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.
- Estimation lead to poor out-of-sample performance (Demiguel (2009)[1])
- Estimation error is more of an issue for high dimensional systems. (Michaud (1989)[10])
- Estimation of covariance matrix is challenging also 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] adopted **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].
- 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.

Shrinkage target : TBN

- 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

- We consider the performance of Global Minimum Variance Portfolio(GMVP) $\mathbf{x}(\alpha)$

- To construct shrank GMVP we need to shrink covariance matrix at first

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

- $\mathbf{H}_t$ is covariance matrix
- $\tilde{\mathbf{H}}_t$ shrank covariance matrix

Problem Formulation : Correlation Shrinkage

- We get shrank covariance matrix $\mathbf{H}_t$ by shrinking stock correlation.

$$\begin{aligned} \tilde{\mathbf{H}}_t &= \mathbf{D}_t \tilde{\mathbf{R}}_t \mathbf{D}_t && (2) \\ &= \mathbf{D}_t \left[(1-\alpha)\mathbf{R}_t + \alpha \mathring{\mathbf{R}}_t \right] \mathbf{D}_t && (3) \\ &= (1-\alpha)\mathbf{H}_t + \alpha \mathbf{D}_t \mathring{\mathbf{R}}_t \mathbf{D}_t && (4) \end{aligned}$$

- $\mathbf{D}_t$ is the diagonal matrix of volatilities
- $\tilde{\mathbf{R}}_t$ is shrank correlation matrix
- $\mathbf{R}_t$ is stock correlation matrix
- $\mathring{\mathbf{R}}_t$ is Text-based Network(TBN)

Problem Formulation : Correlation Shrinkage

With shrank covariance matrix $\mathbf{H}_t$, GMV Portfolio $\mathbf{x}(\alpha)$ can be built for each period

$$\mathbf{x}_t(\alpha) = \frac{\tilde{\mathbf{H}}_t^{-1}\mathbf{1}}{\mathbf{1}'\tilde{\mathbf{H}}_t^{-1}\mathbf{1}} \tag{5}$$

$$= \frac{\mathbf{D}_t^{-1}\left[(1-\alpha)\mathbf{R}_t + \alpha\mathring{\mathbf{R}}_t\right]^{-1}\mathbf{D}_t^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{D}_t^{-1}\left[(1-\alpha)\mathbf{R}_t + \alpha\mathring{\mathbf{R}}_t\right]^{-1}\mathbf{D}_t^{-1}\mathbf{1}} \tag{6}$$

- GMV Portfolio and its performance is a function of shrinkage intensity α
- Deciding the optimal shrinkage intensity α is the second crucial problem.

Problem Formulation : Reinforcement Learning

Deploy 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 $\max_{\pi} E_{\pi}\left[R_1 + \gamma R_2 + \cdots + \gamma^{T-1} R_T\right]$

Problem Formulation : Reinforcement Learning

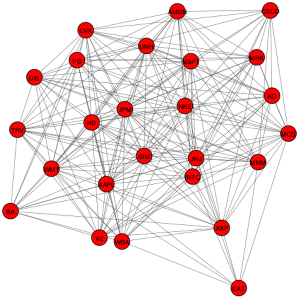
Since action space is continuous, policy gradient methods works better for learning the optimal policy.

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

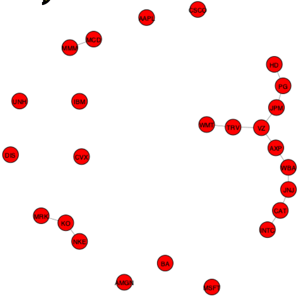
Optimal policy is achieving by updating policy parameter θ_t.

