A Regularized K-means Autoregressive Shrinkage Method to Portfolio Optimization

Xinyu Li

School of Management of Technology, École Polytechnique Fédérale de Lausanne, Vaud, Switzerland

xinyu.li@epfl.ch

Abstract. Estimation is always an issue of portfolio optimization. To address this issue this paper started from addressing global minimum variance (GMV) portfolio, which gives the weights as the function of estimated covariance matrix of the asset net returns. The inverse operation of the matrix always leads to ill-conditioned matrix due to the noise in the return series. Therefore, regularized sample covariance matrix is often used to address this problem. In this work, a regularized autoregressive shrinkage method for covariance matrix estimation with the historical data is developed, which regards the future covariance matrix as the function of the historical covariance matrix. K-means clustering algorithm is also added to enhance the performance of the regularized autoregressive model. The results demonstrate that the performance of GMV portfolio can be enhanced with regularization under autoregressive shrinkage model.

Keywords: Optimization, Regularization, Covariance Matrix, Autoregressive, K-means.

1. Introduction

The ever-changing nature of financial markets makes investment difficult, especially under the background of the ramification of COVID-19 which exacerbated market volatility, Therefore the portfolio optimization attracts investors, analysts and researchers to accomplish higher possible return with less exposure to the risk.

The initial step solving this problem was proposed by Markowitz with an introduction of mean-variance (MV) portfolio [1]. Later, different portfolio such as global minimum-variance (GMV) portfolio has been discussed. The most important part of those portfolio optimizations is estimation of covariance matrix based on market data. However, it well known that errors in estimates of returns have larger impact on portfolio weights than estimates of covariance matrix [2, 3], and therefore mean-variance portfolio tends to perform worse than GMV or Sharpe ratio portfolio. However, even the performance of GMV is highly dependent on estimates of covariance matrix. As DeMiguel [4] pointed out, estimation error associated with sample covariance matrix can still be substantial.

In the literature, several approaches have been proposed to solve such problems. One approach is to apply norm constraints to regularize the optimization problem [4]. Researchers also start from maximum likelihood function and adding norm constraints for estimating covariance matrix [5]. As for the parameter of regularization, cross-validation approach is applied to derive a robust parameter [6]. The second approach is to model the covariance matrix as a time series and estimate through minimizing the loss function. Based on the traditional shrinkage method which used identity matrix as shrinkage target [7], instead of minimizing the total estimation loss, estimation errors of all uncertainty variables are analyzed [8, 9]. The third approach is to consider the uncertainty of estimation of mean and covariance. A confidence set is built for uncertainty of estimation of mean and the optimization problem can be solved in polynomial time for a large range of objective functions [10]. What's more, an approximation method of sample variance of estimated variance is proposed, to regularize the optimization problem with lower estimation uncertainty. They also propose performance-based regularization method, determining the constraints threshold using cross validation [11].

This paper also developed new regularization methods of covariance matrix estimation. Following the idea proposed in the paper of Tong et al. [12], this paper starts from the shrinkage method

proposed in Ledoit of covariance matrix estimation [7]. But different from only using the sample covariance matrix as the base, the past covariance matrix is also taken into consideration. What's more, it utilizes K-means clustering algorithm to construct multi-shrinkage target. Therefore, autoregressive shrinkage models with single or multi shrinkage target are developed, with past covariance matrix and shrinkage parameters. In order to keep robustness, based on the autoregressive shrinkage models, regularization functions are added to the loss function. However, rather than treating weight as objects of interest as introduced in the paper of DeMiguel et al. [4] or focusing on covariance matrix itself, this paper uses l_2 norm as penalty function to the autoregressive coefficients. This paper managed to derive the closed-form solution to the model and applied such frameworks on solving the traditional minimum variance problem.

The contributions to optimal portfolio choice are in several ways. First, a new shrinkage method estimating covariance matrix is proposed. This paper introduces regularized autoregressive model to the shrinkage model and add k-means clustering algorithm to further capture more features of covariance matrix as well as improve the robustness of estimation and portfolio performance. Secondly, this paper shows that k-means clustering method can help capture the feature of correlation relationship among assets through grouping the assets with the patterns of the historical return series.

This paper is organized as follows. In section 2, both single and multi-target autoregressive shrinkage model are analyzed theoretically. The closed form solutions of those models are also provided. In section 3, estimation formula as well as the algorithm of k-means autoregressive shrinkage model based on the past data is derived based on section 2. Empirical tests of both simple and k-means autoregressive shrinkage model have been carried out and analyzed

2. Methods

2.1. AR(1) shrinkage model

The process of return series can be decomposed in the combination of both the random noises and regular patterns. With mature theoretical foundation, ARIMA now is a usual model to deal with time series data for predicting short-time result [13]. Therefore, this paper will introduce the AR (autoregressive) model to the shrinkage model for covariance matrix estimation and first start with the simplest case, the AR(1) shrinkage model.

