PORTFOLIO SELECTION USING TIKHONOV FILTERING TO ESTIMATE THE COVARIANCE MATRIX

SUNGWOO PARK*

Abstract. Markowitz's portfolio selection problem chooses weights for stocks in a portfolio based on a covariance matrix of stock returns. Our study proposes to reduce noise in the estimated covariance matrix using a Tikhonov filter function. In addition, we propose a new strategy to resolve the rank deficiency of the covariance matrix, and a method to choose a Tikhonov parameter which determines a filtering intensity. We put the previous estimators into a common framework and compare their filtering functions for eigenvalues of the correlation matrix. Experiments using the daily return data of the most frequently traded stocks in NYSE, AMEX, and NASDAQ show that Tikhonov filtering estimates the covariance matrix better than methods of Sharpe who applies a market-index model, Ledoit et al. who shrink the sample covariance matrix to the market-index covariance matrix, Elton and Gruber, who suggest truncating the smallest eigenvalues, Bengtsson and Holst, who decrease small eigenvalues at a single rate, and Plerou et al. and Laloux et al., who use a random matrix approach.

Key words. Tikhonov regularization, covariance matrix estimate, Markowitz portfolio selection, ridge regression

1. Introduction.

A stock investor might want to construct a portfolio of stocks whose return has a small variance, because large variance implies high risk. Given a target portfolio return q, Markowitz's portfolio selection problem [22] finds a stock weight vector $\boldsymbol{w}$ to determine a portfolio that minimizes the expected variance of the return. Let $\boldsymbol{\mu}$ be a vector of expected returns for each of N stocks, and let $\boldsymbol{\Sigma}$ be an $N \times N$ covariance matrix for the returns. The portfolio selection problem can be written as

$$\min_{\boldsymbol{w}} \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w} \text{ subject to } \quad \boldsymbol{w}^T \mathbf{1} = 1, \quad \boldsymbol{w}^T \boldsymbol{\mu} = q, \tag{1.1}$$

where $\mathbf{1}$ is a vector of N ones. In order to construct a good portfolio using this formulation, the covariance matrix $\boldsymbol{\Sigma}$ must be well-estimated. Let $\boldsymbol{R} = [\boldsymbol{r}(1), \cdots, \boldsymbol{r}(T)]$ be an $N \times T$ matrix containing observations on N stocks' returns for each of T times. A conventional estimator – a sample covariance matrix $\boldsymbol{\Sigma}_{sample}$ – can be computed from the stock return matrix $\boldsymbol{R}$ as

$$\boldsymbol{\Sigma}_{sample} = \frac{1}{T-1} \boldsymbol{R}(\boldsymbol{I} - \frac{1}{T}\mathbf{1}\mathbf{1}^T)\boldsymbol{R}^T. \tag{1.2}$$

However, since the stock return matrix $\boldsymbol{R}$ contains noise, the sample covariance matrix $\boldsymbol{\Sigma}_{sample}$ might not estimate the true covariance matrix well. This paper uses principal component analysis and reduces the noise in the covariance matrix estimate by using a Tikhonov regularization method. We demonstrate experimentally that this improves the portfolio weight $\boldsymbol{w}$ obtained from (1.1).

Our study is closely related to factor analysis and principal component analysis, which were previously applied to explain interdependency of stock returns and classify the securities into appropriate subgroups. Sharpe [32] first proposed a single-factor model in this context using market returns. King [17] analyzed stock behaviors with

*Computer Science Dept., Univ. of Maryland, College Park, MD 20742; swpark81@gmail.com
This study was advised by Professor Dianne P. O'Leary, and was supported in part by NSF CCF 05 14213.

both multiple factors and multiple principal components. These factor models established a basis for the asset pricing models CAPM [21, 24, 33, 36] and APT [28, 29].

There have been previous efforts to improve the estimate of $\mathbf{\Sigma}$. Sharpe [32] proposed a market-index covariance matrix $\mathbf{\Sigma}_{market}$ derived from a single-factor model of market returns. Ledoit et al. [20] introduced a shrinkage method that averages $\mathbf{\Sigma}_{sample}$ and $\mathbf{\Sigma}_{market}$. Elton and Gruber [6] used a few principal components from a correlation matrix. More recently, Plerou et al. [27], Laloux et al. [19], Conlon et al. [4], and Kwapień [18] applied random matrix theory [23] to this problem. They found that most eigenvalues of correlation matrices from stock return data lie within the bound for a random correlation matrix and hypothesized that eigencomponents (principal components) outside this interval contain true information. Bengtsson and Holst [2] generalized the approach of Ledoit et al. by damping all but the k largest eigenvalues by a single rate. In summary, the estimator of Sharpe [32] uses $\mathbf{\Sigma}_{market}$, the estimator of Ledoit et al. [20] takes the weighted average of $\mathbf{\Sigma}_{market}$ and $\mathbf{\Sigma}_{sample}$, the estimator of Elton and Gruber [6] truncates the smallest eigenvalues, the estimators of Plerou et al. [27], Laloux et al. [19], Conlon et al. [4], and Kwapień [18] adjust principal components in some interval, and the estimator of Bengtsson and Holst [2] attenuates the smallest eigenvalues by a single rate.

We propose to decrease the contribution of the smaller eigenvalues gradually by using a *Tikhonov filtering function*. In addition, the covariance matrix is usually rank deficient due to the lack of the trading price data, which causes non-unique solutions for the Markowitz portfolio selection problem. Most other estimators resolve this issue by replacing the diagonal elements of the estimated covariance matrix with the diagonal elements of the sample covariance matrix. We explain why this replacement is not desirable from the viewpoint of noise filtering and propose a new strategy to fix the rank deficiency.

This paper is organized as follows. In Section 2, we introduce Tikhonov regularization to reduce noise in the stock return data. In Section 3, we show that applying Tikhonov regularization results in filtering the eigenvalues of the correlation matrix for the stock returns. We also propose a new method to resolve the rank deficiency of the covariance matrix, and verify its reasonableness. In Section 4, we discuss how we can choose a Tikhonov parameter that determines the intensity of Tikhonov filtering. In Section 5, we put all of the estimators into a common framework, and compare the characteristics of their filtering functions for the eigenvalues of the correlation matrix. In Section 6, we show the results of numerical experiments of Markowitz portfolio construction comparing different covariance estimators, using the daily return data of the most frequently traded 112 stocks in NYSE, AMEX, and NASDAQ. In Section 7, based on the result of the experiments, we conclude that gradual attenuation by Tikhonov filtering outperforms shrinkage, truncation, single-rate attenuation, and the random matrix approach.

2. Tikhonov filtering. We apply a principal component analysis to find an orthogonal basis that maximizes the variance of the projected data into the basis. Based on the analysis, we use the Tikhonov regularization method to filter out the noise from the data. Next, we explain the feature of gradual down-weighting, which is the key difference between Tikhonov filtering and other methods.

2.1. Principal component analysis. First, we establish some notation. We use a 2-norm $||\cdot||$ for vectors, and a Frobenius norm $||\cdot||_F$ for matrices, defined as

$$||\boldsymbol{a}||^2 = \boldsymbol{a}^T\boldsymbol{a} \text{ for a given vector } \boldsymbol{a},$$
$$||\boldsymbol{A}||_F^2 = \left(\sum_{i=1}^{M}\sum_{j=1}^{N} a_{ij}^2\right) \text{ for a given } M \times N \text{ matrix } \boldsymbol{A},$$

where a_{ij} is the (i,j) element of $\boldsymbol{A}$.