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

Text Based Network Graph

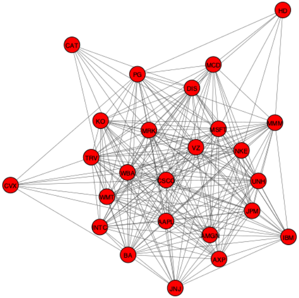


(a) year 1996 with 0 threshold

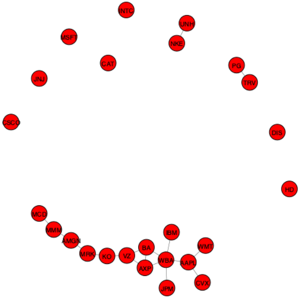


(b) year 1996 with 0.1 threshold

Text Based Network Graph

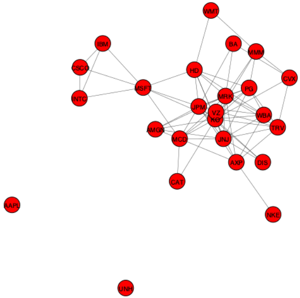


(c) year 2008 with 0 threshold

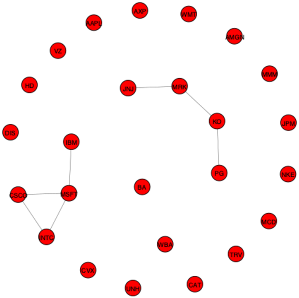


(d) year 2008 with 0.1 threshold

Stock Correlation Graph

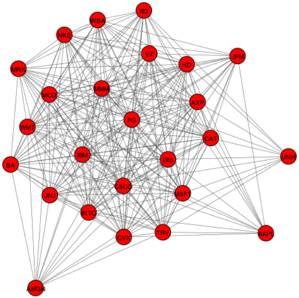


(e) year 1996 with 0.3 threshold

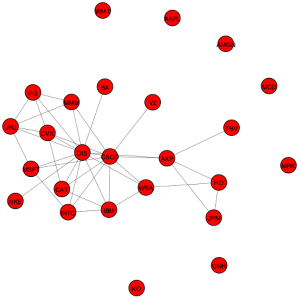


(f) year 1996 with 0.5 threshold

Stock Correlation Graph



(g) year 2008 with 0.5 threshold



(h) year 2008 with 0.7 threshold

Preliminary Results

- Shrinkage Intensity Matters.

Optimal Shrinkage Intensity

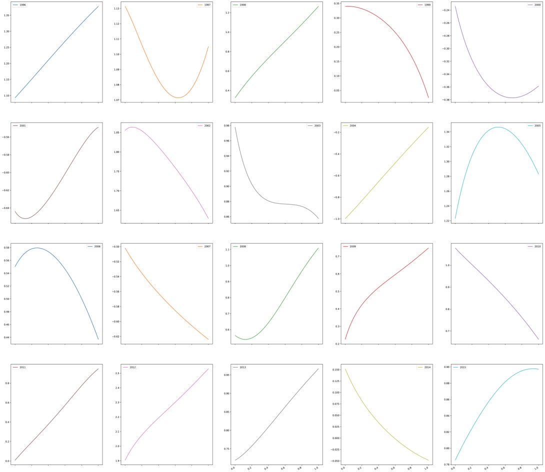


FIGURE – Out-of-sample Sharpe ratio SR(alpha) on years from 1996 to 2015

Optimal Shrinkage Intensity

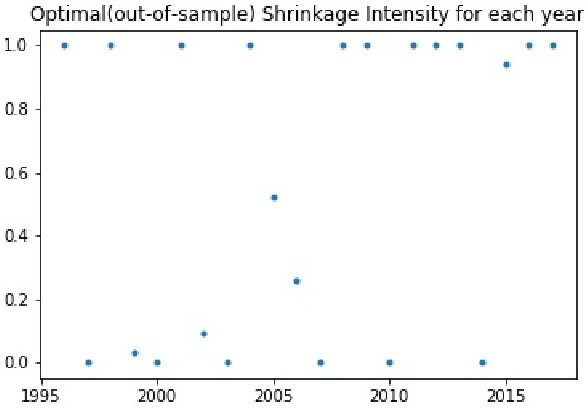


FIGURE – Optimal(in-sample) Shrinkage Intensity for each year

- Shrinkage Target
  - Text-based Network(TBN)
  - Scaled TBN
  - News-based Network
- Shrinkage Methods
  - (Non)-linear Shrinkage (Bench mark)
  - Naive Approach (Bench mark)
  - Reinforcement Learning Control

Extension on Shrinkage Target

- **Text-based Network** is proposed by Hoberg and Phillips (2016)[4]. Correlation matrix created from 10-K reports using cosine similarity.