$$\Sigma_t = \beta_t \Sigma_{t-1} + \alpha_t I + \epsilon_t \tag{1}$$

where Σ_t and Σ_{t-1} is covariance matrix at t and t-1 respectively, $I \in \mathbb{R}^{P\times P}$ is identity matrix, also known as the shrinkage target in Ledoit & Wolf [7], and $\epsilon_t \in \mathbb{R}^{P\times P}$ is residual matrix. Therefore, in order to solve for the coefficients β_t and α_t, it need to minimize the loss functions:

$$\mathcal{L}(\beta_t, \alpha_t) = \left|\left|\Sigma_t - \beta_t \Sigma_{t-1} - \alpha_t I\right|\right|_F^2 \tag{2}$$

where $||\cdot||_F$ denotes Frobenius norm. The loss function (2) can be written in quadratic function of the coefficients:

$$\mathcal{L}(\beta_t, \alpha_t) = (\beta_t \quad \alpha_t) A_0 \begin{pmatrix} \beta_t \\ \alpha_t \end{pmatrix} - 2(\beta_t \quad \alpha_t) b_0 + tr(\Sigma^2) \tag{3}$$

where the A_0 and B_0 is:

$$A_0 = \begin{bmatrix} tr(\Sigma_{t-1}^2) & tr(\Sigma_{t-1}) \\ tr(\Sigma_{t-1}) & tr(I) \end{bmatrix} \tag{4}$$

$$b_0 = \begin{pmatrix} tr(\Sigma_t \Sigma_{t-1}) \\ tr(\Sigma_t) \end{pmatrix} \tag{5}$$

where $\mathrm{tr}(\cdot)$ is the trace matrix. Because A_0 is semi-definite positive matrix, it can find the coefficients through solving the convex optimization problem (2) above with letting $\frac{\partial \mathcal{L}}{\partial \beta_t} = \frac{\partial \mathcal{L}}{\partial \alpha_t} = 0$ in (3), which leads to

$$\begin{cases} \frac{tr(\Sigma_{t-1}^2)}{tr(\Sigma_{t-1}\Sigma_t)}\beta_t + \frac{tr(\Sigma_{t-1})}{tr(\Sigma_{t-1}\Sigma_t)}\alpha_t = 1 \\ \frac{tr(\Sigma_{t-1})}{tr(\Sigma_t)}\beta_t + \frac{tr(I)}{tr(\Sigma_t)}\alpha_t = 1 \end{cases} \tag{6}$$

The coefficients can be obtained through solving (6):

$$\begin{pmatrix} \beta_t^* \\ \alpha_t^* \end{pmatrix} = A_0^{-1} b_0 \tag{7}$$

Coefficients from (7) can be negative, which cannot guarantee the estimation of covariance matrix to be positive definite. Therefore, to keep the estimated covariance matrix positive definite, the negative coefficient is clipped to zero respectively, say $\beta_t = 0$ and $\alpha_t = 0$.

2.2. AR(p) shrinkage model

For a more general case, AR(1) Shrinkage Model above is extended to AR(p) Shrinkage Model, that is covariance matrix from 2 to p lags are also taken into consideration.

$$\Sigma_t = \beta_{t1}\Sigma_{t-1} + \beta_{t2}\Sigma_{t-2} + \cdots + \beta_{tp}\Sigma_{t-p} + \alpha_t I + \epsilon_t = \sum_{i=1}^{p} \beta_{ti}\Sigma_{t-i} + a_t I + \epsilon_t \tag{8}$$

where all coefficients need to be positive to make sure the covariance matrix positive definite, i.e.

$$\alpha_t \geq 0 \text{ and } \beta_{ti} \geq 0 \; \forall i = 1, \cdots, p \tag{9}$$

The loss function is in squared Frobenius norm form:

$$\mathcal{L}(\boldsymbol{\beta}_t, \alpha_t) = \left\| \Sigma_t - \sum_{i=1}^{p} \beta_{ti} \Sigma_{t-i} - \alpha_t I \right\|_F^2 \tag{10}$$

Writing the loss function in quadratic function of coefficients:

$$\mathcal{L}(\boldsymbol{\beta_t}, \alpha_t) = (\boldsymbol{\beta_t'} \quad \alpha_t) A_0 \begin{pmatrix} \boldsymbol{\beta_t} \\ \alpha_t \end{pmatrix} - 2(\boldsymbol{\beta_t'} \quad \alpha_t) b_0 + tr(\Sigma^2) \tag{11}$$

where the A_0 and B_0 is:

$$A_0 = \begin{bmatrix} tr(\Sigma_{t-1}^2) & tr(\Sigma_{t-2}\Sigma_{t-1}) & \cdots & tr(\Sigma_{t-p}\Sigma_{t-1}) & tr(\Sigma_{t-1}) \\ tr(\Sigma_{t-1}\Sigma_{t-2}) & tr(\Sigma_{t-2}^2) & \cdots & tr(\Sigma_{t-p}\Sigma_{t-2}) & tr(\Sigma_{t-2}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ tr(\Sigma_{t-1}\Sigma_{t-p}) & tr(\Sigma_{t-2}\Sigma_{t-p}) & \cdots & tr(\Sigma_{t-p}^2) & tr(\Sigma_{t-p}) \\ tr(\Sigma_{t-1}) & tr(\Sigma_{t-2}) & \cdots & tr(\Sigma_{t-p}) & tr(I) \end{bmatrix} \tag{12}$$

$$b_0 = \begin{pmatrix} tr(\Sigma_{t-1}\Sigma_t) \\ \vdots \\ tr(\Sigma_{t-p}\Sigma_t) \\ tr(\Sigma_t) \end{pmatrix} \tag{13}$$

The coefficients can then be obtained by solving the problem of minimizing the loss function of (11), which is a strictly convex quadratic program with p+1 variables.