Define a collection of observations $\boldsymbol{X} = [\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T)]$ for N objects during T times where $\boldsymbol{x}(t) = [x_1(t), \ldots, x_N(t)]^T$. Let $\mathbb{E}(x_i(t))$ and $\mathrm{Var}(x_i(t))$ denote the expected value and the variance of $x_i(t)$ for a given time T, defined so that

$$\mathbb{E}(x_i(t)) = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_i(t),$$
$$\mathrm{Var}(x_i(t)) = \frac{1}{T-1}\sum_{t=1}^{T} (\, x_i(t) - \mathbb{E}(x_i(t)))^2$$
$$= \frac{1}{T-1}\left(\sum_{t=1}^{T} \, x_i^2(t)\right) - \mathbb{E}(x_i(t))^2.$$

For two variables $x_i(t)$ and $x_j(t)$, let $\mathrm{Cov}(x_i(t), x_j(t))$ and $\mathrm{Corr}(x_i(t), x_j(t))$ be the covariance and correlation coefficient for them, defined so that

$$\mathrm{Cov}\,(x_i(t), x_j(t)) = \frac{1}{T-1}\left(\sum_{t=1}^{T} x_i(t)x_j(t)\right) - \mathbb{E}(x_i(t))\; \mathbb{E}(x_j(t)),$$
$$\mathrm{Corr}\,(x_i(t), x_j(t)) = \mathrm{Var}(x_i(t))^{-\frac{1}{2}}\; \mathrm{Cov}(x_i(t), x_j(t))\; \mathrm{Var}(x_j(t))^{-\frac{1}{2}}.$$

In order to distinguish multivariate statistics from single-variate statistics, we use square brackets like $\mathbb{E}[\boldsymbol{x}(t)]$, $\mathrm{Var}[\boldsymbol{x}(t)]$, $\mathrm{Cov}[\boldsymbol{x}(t)]$, and $\mathrm{Corr}[\boldsymbol{x}(t)]$. We define the expected value $\mathbb{E}[\boldsymbol{x}(t)]$ and the variance $\mathrm{Var}[\boldsymbol{x}(t)]$ by

$$\mathbb{E}[\boldsymbol{x}(t)] = \begin{bmatrix} \mathbb{E}(x_1(t)) \\ \vdots \\ \mathbb{E}(x_N(t)) \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}(t),$$

$$\mathrm{Var}[\boldsymbol{x}(t)] = \begin{bmatrix} \mathrm{Var}(x_1(t)) \\ \vdots \\ \mathrm{Var}(x_N(t)) \end{bmatrix} = \frac{1}{T-1}\sum_{t=1}^{T} (\boldsymbol{x}(t) - \mathbb{E}[\boldsymbol{x}(t)])^2.$$

$\mathrm{Cov}[\boldsymbol{x}(t)]$ will be an $N \times N$ matrix whose (i,j) element is $\mathrm{Cov}(x_i(t), x_j(t))$. $\mathrm{Cov}[\boldsymbol{x}(t)]$ can be computed from the $N \times T$ observation matrix $\boldsymbol{X}$ as

$$\mathrm{Cov}[\boldsymbol{x}(t)] = \frac{1}{T-1}\boldsymbol{X}(\boldsymbol{I} - \frac{1}{T}\boldsymbol{1}\boldsymbol{1}^T)\boldsymbol{X}^T.$$

$\mathrm{Corr}[\boldsymbol{x}(t)]$ will be an $N \times N$ matrix whose (i,j) element is $\mathrm{Corr}(x_i(t), x_j(t))$. $\mathrm{Corr}[\boldsymbol{x}(t)]$ can be computed from $\mathrm{Cov}[\boldsymbol{x}(t)]$ and $\mathrm{Var}[\boldsymbol{x}(t)]$ as

$$\mathrm{Corr}[\boldsymbol{x}(t)] = \mathrm{diag}(\mathrm{Var}[\boldsymbol{x}(t)])^{-\frac{1}{2}}\; \mathrm{Cov}[\boldsymbol{x}(t)]\;\; \mathrm{diag}(\mathrm{Var}[\boldsymbol{x}(t)])^{-\frac{1}{2}}.$$

Now we apply principal component analysis (PCA) to the stock return data $\boldsymbol{R}$. Let $\boldsymbol{Z} = [\boldsymbol{z}(1), \ldots, \boldsymbol{z}(T)]$ be an $N \times T$ matrix of a normalized stock returns derived from $\boldsymbol{R}$, defined so that

$$\mathbb{E}[\boldsymbol{z}(t)] = \boldsymbol{0}, \quad \mathrm{Var}[\boldsymbol{z}(t)] = \boldsymbol{1}, \tag{2.1}$$

where $\boldsymbol{0}$ is a vector of N zeros. We can compute $\boldsymbol{Z}$ as

$$\boldsymbol{Z} = \boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}^{-\frac{1}{2}}(\boldsymbol{R} - \frac{1}{T}\boldsymbol{R}\,\boldsymbol{1}\boldsymbol{1}^T), \tag{2.2}$$

where $\boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]} = \mathrm{diag}(\mathrm{Var}[\boldsymbol{r}(t)])$ is an $N \times N$ diagonal matrix containing the N variances for the N stock returns. The covariance matrix for the normalized stock return $\boldsymbol{z}(t)$ is equal to the correlation matrix for $\boldsymbol{r}(t)$. Thus, $\mathrm{Cov}[\boldsymbol{r}(t)]$ can be calculated from $\mathrm{Cov}[\boldsymbol{z}(t)]$ as

$$\begin{aligned} \mathrm{Cov}[\boldsymbol{r}(t)] &= \boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}^{\frac{1}{2}} \mathrm{Corr}[\boldsymbol{r}(t)] \boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}^{\frac{1}{2}} \\ &= \boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}^{\frac{1}{2}} \mathrm{Cov}[\boldsymbol{z}(t)] \boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}^{\frac{1}{2}} . \end{aligned} \tag{2.3}$$

Therefore, we can transfer the problem of improving our estimate of $\mathrm{Cov}[\boldsymbol{r}(t)]$ into the problem of improving our estimate of $\mathrm{Cov}[\boldsymbol{z}(t)]$. By using the normalized stock return matrix $\boldsymbol{Z}$ rather than $\boldsymbol{R}$, we can make the PCA independent of the different variance of each stock return [16, pp.64-66].

PCA finds an orthogonal basis $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k] \in \mathbb{R}^{N \times k}$ for $\boldsymbol{Z}$ where $k = \mathrm{rank}(\boldsymbol{Z})$. Each basis vector $\boldsymbol{u}_i$ maximizes the variance of the projected data $\boldsymbol{u}_i^T \boldsymbol{Z}$, while maintaining orthogonality to all the preceding basis vectors $\boldsymbol{u}_j$ $(j < i)$. By PCA, we can represent the given data $\boldsymbol{Z} = [\boldsymbol{z}(1), \ldots, \boldsymbol{z}(T)]$ as

$$\boldsymbol{Z} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k]\ \boldsymbol{X} = \boldsymbol{U}\boldsymbol{X}, \tag{2.4}$$

$$\boldsymbol{z}(t) = \boldsymbol{U}\ \boldsymbol{x}(t) = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k]\ \boldsymbol{x}(t) = \sum_{i=1}^{k} x_i(t)\boldsymbol{u}_i, \tag{2.5}$$

where $\boldsymbol{x}(t) = [x_1(t), \ldots, x_k(t)]^T$ is the projected data at time t, and $\mathrm{Var}(x_1(t)) \geq \mathrm{Var}(x_2(t)) \geq \cdots \geq \mathrm{Var}(x_k(t))$. From now on, we call the projected data $x_i(t)$ the i-th principal component at time t. Larger $\mathrm{Var}(x_i(t))$ implies that the corresponding $\boldsymbol{u}_i$ plays a more important role in representing $\boldsymbol{Z}$. The orthogonal basis $\boldsymbol{U}$ and the projected data $\boldsymbol{X}$ can be obtained by the singular value decomposition (SVD) of $\boldsymbol{Z}$,

$$\boldsymbol{Z} = \boldsymbol{U}_k\ \boldsymbol{S}_k\ \boldsymbol{V}_k^T, \tag{2.6}$$

where k is the rank of Z,

$\boldsymbol{U}_k = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k] \in \mathbb{R}^{N \times k}$ is a matrix of left singular vectors,
$\boldsymbol{S}_k = \mathrm{diag}(s_1, \ldots, s_k) \in \mathbb{R}^{k \times k}$ is a diagonal matrix of singular values s_i,
and $\boldsymbol{V}_k = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k] \in \mathbb{R}^{T \times k}$ is a matrix of right singular vectors.

The left and right singular vector matrices $\boldsymbol{U}_k$ and $\boldsymbol{V}_k$ are orthonormal :

$$\boldsymbol{U}_k^T\,\boldsymbol{U}_k = \boldsymbol{I}_k\ , \quad \text{and} \quad \boldsymbol{V}_k^T\,\boldsymbol{V}_k = \boldsymbol{I}_k, \tag{2.7}$$

where $\boldsymbol{I}_k$ is a $k \times k$ identity matrix.