- **Scaled TBN** is transformation of TBN. Changing the distribution of TBN. Trying to over[illegible]e TBN's drawback.

- **News-based Network** is a new method of constructing correlation matrix utilizing news data. Having more flexibility of updating frequency.

Comparison of three different targets

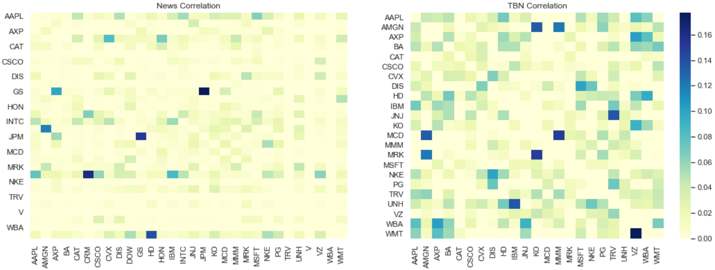


FIGURE – Heat map for News correlation and TBN

Comparison of three different targets

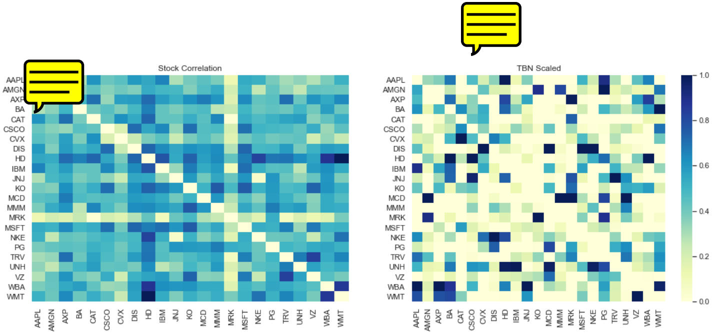


FIGURE – Heat map for Stock correlation and scaled TBN

Extension on Shrinkage Methods

- Reinforcement Learning Control
  - Deep Q Network(DQN)
  - REINFORCE
- Bench Mark
  - Linear Shrinkage[7]
  - Non-linear Shrinkage[7]
  - Naive Approach

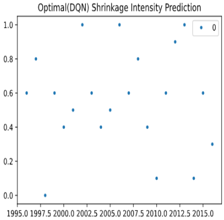
Reinforcement Learning Control

- Deep Q Network(DQN)

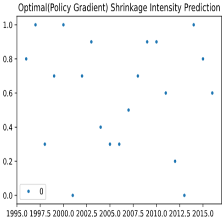
$$\mathcal{L}(w) = \mathbb{E}_{s,a,r,s'\sim D}\left[\left(r + \gamma \max_{a'} Q\left(s', a'; w\right) - Q\left(s, a; w\right)\right)^2\right]$$

- REINFORCE
    - Objective function : $J(\theta) = E_{\pi_\theta}\left[R_1 + \gamma R_2 + \cdots + \gamma^{T-1} R_T\right]$
    - updating rule : $\theta_{t+1} = \theta_t + \alpha \widehat{\nabla J\left(\theta_t\right)}$

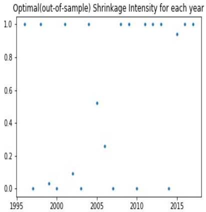
RL Performance



(a) DQN controlled intensity



(b) PG controlled intensity



(c) Optimal intensity

Performance(SR) Table

	TBN	Scaled TBN	Identity
shrink 0 pct	0.444084	0.444084	0.444084
shrink 50 pct	0.349595	-0.130572	0.507769
shrink 100 pct	-0.430668	0.387176	0.573233
linear shrinkage	-	-	0.471745
non-linear	-	-	0.449374
OAS	-	-	0.459414
DQN	0.122987	-0.286750	0.506485
REINFORCE	-0.490042	0.185471	0.503571

TABLE – Performance(SR) of 8 methods against 3 shrinkage targets