2.3. Multi-target shrinkage model

The single shrinkage target matrix introduced in (1) and (8) can only capture features of the diagonal elements, which are the sample variance of assets in the portfolio. However, considering that assets in the portfolio can be grouped because certain assets can have the similar patterns in the return series, features of the covariance among assets in different group should also be captured. Therefore, equation (1) should be extended to

$$\Sigma_t = \beta_t \Sigma_{t-1} + \alpha_{0t} I + \sum_{i=1}^{m} \alpha_{it} T_i + \epsilon_t \tag{14}$$

where T_i represents the shrinkage target matrix capturing the covariance among assets in the same group. Like (11), the loss function now can be written as

$$\mathcal{L}(\beta_t, \boldsymbol{\alpha_t}) = (\beta_t \quad \boldsymbol{\alpha_t'}) A_0 \begin{pmatrix} \beta_t \\ \boldsymbol{\alpha_t} \end{pmatrix} - 2(\beta_t \quad \boldsymbol{\alpha_t'}) b_0 + tr(\Sigma^2) \tag{15}$$

where the A_0 and B_0 is:

$$A_0 = \begin{bmatrix} tr(\Sigma_{t-1}^2) & tr(\Sigma_{t-1}T_1) & \cdots & tr(\Sigma_{t-1}T_m) & tr(\Sigma_{t-1}) \\ tr(\Sigma_{t-1}T_1) & tr(T_1^2) & \cdots & tr(T_1T_m) & tr(\mathrm{T}_1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ tr(\Sigma_{t-1}T_m) & tr(T_1T_m) & \cdots & tr(T_m^2) & tr(T_m) \\ tr(\Sigma_{t-1}) & tr(\mathrm{T}_1) & \cdots & tr(T_m) & tr(I) \end{bmatrix} \tag{16}$$

$$b_0 = \begin{pmatrix} tr(\Sigma_{t-1}\Sigma_t) \\ tr(T_1\Sigma_t) \\ \vdots \\ tr(T_m\Sigma_t) \\ tr(\Sigma_t) \end{pmatrix} \tag{17}$$

which can be solved in the form of (7).

2.4. Regularized shrinkage model

Based on AR(1) shrinkage model, l_2 norm of coefficients is added to the loss function as regularization. Now the regularized loss function becomes

$$\mathcal{L}^{Ridge}(\beta_t, \alpha_t) = \left|\left|\Sigma_t - \beta_t\Sigma_{t-1} - \alpha_t I\right|\right|_F^2 + \gamma(\beta_t^2 + \alpha_t^2) \tag{18}$$

where $\gamma > 0$ is the regularized parameter, representing how sensitive the function to parameters. and the regularized loss function (18) can be written in quadratic function of the coefficients:

$$\mathcal{L}^{Ridge}(\beta_t, \alpha_t) = (\beta_t \quad \alpha_t)A_0\begin{pmatrix}\beta_t \\ \alpha_t\end{pmatrix} - 2(\beta_t \quad \alpha_t)b_0 + tr(\Sigma^2) + \gamma(\beta_t \quad \alpha_t)\begin{pmatrix}\beta_t \\ \alpha_t\end{pmatrix} \tag{19}$$

where A_0 and b_0 equal to function (4) and (5). It can find the minimizer of $\mathcal{L}^{Ridge}(\beta_t, \alpha_t)$ by letting $\frac{\partial\mathcal{L}^{Ridge}}{\partial\beta_t} = \frac{\partial\mathcal{L}^{Ridge}}{\partial\alpha_t} = 0$ and solving the above bivariate convex optimization problem.

$$\begin{cases} \dfrac{tr(\Sigma_{t-1}^2) + \gamma}{tr(\Sigma_{t-1}\Sigma_t)}\beta_t + \dfrac{tr(\Sigma_{t-1})}{tr(\Sigma_{t-1}\Sigma_t)}\alpha_t = 1 \\ \dfrac{tr(\Sigma_{t-1})}{tr(\Sigma_t)}\beta_t + \dfrac{tr(I) + \gamma}{tr(\Sigma_t)}\alpha_t = 1 \end{cases} \tag{20}$$

From (20) the explicit form of coefficients can be derived as a function of γ

$$\begin{pmatrix}\beta_t^* \\ \alpha_t^*\end{pmatrix} = (A_0 + \gamma I)^{-1}b_0 \tag{21}$$

3. Empirical Study

3.1. Data-based estimation

To avoid the impact of noise in the return series when estimating the covariance matrix, methods introduced in Arlot are applied [6]. Denote return series from time 1 to T as $[r_1, r_2, \cdots, r_{T-1}, r_T]$ where r_t as the return vector of assets in time t. Let y_t be the return series dropping out r_t and the estimated covariance matrix $\hat{\Sigma}_T = \frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y}_t)(y_t - \bar{y}_t)^T$. Therefore, the cost function can be written as

$$\mathcal{L}(\boldsymbol{\beta_t}, \boldsymbol{\alpha_t}) = \frac{1}{T-p}\sum_{t=p+1}^{T}\left\|\sum_{i=1}^{p}\beta_{ti}\hat{\Sigma}_{t-i} + \alpha_{0t}I + \sum_{i=1}^{m}\alpha_{it}T_i - \Sigma_t\right\|_F^2 + p(\boldsymbol{\beta_t}, \boldsymbol{\alpha_t}) \tag{22}$$

$$\Rightarrow \mathcal{L}(\boldsymbol{\beta_t}, \boldsymbol{\alpha_t}) = (\boldsymbol{\beta_t'} \quad \boldsymbol{\alpha_t'})\hat{A}_0\begin{pmatrix}\boldsymbol{\beta_t}\\ \boldsymbol{\alpha_t}\end{pmatrix} - 2(\boldsymbol{\beta_t} \quad \boldsymbol{\alpha_t})\hat{b}_0 + \frac{1}{T-p}\sum_{t=p+1}^{T} tr(\Sigma_t^2) + p(\boldsymbol{\beta_t}, \boldsymbol{\alpha_t}) \tag{23}$$