In PCA, the orthogonal basis matrix $\boldsymbol{U}$ corresponds to $\boldsymbol{U}_k$, and the projected data $\boldsymbol{X}$ corresponds to $(\boldsymbol{S}_k \boldsymbol{V}_k^T)$ [16, p.193]. Moreover, the standard deviation of the projected data $x_i(t)$ is proportional to the square of singular value s_i^2 as we now show. Since $\boldsymbol{z}(t) = \boldsymbol{U}\,\boldsymbol{x}(t)$ and the normalized data $\boldsymbol{z}(t)$ has zero-mean,

$$\boldsymbol{0} = \mathbb{E}[\boldsymbol{z}(t)] = \mathbb{E}[\boldsymbol{U}\boldsymbol{x}(t)] = \boldsymbol{U}\,\mathbb{E}[\boldsymbol{x}(t)].$$

Because $\boldsymbol{U}$ is a full rank matrix,

$$\mathbb{E}[\boldsymbol{x}(t)] = \boldsymbol{0}. \tag{2.8}$$

By the definition of $\mathrm{Var}(x_i(t))$,

$$\mathrm{Var}(x_i(t)) = \frac{1}{T-1}\sum_{t=1}^{T}\left(x_i(t) - \mathbb{E}(x_i(t))\right)^2 = \frac{1}{T-1}\sum_{t=1}^{T} x_i^2(t).$$

Since $\boldsymbol{X}$ is equal to $\boldsymbol{S}_k \boldsymbol{V}_k^T$,

$$x_i(t) = s_i v_i(t) \tag{2.9}$$

where $v_i(t)$ is the (t, i) element of $\boldsymbol{V}_k$. Thus,

$$\mathrm{Var}(x_i(t)) = \frac{1}{T-1}\sum_{t=1}^{T}(s_i v_i(t))^2 = \frac{1}{T-1} s_i^2 (\boldsymbol{v}_i^T \boldsymbol{v}_i) = \frac{s_i^2}{T-1}, \tag{2.10}$$

by the orthonormality of $\boldsymbol{v}_i$. Thus, the singular value s_i determines the magnitude of $\mathrm{Var}(x_i(t))$, so it measures the contribution of the projected data $x_i(t)$ to $\boldsymbol{z}(t)$.

2.2. Tikhonov regularization. $\boldsymbol{U}$ and $\boldsymbol{x}(t)$ in (2.5) form a linear model with a k–dimensional orthogonal basis for the normalized stock return $\boldsymbol{Z}$, where $k = \mathrm{rank}(\boldsymbol{Z})$. As mentioned in the previous section, the singular value s_i determines how much the principal component $x_i(t)$ contributes to $\boldsymbol{z}(t)$. However, since noise is included in $\boldsymbol{z}(t)$, the k–dimensional model is overfitted, containing unimportant principal components possibly corresponding to the noise. We use a Tikhonov regularization method [25, 35, 38], sometimes called ridge regression [14, 15], to reduce the contribution of unimportant principal components to the normalized stock return $\boldsymbol{Z}$. Eventually, we construct a filtered principal component $\widetilde{\boldsymbol{x}}(t)$ and a filtered market return $\widetilde{\boldsymbol{Z}}$.

Originally, regularization methods were developed to reduce the influence of noise when solving a discrete ill-posed problem $\boldsymbol{b} \approx \boldsymbol{A}\boldsymbol{x}$, where the $M \times N$ matrix $\boldsymbol{A}$ has some singular values close to 0 [11, pp.71-86]. If we write the SVD of A as

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\,\boldsymbol{V}^T = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N]\begin{bmatrix} s_1 & & \\ & \ddots & \\ & & s_N \end{bmatrix}\begin{bmatrix} \boldsymbol{v}_1^T \\ \vdots \\ \boldsymbol{v}_N^T \end{bmatrix},$$

then the minimum norm least square solution $\boldsymbol{x}_{LS}$ to $\boldsymbol{b} \approx \boldsymbol{A}\boldsymbol{x}$ is

$$\boldsymbol{x}_{LS} = \boldsymbol{A}^{\dagger}\boldsymbol{b} = \boldsymbol{V}\boldsymbol{S}^{\dagger}\,\boldsymbol{U}^T\boldsymbol{b} = \sum_{i=1}^{\mathrm{rank}(\boldsymbol{A})} \frac{\boldsymbol{u}_i^T\boldsymbol{b}}{s_i}\boldsymbol{v}_i. \tag{2.11}$$

If $\boldsymbol{A}$ has some small singular values, then $\boldsymbol{x}_{LS}$ is dominated by the corresponding singular vectors $\boldsymbol{v}_i$. Two popular methods are used for regularization to reduce the influence of components $\boldsymbol{v}_i$ corresponding to small singular values: a truncated SVD method (TSVD) [7, 13] and a Tikhonov method [35]. Briefly speaking, the TSVD simply truncates terms in (2.11) corresponding to singular values close to 0. In contrast, Tikhonov regularization solves the least squares problem

$$\min_{\boldsymbol{x}} ||\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}||^2 + \alpha^2 ||\boldsymbol{L}\boldsymbol{x}||^2, \tag{2.12}$$

where α and $\boldsymbol{L}$ are predetermined. The penalty term $||\boldsymbol{L}\boldsymbol{x}||^2$ restricts the magnitude of the solution $\boldsymbol{x}$ so that the effects of small singular values are reduced.

Returning to our original problem, we use regularization in order to filter out the noise from the principal component $\boldsymbol{x}(t)$. We formulate the linear problem to find a filtered $\widetilde{\boldsymbol{x}}(t)$ as

$$\widetilde{\boldsymbol{z}}(t) = \boldsymbol{U}\,\widetilde{\boldsymbol{x}}(t), \tag{2.13}$$

$$\boldsymbol{z}(t) = \widetilde{\boldsymbol{z}}(t) + \boldsymbol{\epsilon}(t) = \boldsymbol{U}\,\widetilde{\boldsymbol{x}}(t) + \boldsymbol{\epsilon}(t), \tag{2.14}$$

where $\widetilde{\boldsymbol{x}}(t)$ is the filtered principal component,
$\widetilde{\boldsymbol{z}}(t)$ is the resulting filtered data,
$\boldsymbol{\epsilon}(t)$ is the extracted noise.
$\boldsymbol{x}(t)$ in (2.5) is the exact solution of (2.14) when $\boldsymbol{\epsilon}(t) = 0$. By (2.9), we can express $\boldsymbol{x}(t)$ as

$$\boldsymbol{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_k(t) \end{bmatrix} = \begin{bmatrix} s_1\ v_1(t) \\ \vdots \\ s_k v_k(t) \end{bmatrix} = \sum_{i=1}^{k} (s_i v_i(t))\boldsymbol{e}_i,$$

where $\boldsymbol{e}_i$ is the i-th column of the identity matrix. Since we expect that the unimportant principal components $x_i(t)$ are more contaminated by the noise, we reduce the contribution of these principal components. We apply a filtering matrix $\boldsymbol{\Phi} = \text{diag}(\phi_1, \ldots, \phi_k)$ to $\boldsymbol{x}(t)$ with each $\phi_i \in [0, 1]$ so that

$$\widetilde{\boldsymbol{x}}(t) = \boldsymbol{\Phi}\ \boldsymbol{x}(t).$$

The element ϕ_i should be small when s_i is small. The resulting filtered data are

$$\widetilde{\boldsymbol{z}}(t) = \boldsymbol{U}\,\boldsymbol{\Phi}\ \boldsymbol{x}(t), \tag{2.15}$$

$$\widetilde{\boldsymbol{Z}} = \boldsymbol{U}\,\boldsymbol{\Phi}\boldsymbol{X}. \tag{2.16}$$

We introduce two different filtering matrices, $\boldsymbol{\Phi}_{trun}(p)$ and $\boldsymbol{\Phi}_{tikh}(\alpha)$, which correspond to truncated SVD and Tikhonov regularization.

First, we can simply truncate all but p most important components as Elton and Gruber [6] did by using a filtering matrix of $\boldsymbol{\Phi}_{trun}(p) = \text{diag}(\underbrace{1, \ldots, 1}_{p}, \underbrace{0, \ldots, 0}_{k-p})$, so the truncated principal component $\widetilde{\boldsymbol{x}}_{trun}(t)$ is

$$\widetilde{\boldsymbol{x}}_{trun}(t) = \boldsymbol{\Phi}_{trun}(p)\boldsymbol{x}(t).$$

By (2.15) and (2.16), the resulting filtered data are $\widetilde{\boldsymbol{z}}_{trun}(t) = \boldsymbol{U}\boldsymbol{\Phi}_{trun}(p)\boldsymbol{x}(t)$ and $\widetilde{\boldsymbol{Z}}_{trun} = \boldsymbol{U}\,\boldsymbol{\Phi}_{trun}(p)\boldsymbol{X}$. Since $\boldsymbol{X} = \boldsymbol{S}_k\,\boldsymbol{V}_k^T$, we can rewrite $\widetilde{\boldsymbol{Z}}_{trun}$ as

$$\widetilde{\boldsymbol{Z}}_{trun} = \boldsymbol{U}\,\boldsymbol{\Phi}_{trun}(p)(\boldsymbol{S}_k\,\boldsymbol{V}_k^T) = \sum_{i=1}^{p} s_i \boldsymbol{u}_i \boldsymbol{v}_i^T. \tag{2.17}$$

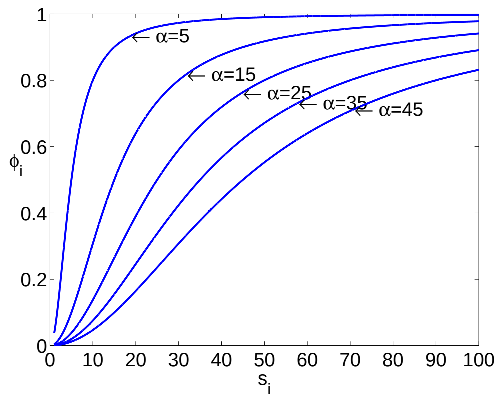


FIG. 2.1. *Tikhonov filtering as a function of s_i for various values of α*

From (2.17), we can see that this truncation method corresponds to the truncated SVD regularization (TSVD) [7, 13].

