}$$

$\mathbf{D}_t$ is the diagonal matrix of volatilities and $\tilde{\mathbf{R}}_t$ is shrank correlation matrix. $\mathbf{R}_t$ is stock correlation matrix and $\mathring{\mathbf{R}}_t$ is Text-based Network(TBN). We get shrank covariance matrix $\mathbf{H}_t$ by shrinking stock correlation.

Problem Formulation : Correlation Shrinkage

With shrank covariance matrix $\mathbf{H}_t$, we can easily get GMV Portfolio for each period $\mathbf{x}(\alpha)$.

$$\mathbf{x}_t(\alpha) = \frac{\tilde{\mathbf{H}}_t^{-1}\mathbf{1}}{\mathbf{1}'\tilde{\mathbf{H}}_t^{-1}\mathbf{1}} \tag{4}$$

$$= \frac{\mathbf{D}_t^{-1}\left[(1-\alpha)\mathbf{R}_t + \alpha\mathring{\mathbf{R}}_t\right]^{-1}\mathbf{D}_t^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{D}_t^{-1}\left[(1-\alpha)\mathbf{R}_t + \alpha\mathring{\mathbf{R}}_t\right]^{-1}\mathbf{D}_t^{-1}\mathbf{1}} \tag{5}$$

Obviously, GMV Portfolio and its performance is a function of shrinkage intensity α. How to decide the optimal value for the shrinkage intensity α is the second crucial problem.

Problem Formulation : Reinforcement Learning

We propose to use Reinforcement Learning (RL) to control the shrinkage intensity α. The concrete problem is defined as following.

- State space $S = \{r_{p,t}\}$
- Action space $A = [0, 1]$
- Reward $R_t = \frac{\mathbf{E}[r_p - r_f]}{\sigma[r_p - r_f]}$
- Objective function $\max_{\pi} E_{\pi}\left[R_1 + \gamma R_2 + \cdots + \gamma^{T-1} R_T\right]$

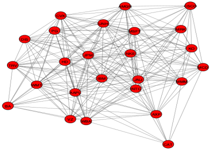
Since action space is continuous, policy gradient methods works better for learning the optimal policy.

$$J(\theta) = E_{\pi_\theta}\left[R_1 + \gamma R_2 + \cdots + \gamma^{T-1} R_T\right]$$

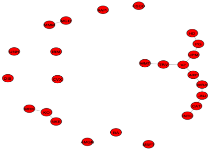
Optimal policy is achieving by updating policy parameter θ_t.