Where

$$\hat{A}_0 = \begin{bmatrix}\hat{A}_0^{11} & \hat{A}_0^{12}\\ \hat{A}_0^{21} & \hat{A}_0^{22}\end{bmatrix} \tag{24}$$

$$\begin{cases}\hat{A}_0^{11} = \frac{1}{T-p}\begin{bmatrix}\sum_{t=p+1}^{T} tr(\hat{\Sigma}_{t-1}^2) & \cdots & \sum_{t=p+1}^{T} tr(\hat{\Sigma}_{t-p}\hat{\Sigma}_{t-1})\\ \vdots & \ddots & \vdots\\ \sum_{t=p+1}^{T} tr(\hat{\Sigma}_{t-1}\hat{\Sigma}_{t-p}) & \cdots & \sum_{t=p+1}^{T} tr(\hat{\Sigma}_{t-p}^2)\end{bmatrix}\\ \hat{A}_0^{12} = \hat{A}_0^{21} = \frac{1}{T-p}\begin{bmatrix}\sum_{t=p+1}^{T} tr(\hat{\Sigma}_{t-1}T_1) & \cdots & \sum_{t=p+1}^{T} tr(\hat{\Sigma}_{t-1}T_m) & \sum_{t=p+1}^{T} tr(\hat{\Sigma}_{t-1}\\ \vdots & \ddots & \vdots & \vdots\\ \sum_{t=p+1}^{T} tr(\hat{\Sigma}_{t-p}T_1) & \cdots & \sum_{t=p+1}^{T} tr(\hat{\Sigma}_{t-p}T_m) & \sum_{t=p+1}^{T} tr(\hat{\Sigma}_{t-p}\end{bmatrix}\\ \hat{A}_0^{22} = \begin{bmatrix}tr(T_1^2) & \cdots & tr(T_1T_m) & tr(T_1)\\ \vdots & \ddots & \vdots & \vdots\\ tr(T_mT_1) & \cdots & tr(T_m^2) & tr(T_m)\\ tr(T_1) & \cdots & tr(T_m) & tr(I)\end{bmatrix}\end{cases}$$

$$\hat{b}_0 = \frac{1}{T-p}\begin{pmatrix}\hat{b}_0^1\\ \hat{b}_0^2\end{pmatrix} \tag{25}$$

$$\hat{b}_0^1 = \begin{pmatrix} \sum_{t=p+1}^{T} tr\left(\hat{\Sigma}_{t-1}\hat{\Sigma}_t\right) \\ \vdots \\ \sum_{t=p+1}^{T} tr\left(\hat{\Sigma}_{t-p}\hat{\Sigma}_t\right) \end{pmatrix} \text{ and } \hat{b}_0^2 = \begin{pmatrix} \sum_{t=p+1}^{T} tr\left(T_1\hat{\Sigma}_t\right) \\ \vdots \\ \sum_{t=p+1}^{T} tr\left(T_m\hat{\Sigma}_t\right) \\ \sum_{t=p+1}^{T} tr\left(\hat{\Sigma}_t\right) \end{pmatrix}$$

And $p(\boldsymbol{\beta}_t, \boldsymbol{\alpha}_t)$ is regularization function of coefficients (in l_2 norm form as equation (22)). Combining (23), (24) and (25), the minimizer of loss function (23) can be directly solved in explicit form as (7)

$$\begin{pmatrix} \boldsymbol{\beta}_t^* \\ \boldsymbol{\alpha}_t^* \end{pmatrix} = \left(\hat{A}_0 + \gamma I\right)^{-1} \hat{b}_0 \tag{26}$$

which is the function of $T,\ p,\ \gamma$.

3.2. Global minimum-variance portfolio

Considering that it is more difficult to estimate expected return than covariance matrix, in this paper global minimum variance (GMV) portfolio will be focused on because of its simplicity and satisfactory result. The GMV aims at finding the portfolio giving the least risk level with the risk measurement as variance. There is a closed-form solution for GMV problem, which is nevertheless dependent on estimation of covariance matrix so that the error is diminished. To further illustrate the GMV portfolio choice, denote $w \in \mathbb{R}^p$ as the weight vector indicating the percentage of allocation for investment in each asset, and $\Sigma \in \mathbb{R}^{p \times p}$ as the covariance matrix of the net returns. The weights for the GMV portfolio can be obtained by solving the following optimization problem:

$$\begin{aligned} &\min_{w \in \mathrm{R}^{\mathrm{p}}} w'\Sigma w \\ &s.t.\, \mathbf{1}'w = 1 \end{aligned} \tag{27}$$

which can be solved using the Lagrange-multiplier method to obtain the weights:

$$w_o = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \tag{28}$$

The solution in (28) shows that the solution only depends on covariance matrix which cannot be observed in practice. Therefore, it will substitute Σ to the estimated value obtained in (19) and (26) to get the estimated portfolio weight vector.