Second, we can apply the Tikhonov method, and this is our approach to estimating the covariance matrix. We formulate the regularized least squares problem to solve (2.12) as

$$\min_{\widetilde{\boldsymbol{x}}(t)} M(\widetilde{\boldsymbol{x}}(t)) \tag{2.18}$$

with

$$M(\widetilde{\boldsymbol{x}}(t)) = ||\boldsymbol{z}(t) - \boldsymbol{U}\widetilde{\boldsymbol{x}}(t)||^2 + \alpha^2||\boldsymbol{L}\widetilde{\boldsymbol{x}}(t)||^2,$$

where α^2 is a penalty parameter and $\boldsymbol{L}$ is a penalty matrix. The first term $||\boldsymbol{z}(t) - \boldsymbol{U}\widetilde{\boldsymbol{x}}(t)||^2$ forces $\widetilde{\boldsymbol{x}}(t)$ to be close to the exact solution $\boldsymbol{x}(t)$. The second term $||\boldsymbol{L}\widetilde{\boldsymbol{x}}(t)||^2$ controls the size of $\widetilde{\boldsymbol{x}}(t)$. We can choose, for example,

$$\boldsymbol{L} = \text{diag}(s_1^{-1}, \ldots, s_k^{-1}).$$

Let $\widetilde{x}_i(t)$ denote the i-th element of $\widetilde{\boldsymbol{x}}(t)$. The matrix $\boldsymbol{L}$ scales each $\widetilde{x}_i(t)$ by s_i^{-1}, so the unimportant principal components corresponding to small s_i are penalized more than the more important principal components, since we expect that the unimportant principal components $x_i(t)$ are more contaminated by the noise. Thus, the penalty term prevents $\widetilde{\boldsymbol{x}}(t)$ from containing large amounts of unimportant principal components. As we showed before, s_i is proportional to the standard deviation of the i-th principal component $x_i(t)$. Therefore, this penalty matrix $\boldsymbol{L}$ is statistically meaningful considering that the values of $\widetilde{x}_i(t)/s_i$ are in proportion to the normalized principal components $\widetilde{x}_i(t)/\text{std}(x_i(t))$.

The penalty parameter α balances the minimization between the error term $||\boldsymbol{z}(t) - \boldsymbol{U}\widetilde{\boldsymbol{x}}(t)||^2$ and the penalty term $||\boldsymbol{L}\widetilde{\boldsymbol{x}}(t)||^2$. Therefore, as α increases, the regularized solution $\widetilde{\boldsymbol{x}}(t)$ moves away from the exact solution $\boldsymbol{x}(t)$, but should discard more of $\boldsymbol{x}(t)$ as noise. We can quantify this property by determining the solution to (2.18). At the minimizer of (2.18), the gradient of $M(\widetilde{\boldsymbol{x}}(t))$ with respect to each $\widetilde{x}_i(t)$ becomes zero, so

$$\nabla M(\widetilde{\boldsymbol{x}}(t)) = 2\,\boldsymbol{U}^T\boldsymbol{U}\widetilde{\boldsymbol{x}}(t) - 2\,\boldsymbol{U}^T\boldsymbol{z}(t) + 2\alpha^2\boldsymbol{L}^T\boldsymbol{L}\,\widetilde{\boldsymbol{x}}(t) = 0,$$

and thus

$$(\boldsymbol{U}^T \boldsymbol{U} + \alpha^2 \boldsymbol{L}^T \boldsymbol{L})\, \widetilde{\boldsymbol{x}}(t) = \boldsymbol{U}^T \boldsymbol{z}(t).$$

Since $\boldsymbol{U}^T \boldsymbol{U} = \boldsymbol{I}_k,\ \boldsymbol{L} = \mathrm{diag}(s_1^{-1}, \ldots, s_k^{-1})$, and $\boldsymbol{z}(t) = \boldsymbol{U}\boldsymbol{x}(t)$, this becomes

$$\left(\boldsymbol{I}_k^2 + \alpha^2 \,\mathrm{diag}(s_1^{-2}, \ldots, s_k^{-2})\right) \widetilde{\boldsymbol{x}}(t) = \boldsymbol{U}^T \left(\boldsymbol{U}\boldsymbol{x}(t)\right).$$

Therefore,

$$\mathrm{diag}\left(\frac{s_1^2 + \alpha^2}{s_1^2}, \ldots, \frac{s_k^2 + \alpha^2}{s_k^2}\right) \widetilde{\boldsymbol{x}}(t) = \boldsymbol{x}(t),$$

and

$$\widetilde{\boldsymbol{x}}(t) = \mathrm{diag}\left(\frac{s_1^2}{s_1^2 + \alpha^2}, \ldots, \frac{s_k^2}{s_k^2 + \alpha^2}\right) \boldsymbol{x}(t).$$

So, our Tikhonov estimate is

$$\widetilde{\boldsymbol{x}}_{tikh}(t) = \boldsymbol{\Phi}_{tikh}(\alpha)\, \boldsymbol{x}(t),$$

where $\boldsymbol{\Phi}_{tikh}(\alpha)$, called the Tikhonov filtering matrix, denotes $(\boldsymbol{S}_k^2 + \alpha^2 \boldsymbol{I}_k)^{-1} \boldsymbol{S}_k^2$. Thus, we can see that the regularized principal component $\widetilde{\boldsymbol{x}}_{tikh}(t)$ is the result after filtering the original principal component $\boldsymbol{x}(t)$ with the diagonal matrix $\boldsymbol{\Phi}_{tikh}(\alpha)$, whose diagonal elements $\phi_i^{tikh}(\alpha) = \dfrac{s_i^2}{s_i^2 + \alpha^2}$ lie in $[0, 1]$. By (2.15) and (2.16), the resulting filtered data become $\widetilde{\boldsymbol{z}}_{tikh}(t) = \boldsymbol{U}\boldsymbol{\Phi}_{tikh}(\alpha)\, \boldsymbol{x}(t)$ and $\widetilde{\boldsymbol{Z}}_{tikh} = \boldsymbol{U}\boldsymbol{\Phi}_{tikh}(\alpha)\, \boldsymbol{X}$. Let us see how $\phi_i^{tikh}(\alpha)$ changes as α and s_i vary. First, as α increases, $\phi_i^{tikh}(\alpha)$ decreases, as illustrated in Fig. 2.1. This is reasonable since α balances the error term and the penalty term. Second, $\phi_i^{tikh}(\alpha)$ monotonically increases as s_i increases, so the Tikhonov filter matrix reduces the less important principal components more intensely. The main difference between the Tikhonov method and TSVD is that Tikhonov preserves some information from the least important principal components while TSVD discards all of it.

3. Estimate of the covariance matrix $\widetilde{\boldsymbol{\Sigma}}$. Now, we study how $\mathrm{Cov}[\widetilde{\boldsymbol{z}}(t)]$ differs from $\mathrm{Cov}[\boldsymbol{z}(t)]$ after filtering noise, how we can derive a covariance matrix estimate $\widetilde{\boldsymbol{\Sigma}}$ of stock returns from $\mathrm{Cov}[\widetilde{\boldsymbol{z}}(t)]$, and what changes must be applied to make it full rank.