$$\theta_{t+1} = \theta_t + \alpha \widehat{\nabla J(\theta_t)}$$

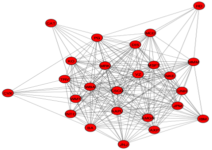
Text Based Network Graph



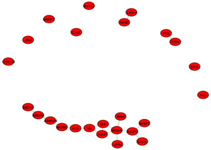
(a) year 1996 with 0.3 threshold



(b) year 2008 with 0.5 threshold

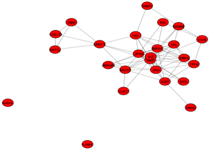


(c) year 1996 with 0.5

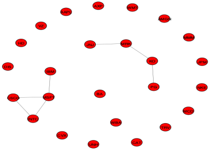


(d) year 2008 with 0.7

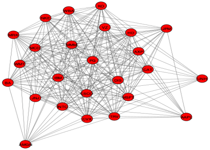
Stock Correlation Graph



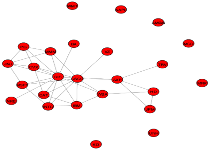
(e) year 1996 with 0 threshold



(f) year 2008 with 0 threshold



(g) year 1996 with 0.1 threshold



(h) year 2008 with 0.1 threshold

Optimal Shrinkage Intensity

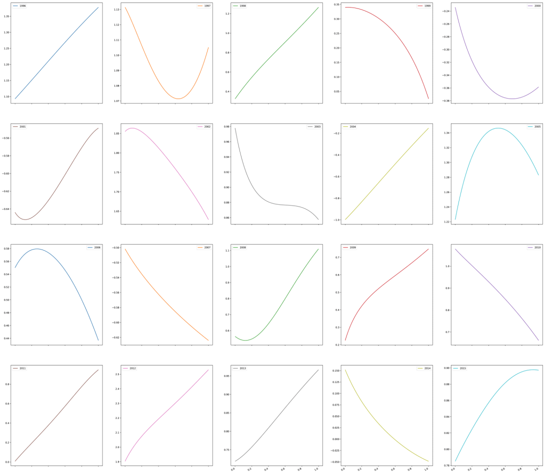


FIGURE – Out-of-sample Sharpe ratio SR(alpha) on years from 1996 to 2015

Optimal Shrinkage Intensity

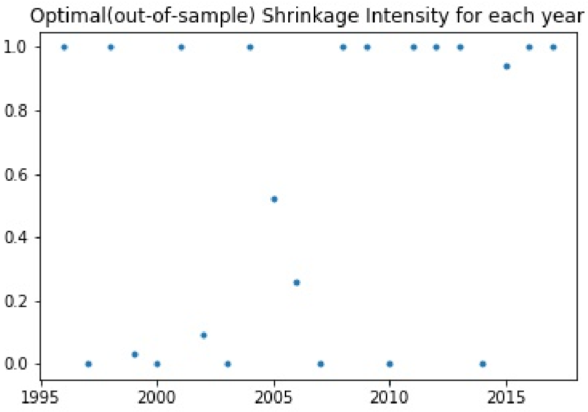


FIGURE – Optimal(out-of-sample) Shrinkage Intensity for each year

Future Work

- Get optimal policy of shrinkage from performing RL. Compare it with preliminary shrinkage intensity result. Investigate RL's contribute to the shrinkage method.
- Move beyond the linear shrinkage method to non-linear shrinkage. Explore the contribution of non-linear shrinkage method.
- Move beyond the Text-Based Network. Investigating the contribution of shrinkage target.

References I

Victor DeMiguel, Lorenzo Garlappi et Raman Uppal. "Optimal versus naive diversification : How inefficient is the 1/N portfolio strategy ?" In : *Review of Financial Studies* 22.5 (2009), p. 1915-1953.

Bradley Efron et Carl Morris. "Stein's estimation rule and its competitors—an empirical Bayes approach". In : *Journal of the American Statistical Association* 68.341 (1973), p. 117-130.

Donald Goldfarb et Garud Iyengar. "Robust portfolio selection problems". In : *Mathematics of operations research* 28.1 (2003), p. 1-38.

Gerard Hoberg et Gordon Phillips. "Text-based network industries and endogenous product differentiation". In : *Journal of Political Economy* 124.5 (2016), p. 1423-1465.

Roger W Klein et Vijay S Bawa. "The effect of estimation risk on optimal portfolio choice". In : *Journal of financial economics* 3.3 (1976), p. 215-231.

Olivier Ledoit et Michael Wolf. "A well-conditioned estimator for large-dimensional covariance matrices". In : *Journal of multivariate analysis* 88.2 (2004), p. 365-411.

References II

Olivier Ledoit, Michael Wolf et al. "Nonlinear shrinkage estimation of large-dimensional covariance matrices". In : *The Annals of Statistics* 40.2 (2012), p. 1024-1060.

Olivier Ledoit et Michael Wolf. "The power of (non-) linear shrinking : a review and guide to covariance matrix estimation". In : *University of Zurich, Department of Economics, Working Paper* 323 (2020).

Harry Markowitz. "The utility of wealth". In : *Journal of political Economy* 60.2 (1952), p. 151-158.

Richard O Michaud. "The Markowitz optimization enigma : is' optimized'optimal ?" In : *Financial Analysts Journal* 45.1 (1989), p. 31-42.

Charles Stein. *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*. Rapp. tech. Stanford University Stanford United States, 1956.