$$w^* = \frac{\hat{\Sigma}_{\beta_t,\alpha_t}^{-1}\mathbf{1}}{\mathbf{1}'\hat{\Sigma}_{\beta_t,\alpha_t}^{-1}\mathbf{1}} \tag{29}$$

3.3. Data and evaluation

To evaluate the performance of the proposed regularized shrinkage model and autoregressive shrinkage model, in this paper, it will utilize data set similar with previous investigations. The dataset consists of daily net returns of 49 industries based on stocks listed on NYSE, AMEX, and NASDAQ,

covering periods from July 1, 1926, to July 31, 2022. Weekly return of each industry portfolio is calculated based on the day-frequency dataset and 104 weeks (2 year) is used as the estimation period for covariance matrix, that is the covariance matrix is estimated with the past 104-week net return of different industry portfolio. For the back-test period, this paper carries out the back tests from January 2000 to June 2022.

For calibrating the coefficients, rolling-window method is applied, that is, for a given calibrating period T (T = 52, 78, 104), at time t, coefficients (β_t, α_t) are calibrated according to (26) on the data from t-T to t-1. Then with estimated covariance matrix, it can derive the portfolio weights based on (29). The weight is calculated and rebalanced at the end of every week. For evaluation of the performance, several indicators are utilized, such as Sharpe Ratio, Predicted Risk and Realized Risk as the indicators, which are introduced in detail in table 1 below.

Table 1. Details of performance evaluation indicators

Indicator	Formula	Comments
Mean	$\mu = \frac{1}{N}\sum_{t=1}^{N} r_t$	
Standard Deviation	$\sigma(r) = \frac{1}{N-1}\sum_{t-1}^{N}(r_t - \mu)^2$	
Sharpe Ratio	$SR = \frac{\mu - R_0}{\sigma(r)}$	R_0 is average risk-free return which is the return of 3-month US Treasury bills
Predicted Risk	$\widehat{Var}_t = \frac{1}{N}\sum_{t=1}^{N} w_t'\hat{\Sigma}_t w_t$	The average of predicted portfolio variance with estimated covariance matrix
Realized Risk	$Var_t = \frac{1}{N}\sum_{t=1}^{N} w_t'\Sigma_t w_t$	The average of real portfolio variance with realized covariance matrix

3.4. Empirical analysis results for AR(1) shrinkage model

Given that the estimating coefficients are functions of the regularization parameter as (26), performances of the methods under different regularization parameters are also tested. The accumulated return and performance statistics are analyzed with changing γ in function (26) from 0 to 0.9 (0 is the case without regularization). Table 2 and figure 1, 2, 3 are the summary of statistics on the performance of different methods.



Figure 1. Accumulated return of AR(1) shrinkage model (T=52)

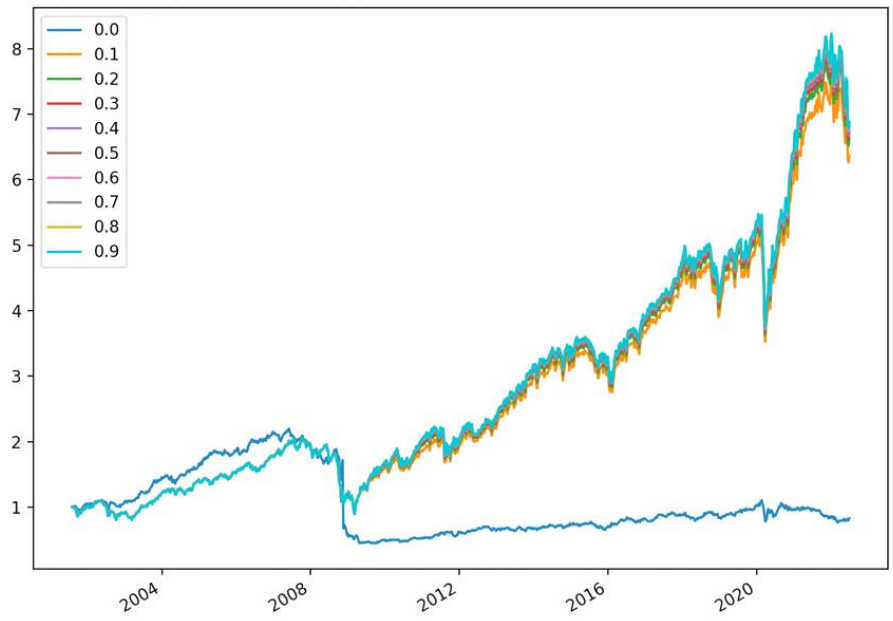


Figure 2. Accumulated return of AR(1) shrinkage model (T=78)