3.1. A covariance estimate. Because $\boldsymbol{z}(t)$ is normalized, the covariance matrix of $\boldsymbol{z}(t)$ is

$$\mathrm{Cov}[\boldsymbol{z}(t)] = \frac{1}{T-1} \boldsymbol{Z}(\boldsymbol{I} - \frac{1}{T}\mathbf{1}\mathbf{1}^T)\boldsymbol{Z}^T = \frac{1}{T-1} \boldsymbol{Z}\boldsymbol{Z}^T.$$

Since $\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{X}$ and $\boldsymbol{X} = \boldsymbol{S}_k \boldsymbol{V}_k^T$,

$$\begin{aligned}
\mathrm{Cov}[\boldsymbol{z}(t)] &= \frac{1}{T-1}(\boldsymbol{U}\boldsymbol{X})\,(\boldsymbol{U}\boldsymbol{X})^T \\
&= \frac{1}{T-1}\boldsymbol{U}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{U}^T \\
&= \frac{1}{T-1}\boldsymbol{U}(\boldsymbol{S}_k \boldsymbol{V}_k^T)(\boldsymbol{S}_k \boldsymbol{V}_k^T)^T\boldsymbol{U}^T \\
&= \frac{1}{T-1}\boldsymbol{U}\boldsymbol{S}_k \boldsymbol{V}_k^T \boldsymbol{V}_k \boldsymbol{S}_k^T \boldsymbol{U}^T \\
&= \frac{1}{T-1}\boldsymbol{U}\boldsymbol{S}_k^2 \boldsymbol{U}^T. \qquad (3.1)
\end{aligned}$$

Now, we calculate the covariance matrices of the filtered data $\widetilde{\boldsymbol{z}}(t)$. First, we compute $\mathbb{E}[\widetilde{\boldsymbol{z}}(t)]$. Since $\widetilde{\boldsymbol{z}}(t) = \boldsymbol{U}\boldsymbol{\Phi}\boldsymbol{x}(t)$ in (2.15) and $\mathbb{E}[\boldsymbol{x}(t)] = \boldsymbol{0}$ in (2.8), the mean of $\widetilde{\boldsymbol{z}}(t)$ is

$$\mathbb{E}[\widetilde{\boldsymbol{z}}(t)] = \mathbb{E}[\boldsymbol{U}\boldsymbol{\Phi}\boldsymbol{x}(t)] = \boldsymbol{U}\boldsymbol{\Phi}\ \mathbb{E}[\boldsymbol{x}(t)] = \boldsymbol{0},$$

so the covariance matrix of $\widetilde{\boldsymbol{z}}(t)$ is

$$\mathrm{Cov}[\widetilde{\boldsymbol{z}}(t)] = \frac{1}{T-1}\widetilde{\boldsymbol{Z}}(\boldsymbol{I} - \frac{1}{T}\boldsymbol{1}\boldsymbol{1}^T)\widetilde{\boldsymbol{Z}}^T = \frac{1}{T-1}\widetilde{\boldsymbol{Z}}\widetilde{\boldsymbol{Z}}^T.$$

Moreover, since $\widetilde{\boldsymbol{Z}} = \boldsymbol{U}\boldsymbol{\Phi}\boldsymbol{X}$,

$$\begin{aligned}
\mathrm{Cov}[\widetilde{\boldsymbol{z}}(t)] &= \frac{1}{T-1}(\boldsymbol{U}\boldsymbol{\Phi}\boldsymbol{X})\,(\boldsymbol{U}\boldsymbol{\Phi}\boldsymbol{X})^T \\
&= \frac{1}{T-1}\boldsymbol{U}\boldsymbol{\Phi}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{\Phi}^T\boldsymbol{U}^T \\
&= \frac{1}{T-1}\boldsymbol{U}\boldsymbol{\Phi}(\boldsymbol{S}_k \boldsymbol{V}_k^T)(\boldsymbol{S}_k \boldsymbol{V}_k^T)^T\boldsymbol{\Phi}^T\boldsymbol{U}^T \\
&= \frac{1}{T-1}\boldsymbol{U}\boldsymbol{\Phi}\boldsymbol{S}_k \boldsymbol{S}_k^T\boldsymbol{\Phi}^T\boldsymbol{U}^T.
\end{aligned}$$

Since $\boldsymbol{\Phi}$ and $\boldsymbol{S}_k$ commute (because they are diagonal),

$$\mathrm{Cov}[\widetilde{\boldsymbol{z}}(t)] = \frac{1}{T-1}\boldsymbol{U}\ \left(\boldsymbol{\Phi}^2\,\boldsymbol{S}_k^2\right)\ \boldsymbol{U}^T. \qquad (3.2)$$

Comparing $\mathrm{Cov}[\boldsymbol{z}(t)]$ in (3.1) and $\mathrm{Cov}[\widetilde{\boldsymbol{z}}(t)]$ in (3.2), we can see that $\mathrm{Cov}[\widetilde{\boldsymbol{z}}(t)]$ is the result of applying the filtering matrix $\boldsymbol{\Phi}^2$ to $\boldsymbol{S}_k^2$ in $\mathrm{Cov}[\boldsymbol{z}(t)]$. Considering that each diagonal element of $\boldsymbol{S}_k^2$ corresponds to an eigenvalue of $\mathrm{Cov}[\boldsymbol{z}(t)]$, the filtering matrix $\boldsymbol{\Phi}^2$ results in attenuating the eigenvalues of $\mathrm{Cov}[\boldsymbol{z}(t)]$. In the previous section, we introduced two filtering matrices :

$$\boldsymbol{\Phi}_{trun}(p) = \mathrm{diag}(\underbrace{1,\ldots,1}_{p},\underbrace{0,\ldots,0}_{k-p}),$$

$$\text{and} \qquad \boldsymbol{\Phi}_{tikh}(\alpha) = \mathrm{diag}(\frac{s_1^2}{s_1^2+\alpha^2},\ldots,\frac{s_k^2}{s_k^2+\alpha^2}).$$

Therefore, $\boldsymbol{\Phi}_{trun}^2(p)$ truncates the eigen-components corresponding to the $(k-p)$ smallest eigenvalues, and $\boldsymbol{\Phi}_{tikh}^2(\alpha)$ down-weights all the eigenvalues at a rate $\left(\dfrac{s_i^2}{s_i^2+\alpha^2}\right)^2 =$

$\left(\dfrac{\lambda_i}{\lambda_i+\alpha^2}\right)^2$ where λ_i is the i-th largest eigenvalue of $\mathrm{Cov}[\boldsymbol{z}(t)]$. Hence, the truncated SVD filtering functions $\phi^2_{trun}(\lambda_i)$ for eigenvalues λ_i become

$$\phi^2_{trun}(\lambda_i) = \begin{cases} 1, & \text{if } i \le p, \\ 0, & \text{otherwise,} \end{cases}$$

and the Tikhonov filtering functions $\phi^2_{tikh}(\lambda_i)$ are

$$\phi^2_{tikh}(\lambda_i) = \left(\frac{\lambda_i}{\lambda_i+\alpha^2}\right)^2.$$

Now we estimate the filtered covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ of stock returns by replacing $\mathrm{Cov}[\boldsymbol{z}(t)]$ in (2.3) with $\mathrm{Cov}[\widetilde{\boldsymbol{z}}(t)]$, so that

$$\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{D}^{\frac{1}{2}}_{\mathrm{Var}[\boldsymbol{r}(t)]}\,\mathrm{Cov}[\widetilde{\boldsymbol{z}}(t)]\,\boldsymbol{D}^{\frac{1}{2}}_{\mathrm{Var}[\boldsymbol{r}(t)]}. \tag{3.3}$$

In addition, by substituting $\mathrm{Cov}[\widetilde{\boldsymbol{z}}(t)]$ in (3.2) into (3.3), we have

$$\widetilde{\boldsymbol{\Sigma}} = \frac{1}{T-1}\left(\boldsymbol{D}^{\frac{1}{2}}_{\mathrm{Var}[\boldsymbol{r}(t)]}\,\boldsymbol{U}\,\left(\boldsymbol{\Phi}^2\,\boldsymbol{S}^2_k\right)\boldsymbol{U}^T\,\boldsymbol{D}^{\frac{1}{2}}_{\mathrm{Var}[\boldsymbol{r}(t)]}\right). \tag{3.4}$$

From now on, we let $\boldsymbol{\Sigma}_{trun}$ and $\boldsymbol{\Sigma}_{tikh}$ denote the estimates resulting from applying $\boldsymbol{\Phi}^2_{trun}(p)$ and $\boldsymbol{\Phi}^2_{tikh}(\alpha)$ to (3.4).

We show the reasonableness of our estimate (3.3) by analyzing how the noise in $\boldsymbol{r}(t)$ affects the covariance estimates. Let $r_i(t)$ denote the observed return of the i-th stock return containing noise. We formulate a noise model for $r_i(t)$ as

$$r_i(t) = r_i^{true}(t) + \nu_i(t),$$

where $r_i^{true}(t)$ is the true i-th stock return, and $\nu_i(t)$ is the noise contained in $r_i(t)$. We assume that $\nu_i(t)$ has zero-mean and is uncorrelated with $r_i^{true}(t)$. With this assumption,

$$\begin{aligned} &\mathbb{E}(r_i^{true}(t)) = \mu_{r_i^{true}}, && \mathrm{Var}(r_i^{true}(t)) = \sigma^2_{r_i^{true}}, \\ &\mathbb{E}(\nu_i(t)) = 0, && \mathrm{Var}(\nu_i(t)) = \sigma^2_{\nu_i}, \end{aligned}$$

so

$$\mathbb{E}(r_i(t)) = \mu_{r_i^{true}}, \quad \mathrm{Var}(r_i(t)) = \sigma^2_{r_i^{true}} + \sigma^2_{\nu_i}.$$

With the notation above, the normalized data $z_i(t)$ become

$$\begin{aligned} z_i(t) &= \frac{r_i(t) - \mathbb{E}(r_i(t))}{Var[r_i(t)]^{\frac{1}{2}}} = \frac{(r_i^{true}(t)+\nu_i(t)) - \mu_{r_i^{true}}}{\sqrt{\sigma^2_{r_i^{true}} + \sigma^2_{\nu_i}}} \\ &= \frac{r_i^{true}(t) - \mu_{r_i^{true}}}{\sqrt{\sigma^2_{r_i^{true}} + \sigma^2_{\nu_i}}} + \frac{\nu_i(t)}{\sqrt{\sigma^2_{r_i^{true}} + \sigma^2_{\nu_i}}}. \end{aligned} \tag{3.5}$$

The first term in (3.5) corresponds to the filtered data $\widetilde{z}_i^{\,true}(t)$ we want to obtain, and the second term in (3.5) corresponds to a noise $\epsilon_i(t)$ we want to remove. Thus, we can define $\widetilde{z}_i^{\,true}(t)$ and $\epsilon_i(t)$ as

$$\widetilde{z}_i^{\,true}(t) = \frac{r_i^{true}(t) - \mu_{r_i^{true}}}{\sqrt{\sigma^2_{r_i^{true}} + \sigma^2_{\nu_i}}} \quad \text{and} \quad \epsilon_i(t) = \frac{\nu_i(t)}{\sqrt{\sigma^2_{r_i^{true}} + \sigma^2_{\nu_i}}}, \tag{3.6}$$

with $z_i(t) = \widetilde{z}_i^{\,true}(t) + \epsilon_i(t)$. Because $\nu_i(t)$ and $z_i(t)$ have zero-mean, $\epsilon_i(t)$ and $\widetilde{z}_i^{\,true}(t)$ in (3.6) also have zero-mean. Hence, $\mathrm{Cov}\left(\widetilde{z}_i^{\,true}(t), \widetilde{z}_j^{\,true}(t)\right)$ becomes

$$\begin{aligned}
\mathrm{Cov}\left(\widetilde{z}_i^{\,true}(t), \widetilde{z}_j^{\,true}(t)\right) &= \frac{1}{T-1}\sum_{t=1}^{T} \widetilde{z}_i^{\,true}(t)\widetilde{z}_j^{\,true}(t) \\
&= \frac{1}{T-1}\sum_{t=1}^{T} \left(\frac{r_i^{true}(t) - \mu_{r_i^{true}}}{\sqrt{\sigma^2_{r_i^{true}} + \sigma^2_{\nu_i}}}\right)\left(\frac{r_j^{true}(t) - \mu_{r_j^{true}}}{\sqrt{\sigma^2_{r_j^{true}} + \sigma^2_{\nu_j}}}\right) \\
&= \frac{\mathrm{Cov}(r_i^{true}(t), r_j^{true}(t))}{\sqrt{\sigma^2_{r_i^{true}} + \sigma^2_{\nu_i}}\sqrt{\sigma^2_{r_j^{true}} + \sigma^2_{\nu_j}}},
\end{aligned}$$

for $i,\ j = 1, \ldots, N$. Thus, we can compute $\mathrm{Cov}(r_i^{true}(t), r_j^{true}(t))$ as

$$\begin{aligned}
\mathrm{Cov}\left(r_i^{true}(t), r_j^{true}(t)\right) &= (\sigma^2_{r_i^{true}} + \sigma^2_{\nu_i})^{\frac{1}{2}}\,\mathrm{Cov}\left(\widetilde{z}_i^{\,true}(t), \widetilde{z}_j^{\,true}(t)\right)(\sigma^2_{r_j^{true}} + \sigma^2_{\nu_j})^{\frac{1}{2}} \\
&= \mathrm{Var}(r_i(t))^{\frac{1}{2}}\ \mathrm{Cov}\left(\widetilde{z}_i^{\,true}(t), \widetilde{z}_j^{\,true}(t)\right)\ \mathrm{Var}(r_j(t))^{\frac{1}{2}}.
\end{aligned}$$

This is the same equation used to compute each element of $\widetilde{\mathbf{\Sigma}}$ in (3.3). Therefore, $\widetilde{\mathbf{\Sigma}}$ from (3.3) should be a good estimator for true stock returns once we successfully filter the noise term $\epsilon_i(t)$ out of $z_i(t)$.

3.2. Rank deficiency of the covariance matrix. Since the covariance matrix is positive semi-definite, the Markowitz portfolio selection problem (1.1) always has a minimizer $\boldsymbol{w}$. However, when the covariance matrix is rank deficient, the minimizer $\boldsymbol{w}$ is not unique, which might not be desirable for investors who want to choose one portfolio. The sample covariance matrix $\mathbf{\Sigma}_{sample}$ from (1.2) has rank $(T-1)$ at most. Therefore, whenever the number of observations T is less than or equal to the number of stocks N, $\mathbf{\Sigma}_{sample}$ is rank deficient. To insure full rank and high quality estimate, we must have at least $N+1$ recent observations of returns, derived from at least $N+1$ recent trades, and this is not always possible.

The covariance matrix estimate $\widetilde{\mathbf{\Sigma}}$ in (3.4) has the same problem. Throughout the normalization, the rank k of the normalized data $\boldsymbol{Z}$ is at most $(T-1)$. By (3.2), the rank of $\mathrm{Cov}[\widetilde{\boldsymbol{z}}(t)]$ is also less than or equal to $(T-1)$ because the filtering matrix $\mathbf{\Phi}^2$ only decreases the eigenvalues of $\mathrm{Cov}[\widetilde{\boldsymbol{z}}(t)]$, so $\widetilde{\mathbf{\Sigma}}$ is also rank deficient by (3.3). Therefore, we may need to modify the filtered covariance matrix to have a unique solution for the Markowitz problem. Prior to resolving this problem, we introduce a perturbation theorem for symmetric matrices.

Theorem 3.1 (Perturbation theorem). *If A and $A+E$ are $n \times n$ symmetric matrices, then*

$$\lambda_k(A) + \lambda_n(E) \le \lambda_k(A+E) \le \lambda_k(A) + \lambda_1(E) \quad \textit{for } k = 1, \ldots, n,$$

where $\lambda_k(X)$ is the k-th largest eigenvalue of a given matrix X.

Proof. See Stewart and Sun [34, p.203]. ∎

Sharpe [32], Ledoit et al. [20], Bengtsson and Holst [2], and Plerou et al. [27] overcome this rank-deficiency problem by replacing the diagonal elements of their modified covariance matrices $\widetilde{\mathbf{\Sigma}}$ with the diagonal elements of the sample covariance matrix $\mathbf{\Sigma}_{sample}$. Let $\widehat{\mathbf{\Sigma}}$ denote the modified covariance matrix after replacing the diagonal elements with the diagonal elements of $\mathbf{\Sigma}_{sample}$. This replacement can be thought of as adding a diagonal matrix $\boldsymbol{D}$ to $\widetilde{\mathbf{\Sigma}}$,

$$\widehat{\mathbf{\Sigma}} = \widetilde{\mathbf{\Sigma}} + \boldsymbol{D}. \tag{3.7}$$

If each diagonal element of $\widetilde{\mathbf{\Sigma}}$ is less than the corresponding diagonal element of $\mathbf{\Sigma}_{sample}$, then $\boldsymbol{D}$ has all positive diagonal elements and $\widehat{\mathbf{\Sigma}}$ has full rank by Theorem 3.1. This strategy assumes that the variances $\text{Var}[\boldsymbol{r}(t)]$ of the stock returns, the diagonal elements of $\mathbf{\Sigma}_{sample}$, are estimated well enough. However, from the viewpoint of noise filtering, $\text{Var}[\boldsymbol{r}(t)]$ is also contaminated by noise if $\boldsymbol{r}(t)$ contains noise. Rather than following this strategy, we propose a different strategy to overcome the rank deficiency.