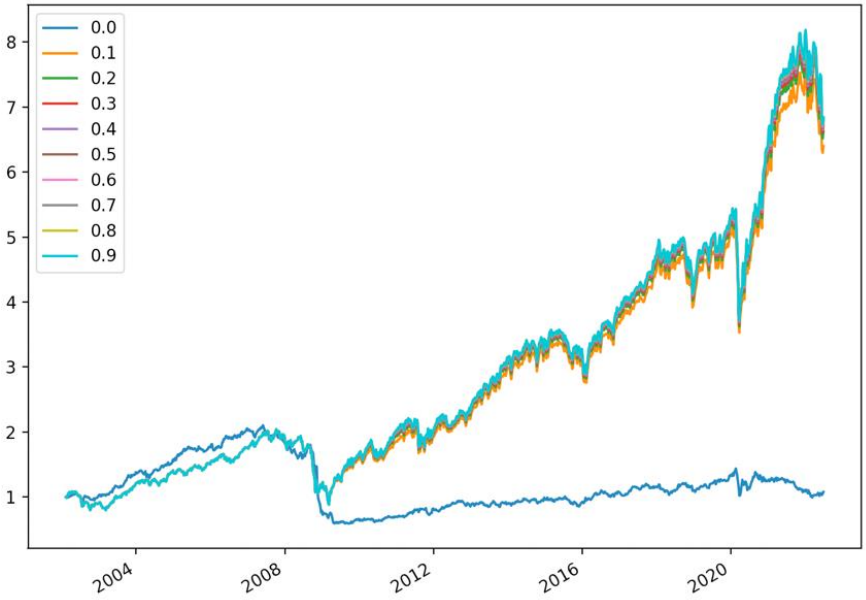


Figure 3. Accumulated return of AR(1) shrinkage model (T=104)

From figure 1, 2, 3 it can observe that different portfolio allocation with regularization can have similar performances, which significantly outperform the allocation strategies without regularization. What's more, it can find that estimation model without regularization can have lower volatility than regularized model, especially in the period after the Global Financial Crisis. What's more, the length of calibrating window (T) also has significant impact on the performance. If the window is too long as shown in figure 2 and 3 (T=78 and 104), the estimation will be affected more by the long-past data and less by the latest data, which will make the estimation less robust.

Table 2. Statistics on performance of AR(1) shrinkage models with different parameters

γ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					T=52					
Mean (%)	3.70	10.56	10.85	10.96	11.03	11.07	11.10	11.12	11.13	11.15
Std (%)	16.09	18.20	18.65	18.83	18.93	18.99	19.03	19.07	19.09	19.11
Sharpe Ratio	0.149	0.509	0.512	0.514	0.514	0.515	0.515	0.515	0.515	0.515
Pred. Risk (%)	0.464	3.183	3.395	3.485	3.536	3.568	3.591	3.607	3.620	3.630
Real. Risk (%)	0.495	3.007	3.277	3.397	3.465	3.509	3.540	3.563	3.580	3.594
					T=78					
Mean (%)	2.02	10.95	11.23	11.34	11.40	11.44	11.46	11.48	11.50	11.51
Std (%)	20.41	18.22	18.67	18.85	18.95	19.02	19.06	19.09	19.11	19.13
Sharpe Ratio	0.039	0.534	0.536	0.536	0.537	0.537	0.537	0.537	0.537	0.537
Pred. Risk (%)	0.528	3.195	3.405	3.494	3.543	3.574	3.596	3.612	3.624	3.634
Real. Risk (%)	0.719	3.021	3.290	3.408	3.475	3.574	3.547	3.569	3.586	3.600
					T=104					
Mean (%)	1.68	11.14	11.40	11.50	11.56	11.59	11.61	11.63	11.64	11.65
Std (%)	16.19	18.21	18.66	18.84	18.94	19.00	19.04	19.07	19.10	19.11
Sharpe Ratio	0.030	0.546	0.547	0.547	0.547	0.547	0.547	0.547	0.547	0.547
Pred. Risk (%)	0.468	3.206	3.414	3.501	3.549	3.580	3.601	3.616	3.628	3.637
Real. Risk (%)	0.486	3.033	3.302	3.418	3.483	3.525	3.554	3.575	3.592	3.605

Table 2 shows the performance of the autoregressive shrinkage model with different parameters (with $\gamma = 0$ indicating the non-regularized model while others regularized models). Comparing the length of calibrating period, it can observe that the longer the calibrating period is, the higher Sharpe Ratio the regularized autoregressive shrinkage model will reach. Therefore, it can conclude that the regularization will enhance performance with longer calibrating period. As for the regularization parameter, the higher γ is, the higher expected return and volatility will be. Such reflects the trade-off between the minimum volatility and the robustness of the estimation.

Comparing the difference between the regularized model and non-regularized model, though regularized model has higher risk level, it can reach significantly higher Sharpe Ratio. Also, compared with non-regularized model whose predicted risk is lower than the realized risk, the regularized model has lower realized risk than predicted risk, indicating that the estimation of covariance matrix with regularization is robust and not under-estimating the risk.

3.5. Empirical analysis results for multi-target AR(1) shrinkage model

To construct the target shrinkage matrix, this paper applied k-means clustering methods first to group the industries. The distance is measured in Euclidean metric. For each group, the elements corresponding to the industries in the group are set to 1 while others 0, and each group corresponds to one shrinkage matrix. The following is the algorithm for target shrinkage matrix construction and covariance matrix estimation.

Initialize
Choose: cluster number N
Choose: $T,\ p,\ \gamma$ in (26)
At time t
Solve clusters on $D_{\{t-T,t\}}^{train}$ with k-means algorithm
Set $\mathcal{H} = \{H_i : i = \{1, \cdots, N\}\}$ where H_i is the set of industries
For $i \leftarrow 1\ to\ N$ **do**:
Initialize $T_i = 0 \in \mathbb{R}^{\#H_i \times \#H_i}$ where $\#H_i$ is the number of elements in H_i
For all $ind_{ik}, ind_{ij} \in H_i,\ T_i[ind_{ik}, ind_{ij}] = 1$
End
Solve $\hat{A}_0$ and $\hat{b}_0$ based on (24) and (25)
Solve coefficients of (23): $(\hat{A}_0 + \gamma I)^{-1}\hat{b}_0$
Return $\hat{\Sigma}_t$

Algorithm 1. Covariance estimation with k-means autoregressive shrinkage model

Similar to analysis in section 3.4, accumulated return and performance statistics of k-means autoregressive model is also analysed. In the back test, the cluster number is set to be 5 for the whole back-test period. Table 3 and figure 4, 5, 6 are the summary of statistics on the performance of different methods.