From (3.6), $\text{Var}(\widetilde{z}_i(t))$ is

$$\text{Var}(\widetilde{z}_i(t)) = \text{Var}\left(\frac{r_i^{true}(t) - \mu_{r_i^{true}}}{\sqrt{\sigma^2_{r_i^{true}} + \sigma^2_{\nu_i}}}\right) = \frac{\sigma^2_{r_i^{true}}}{\sigma^2_{r_i^{true}} + \sigma^2_{\nu_i}} \leq 1. \tag{3.8}$$

By (3.3), we see that the diagonal entries of $\widetilde{\mathbf{\Sigma}}$ are equal to those of $\text{Cov}[\boldsymbol{r}(t)]$ ($= \mathbf{\Sigma}_{sample}$) as assumed in [2, 20, 27, 32], only when the diagonal entries of $\text{Cov}[\widetilde{\boldsymbol{z}}(t)]$ are equal to one. This implies that all $\sigma^2_{\nu_i}$ in (3.8) are zeros. Rather than ignoring the existence of noise in $\text{Var}[\boldsymbol{r}(t)]$, We propose modifying $\text{Cov}[\widetilde{\boldsymbol{z}}(t)]$ as follows :

$$(\text{the } i\text{-th diagonal of } \text{Cov}[\widetilde{\boldsymbol{z}}(t)]) \leftarrow \text{Var}(\widetilde{z}_i(t)) + \delta_i, \tag{3.9}$$

for $i = 1, \dots, N$, where δ_i is a small positive number.

THEOREM 3.2 (Rank modification). *If $Var[\mathbf{r}(t)] > 0$, then replacing the main diagonal of the filtered covariance matrix $Cov[\widetilde{\mathbf{z}}(t)]$ as specified in (3.9) guarantees that the resulting estimate $\widetilde{\mathbf{\Sigma}}$ has full rank.*

Proof. This is a direct consequence of Theorem 3.1 and (3.3). □

As a specific example, we might set

$$(\text{the } i\text{-th diagonal of } \text{Cov}[\widetilde{\boldsymbol{z}}(t)]) \leftarrow (1+\delta) \max_i \left[\text{Var}(\widetilde{z}_i(t))\right], \tag{3.10}$$

for each $i = 1, \dots, N$, where $\delta \ll 1$. This modification makes $\widetilde{\mathbf{\Sigma}}$ have full rank while better conserving the filtered variances.

4. Choice of Tikhonov parameter α. So far, we have seen how to filter noise from the covariance matrix using regularization and how to fix the rank deficiency of the resulting covariance matrix. In order to use Tikhonov regularization, we need to determine the Tikhonov parameter α. In regularization methods for discrete ill-posed problems, there are intensive studies about choosing α using methods such as Generalized Cross Validation [8], L-curves [10, 12], and residual periodograms [30, 31].

In factor analysis and principal component analysis, there are analogous studies to determine the number of factors such as Bartlett's test [1], SCREE test [3], average root [9], partial correlation procedure [39], and cross-validation [40]. More recently, Plerou et al.[26, 27] applied random matrix theory, which will be described in Section 5.6. In the context of arbitrage pricing theory, some different approaches were proposed to determine the number of factors: Trzcinka [37] studied the behavior of

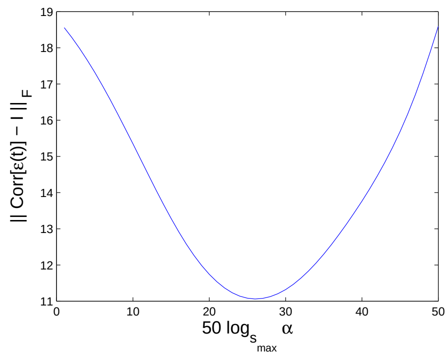


FIG. 4.1. *The difference* $||Corr[\boldsymbol{\epsilon}(t)] - \mathbf{I}_N||_F$ *as a function of log-scaled* α

eigenvalues as the number of assets increases, and Connor and Korajczyk [5] studied the probabilistic behavior of noise factors.

The use of these methods requires various statistical properties for $\boldsymbol{\epsilon}(t)$. We note that since $\mathbb{E}[\boldsymbol{x}(t)] = \mathbf{0}$ by (2.8), the noise $\boldsymbol{\epsilon}(t) = [\epsilon_1(t), \ldots, \epsilon_N(t)]^T$ in (2.14) has zero-mean: By (2.14) and (2.15),

$$\boldsymbol{\epsilon}(t) = \boldsymbol{z}(t) - \boldsymbol{U\Phi x}(t) = \boldsymbol{Ux}(t) - \boldsymbol{U\Phi x}(t) = \boldsymbol{U}(\boldsymbol{I}_N - \boldsymbol{\Phi})\boldsymbol{x}(t).$$

Thus,

$$\mathbb{E}[\boldsymbol{\epsilon}(t)] = \boldsymbol{U}(\boldsymbol{I} - \boldsymbol{\Phi})\ \mathbb{E}[\boldsymbol{x}(t)] = \mathbf{0}. \tag{4.1}$$

We adopt a mutually uncorrelated noise assumption from a factor analysis [16, pp.388-392], so $\text{Corr}[\boldsymbol{\epsilon}(t)] \simeq \boldsymbol{I}_N$. Hence, as a criterion to determine an appropriate parameter α, we formulate an optimization problem minimizing the correlations among the noise,

$$\min_{\alpha \in [s_k, s_1]} ||\ \text{Corr}[\boldsymbol{\epsilon}(t)] - \boldsymbol{I}_N\ ||_F, \tag{4.2}$$

where s_1 and s_k are the largest and the smallest singular values of $\boldsymbol{Z}$ as defined in (2.6). This is similar to Velicer's partial correlation procedure [39] to determine the number of principal components. Fig. 4.1 illustrates an example of $||\text{Corr}[\boldsymbol{\epsilon}(t)] - \boldsymbol{I}_N||_F$ as a function of α in the range $[s_k, s_1]$.

5. Comparison to other estimators. In this section, we compare covariance estimators to our Tikhonov estimator and put them all in a common framework.

5.1. $\boldsymbol{\Sigma}_{sample}$: Sample covariance matrix. A sample covariance matrix is an unbiased covariance matrix estimator that can be computed by (1.2). This is the filtering target of most covariance estimators including our Tikhonov estimator. Thus, the sample covariance matrix $\boldsymbol{\Sigma}_{sample}$ can be thought of as an unfiltered covariance matrix, so the filtering function $\phi_s^2(\lambda_i)$ for eigenvalues of $\text{Cov}[\boldsymbol{z}(t)]$ is

$$\phi_s^2(\lambda_i) = 1 \quad \text{for } i = 1, \ldots, \text{rank}(\boldsymbol{\Sigma}_{sample}).$$

5.2. $\boldsymbol{\Sigma}_{market}$ from the single index market model [32]. Sharpe [32] proposed a single index market model

$$\boldsymbol{r}\ (t) = \boldsymbol{a} + \boldsymbol{b}\ r_m(t) + \boldsymbol{\epsilon}(t), \tag{5.1}$$

where $\boldsymbol{r}(t) \in \mathbb{R}^{N\times 1}$ is stock return at time t,
$r_m(t)$ is market return at time t,
$\boldsymbol{\epsilon}(t) = [\epsilon_1(t), \ldots, \epsilon_N(t)]$ is zero-mean uncorrelated error at time t,
and $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^{N\times 1}$.

This model is based on the assumption that the stock returns $\boldsymbol{r}(t)$ are mainly affected by the market return $r_m(t)$. In this model, the errors $\epsilon_i(t)$ are uncorrelated with each other and with the market return $r_m(t)$. The covariance matrix estimator $\boldsymbol{\Sigma}_{market}$ from the market index model is

$$\boldsymbol{\Sigma}_{market} = \mathrm{Var}(r_m(t))\boldsymbol{b}\boldsymbol{b}^T + \boldsymbol{D}_\epsilon, \tag{5.2}$$

where $\boldsymbol{D}_\epsilon$ is a diagonal matrix containing $\mathrm{Var}[\boldsymbol{\epsilon}(t)]$. By adding the error matrix $\boldsymbol{D}_\epsilon$ in (5.2), $\boldsymbol{\Sigma}_{market}$ has the same diagonal elements as $\boldsymbol{\Sigma}_{sample}$, and it also resolves rank deficiency issues.