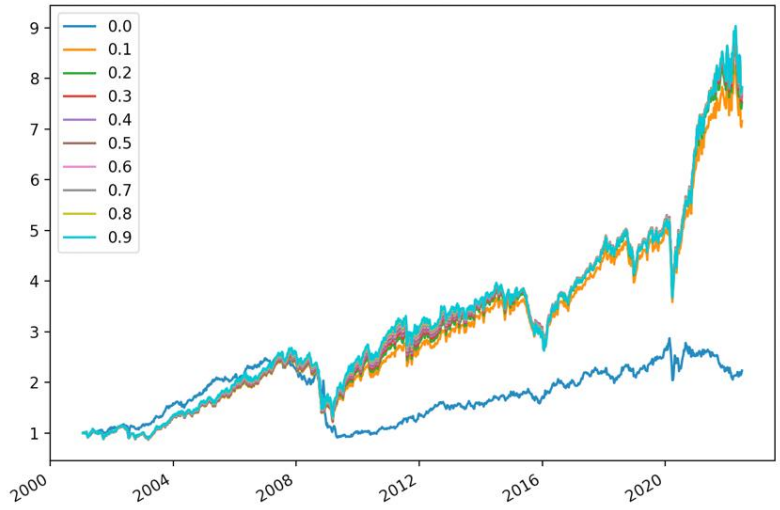


Figure 4. Accumulated return of k-means AR(1) shrinkage model (T=52)

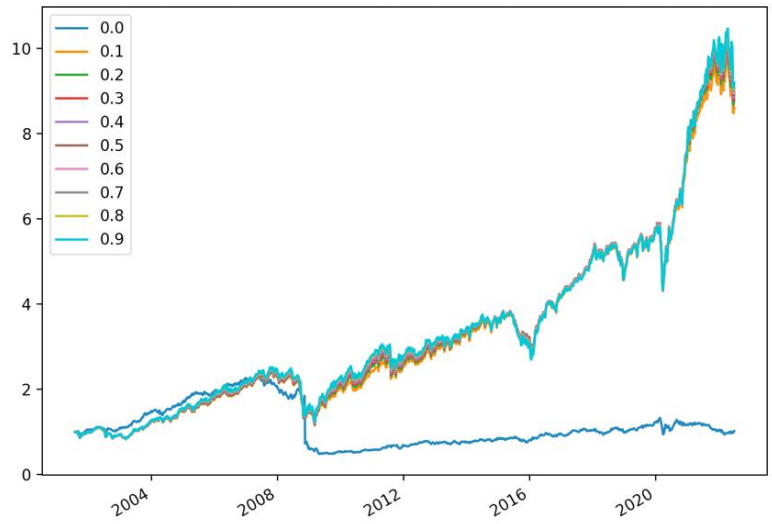


Figure 5. Accumulated return of k-means AR(1) shrinkage model (T=78)

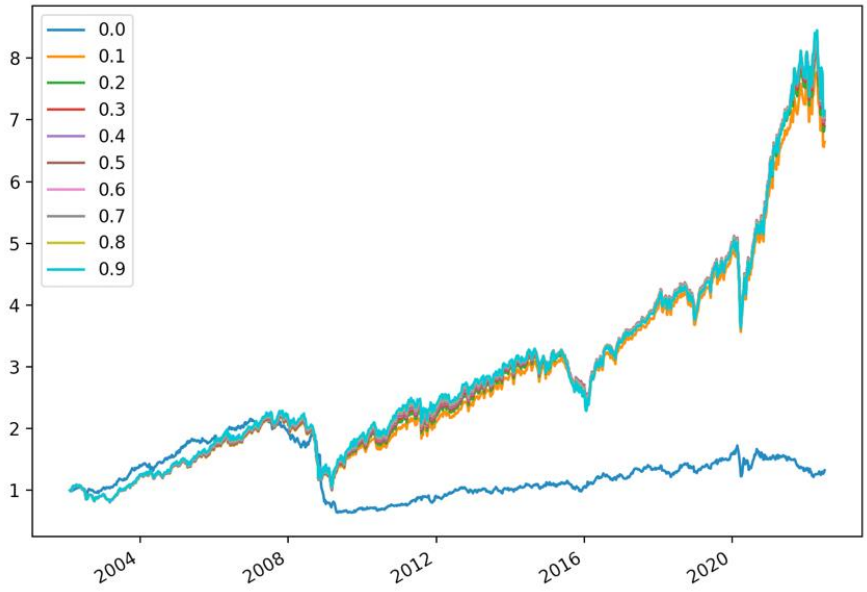


Figure 6. Accumulated return of k-means AR(1) shrinkage model (T=104)

Figure 4, 5 and 6 demonstrates the similar patterns as section 3.4 that (1) regularization can enhance the performance of the model; (2) model without regularization can have lower volatility. The more detailed statistics are shown in table 3 below.

Table 3. Statistics on performance of k-means AR(1) shrinkage models with different parameters

γ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					T=52					
Mean (%)	5.26	11.17	11.48	11.60	11.67	11.71	11.74	11.77	11.79	11.80
Std (%)	16.09	17.90	18.29	18.48	18.62	18.72	18.81	18.89	18.96	19.03
Sharpe Ratio	0.246	0.551	0.557	0.557	0.557	0.556	0.555	0.554	0.553	0.552
Pred. Risk (%)	0.464	2.964	3.136	3.211	3.255	3.285	3.308	3.326	3.341	3.354
Real. Risk (%)	0.494	2.823	3.051	3.161	3.231	3.284	3.326	3.362	3.394	3.424
					T=78					
Mean (%)	3.06	12.36	12.54	12.63	12.69	12.74	12.78	12.82	12.85	12.88
Std (%)	20.39	17.76	18.16	18.36	18.49	18.59	18.68	18.76	18.83	18.90
Sharpe Ratio	0.090	0.627	0.623	0.621	0.620	0.619	0.618	0.618	0.617	0.617
Pred. Risk (%)	0.528	2.984	3.150	3.223	3.266	3.296	3.318	3.337	3.352	3.366
Real. Risk (%)	0.719	2.838	3.057	3.162	3.230	3.281	3.323	3.358	3.390	3.419
					T=104					
Mean (%)	2.72	11.22	11.47	11.58	11.64	11.69	11.72	11.75	11.77	11.79
Std (%)	16.21	17.62	17.98	18.16	18.29	18.39	18.47	18.55	18.63	18.69
Sharpe Ratio	0.094	0.569	0.571	0.572	0.571	0.570	0.570	0.569	0.568	0.566
Pred. Risk (%)	0.459	2.941	3.093	3.160	3.200	3.227	3.248	3.265	3.279	3.292
Real. Risk (%)	0.474	2.782	2.982	3.078	3.142	3.189	3.228	3.262	3.292	3.320

Compared with the back-test result of AR(1) shrinkage model in table 2, k-means AR(1) shrinkage model demonstrates significantly lower predicted risk and realized risk. This shows that k-means clustering method helps to capture the correlation relationship within the groups, which enhances the performance of diversification. What's more, the higher Sharpe Ratio also indicates the better performance with k-means clustering methods. However, compared with simple AR(1) shrinkage model, k-means AR(1) shrinkage model is more sensitive to the choice of T. Such is reasonable because both too long and too short training period for clustering will not capture the right correlation relationship among industries, which will lower the performance.

4. Conclusions

This paper has proposed both simple and k-means autoregressive-based shrinkage methods to estimate the covariance matrix. Several contributions have been made in this paper. Firstly, the explicit form of the covariance matrix estimation model is provided, and it manages to show that the regularized model can outperform the non-regularized model. Secondly, through empirical tests it shows that in a reasonable interval of the regularization parameter, the performance of model is non-sensitive to the value of parameter. Thirdly, a new model including k-means algorithm into estimation has been developed which managed to capture more features than the simple model and enhance the back-test performance. It is hoped that developing a k-means regularized autoregressive shrinkage method for covariance matrix estimation will spur new methods for covariance estimation with better out-sample performance.

References

[1] Markowitz, H. Portfolio selection. Journal of Finance, 1952, 7(1): 77-91.

[2] C.Merton, R. On estimating the expected return on the market: An exploratory investigation. Journal of Financial Economics, 1980, 8(4): 323-361.

[3] Jagannathan, R., Ma, T. Risk reduction in large portfolios: Why imposing the wrong constraints helps. Journal of Finance, 2003, 58(4): 1651-1683.

[4] DeMiguel, V., Garlappi, L., Nogales, F. J., Uppal, R. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. Management Science, 2009, 55(5): 798-812.

[5] Awoye, O. Markowitz minimum variance portfolio optimization using new machine learning methods. (Doctor of Philosophy in Computer Science), University College London, 2016.

[6] Arlot, S., Celisse, A. A survey of cross-validation procedures for model selection. Statistics Surveys, 2010, 4: 40 -79.

[7] Ledoit, O., Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis, 2004, 88(2): 365-411.

[8] Ollila, E., Raninen, E. Optimal shrinkage covariance matrix estimation under random sampling from elliptical distributions. IEEE Transactions on Signal Processing, 2019, 67(10): 2707-2719.

[9] Mahadi, M., Ballal, T., Moinuddin, M., Al-Saggaf, U. M. Portfolio optimization using a consistent vector-based mse estimation approach. IEEE Access, 2022.

[10] Delage, E., Ye, Y. Distributionally robust optimization under moment uncertainty with application to data-driven problems. Operations Research, 2010, 58(3): 595-612.

[11] Ban, G.-Y., Karoui, N. E., Lim, A. E. B. Machine learning and portfolio optimization. Management Science, 2018, 64(3): 1136–1154.

[12] Tong, J., Hu, R., Xi, J., Xiao, Z., Guo, Q., Yu, Y. Linear shrinkage estimation of covariance matrices using low-complexity cross-validation. Signal Processing, 2018, 148: 223-233.

[13] Li, S., Zhang, Q. Empirical research and regulatory thinking on the industry impact of the new crown epidemic on U.S. stocks--Based on the Fama-French three-factor and five-factor model. Wuling Academic Journal, 2021, (05): 48-57.