Interestingly, King [17, p.150] and Plerou et al. [27, p.8] observed that the principal component corresponding to the largest eigenvalue of the correlation matrix $\mathrm{Corr}[\boldsymbol{r}(t)](= \mathrm{Cov}[\boldsymbol{z}(t)])$ is proportional to the entire market returns. This observation is natural in that most stocks are highly affected by the market situation. Based on their observation, we expect that the most important principal component $x_1(t)$ in (2.5) represents the market return $r_m(t)$. Thus, we can represent the relation between $x_1(t)$ and $r_m(t)$ as

$$x_1(t) \simeq C\ r_m(t) \tag{5.3}$$

for some constant C. From (2.2), the return data $\boldsymbol{r}(t)$ and the normalized data $\boldsymbol{z}(t)$ are related by

$$\boldsymbol{r}(t) = \mathbb{E}[\boldsymbol{r}(t)] + \boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}\boldsymbol{z}(t).$$

By (2.5), we can rewrite the equation above as

$$\begin{aligned}\boldsymbol{r}(t) &= \mathbb{E}[\boldsymbol{r}(t)] + \boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}\sum_{i=1}^{k} x_i(t)\boldsymbol{u}_i \\ &= \left(\mathbb{E}[\boldsymbol{r}(t)] + \boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}\sum_{i=2}^{k} x_i(t)\boldsymbol{u}_i\right) + (\boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}\boldsymbol{u}_1)x_1(t).\end{aligned}$$

By substituting (5.3) into the equation above, we have

$$\boldsymbol{r}(t) \simeq \left(\mathbb{E}[\boldsymbol{r}(t)] + \boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}\sum_{i=2}^{k} x_i(t)\boldsymbol{u}_i\right) + \left(C\boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}\boldsymbol{u}_1\right)\ r_m(t). \tag{5.4}$$

Hence, by comparing (5.1) and (5.4), we can see that $\boldsymbol{a}$ in (5.1) corresponds to the first term in (5.4) and $\boldsymbol{b}$ in (5.1) corresponds to $\left(C\boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}\boldsymbol{u}_1\right)$ in (5.4). Moreover, replacing $r_m(t)$ and $\boldsymbol{b}$ in (5.2) with $\dfrac{x_1(t)}{C}$ and $\left(C\boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}\boldsymbol{u}_1\right)$, we have

$$\begin{aligned}\boldsymbol{\Sigma}_{market} &\simeq \mathrm{Var}\left(\frac{x_1(t)}{C}\right)\left(C\boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}\boldsymbol{u}_1\right)\left(C\boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}\boldsymbol{u}_1\right)^T + \boldsymbol{D}_\epsilon \\ &= \mathrm{Var}\left(x_1(t)\right)\boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}\boldsymbol{u}_1\boldsymbol{u}_1^T\boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]} + \boldsymbol{D}_\epsilon.\end{aligned}$$

By the relation of $\mathrm{Var}(x_i(t))$ and s_i in (2.10),

$$\begin{aligned}\boldsymbol{\Sigma}_{market} &\simeq \frac{s_1^2}{T-1}\boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}\boldsymbol{u}_1\boldsymbol{u}_1^T\boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]} + \boldsymbol{D}_\epsilon \\ &= \frac{1}{T-1}\left(\boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}(s_1^2\boldsymbol{u}_1\boldsymbol{u}_1^T)\boldsymbol{D}_{\mathrm{Var}[\boldsymbol{r}(t)]}\right) + \boldsymbol{D}_\epsilon.\end{aligned}$$

By comparing the equation above with (3.4), we can think of $\boldsymbol{\Sigma}_{market}$ as an implicit truncation of all but the largest eigen-component of $\mathrm{Cov}[\boldsymbol{z}(t)]$. Therefore, we can represent the filtering function $\phi_m^2(\lambda_i)$ for $\boldsymbol{\Sigma}_{market}$ as

$$\phi_m^2(\lambda_i) \simeq \begin{cases} 1, & \text{if } i = 1, \\ 0, & \text{otherwise.}\end{cases} \tag{5.5}$$

5.3. $\boldsymbol{\Sigma}_{s\to m}$: Shrinkage toward $\boldsymbol{\Sigma}_{market}$ [20]. Ledoit et al. propose a shrinkage method from $\boldsymbol{\Sigma}_{sample}$ to $\boldsymbol{\Sigma}_{market}$ as

$$\boldsymbol{\Sigma}_{s\to m} = \beta\ \boldsymbol{\Sigma}_{market} + (1-\beta)\boldsymbol{\Sigma}_{sample}, \tag{5.6}$$

where $0 \le \beta \le 1$. Thus, the shrinkage estimator is the weighed average of $\boldsymbol{\Sigma}_{sample}$ and $\boldsymbol{\Sigma}_{market}$. In order to find an optimal weight β, they minimize the distance between $\boldsymbol{\Sigma}_{s\to m}$ and the true covariance matrix $\boldsymbol{\Sigma}_{true}$:

$$\min_\beta ||\boldsymbol{\Sigma}_{s\to m} - \boldsymbol{\Sigma}_{true}||_F^2.$$

Since the true covariance matrix $\boldsymbol{\Sigma}_{true}$ is unknown, they use an asymptotic variance to determine an optimal β. (Refer to [20, Section 2.5-6] for a detailed description.) Considering that $\boldsymbol{\Sigma}_{market}$ is the result of the implicit truncation method, we can think of this shrinkage method as implicitly down-weighting all eigenvalues but the largest at a rate $(1-\beta)$. Therefore, we can represent the filtering function $\phi_{s\to m}^2(\lambda_i)$ as

$$\phi_{s\to m}^2(\lambda_i) \simeq \begin{cases} 1, & \text{if } i = 1, \\ 1-\beta,\ \text{where } 0 \le \beta \le 1 & \text{otherwise.}\end{cases} \tag{5.7}$$

The full rank of $\boldsymbol{\Sigma}_{s\to m}$ comes from the full rank of $\boldsymbol{\Sigma}_{market}$ in (5.6) which replaces the diagonal elements with diagonal elements of $\boldsymbol{\Sigma}_{sample}$.

5.4. Truncated covariance matrix $\boldsymbol{\Sigma}_{trun}$ [6]. As mentioned in Section 3.1, the truncated covariance matrix $\boldsymbol{\Sigma}_{trun}$ has the filtering function $\phi_{trun}^2(\lambda_i)$ for the eigenvalues λ_i of $\mathrm{Cov}[\boldsymbol{z}(t)]$, where

$$\phi_{trun}^2(\lambda_i) \simeq \begin{cases} 1, & \text{if } i = 1,\dots,p, \\ 0, & \text{otherwise.}\end{cases} \tag{5.8}$$

Thus, the model of Elton and Gruber [6] truncates all but the p largest eigen-components of $\mathrm{Cov}[\boldsymbol{z}(t)]$. They did not comment about the rank-deficiency problem.

5.5. $\boldsymbol{\Sigma}_{s\to trun}$: Shrinkage toward $\boldsymbol{\Sigma}_{trun}$ [2]. Bengtsson and Holst propose a shrinkage estimator from $\boldsymbol{\Sigma}_{sample}$ to $\boldsymbol{\Sigma}_{trun}$ as

$$\boldsymbol{\Sigma}_{s\to trun} = \beta\ \boldsymbol{\Sigma}_{trun} + (1-\beta)\boldsymbol{\Sigma}_{sample}, \tag{5.9}$$

where $0 \le \beta \le 1$. They determine the parameter β in a way similar to [20]. (Refer to [2, Section 4.1-4.2] for detailed description.) Therefore, $\boldsymbol{\Sigma}_{s\to trun}$ is a variant of the shrinkage method toward $\boldsymbol{\Sigma}_{trun}$. Because $\boldsymbol{\Sigma}_{trun}$ is the truncated covariance matrix containing the p most significant eigen-components of $\mathrm{Cov}[\boldsymbol{z}(t)]$, we can regard $\boldsymbol{\Sigma}_{s\to trun}$ as damping the smallest eigenvalues by $(1-\beta)$. Thus, the filtering function corresponding to this approach is

$$\phi^2_{s\to trun}(\lambda_i) = \begin{cases} 1, & \text{if } i = 1,\ldots,p\ , \\ 1-\beta, \ \text{where } 0 \le \beta \le 1, & \text{otherwise.} \end{cases} \tag{5.10}$$

Rather than removing all the least important principal components as Elton and Gruber did, Bengtsson and Holst try to preserve the potential information of unimportant principal components by this single-rate attenuation. They ensure that $\boldsymbol{\Sigma}_{s\to trun}$ has full rank by replacing the diagonal elements of $\boldsymbol{\Sigma}_{trun}$ with the diagonal elements of $\boldsymbol{\Sigma}_{sample}$ [2, p.8]. Bengtsson and Holst conclude that their shrinkage matrix $\boldsymbol{\Sigma}_{s\to trun}$ performed best in the Swedish stock market when the shrinkage target $\boldsymbol{\Sigma}_{trun}$ takes only the most significant principal component ($p = 1$). They also mention that the result is consistent with RMT because only the largest eigenvalue deviates far from the range of $[\lambda_{\min}, \lambda_{\max}]$.

5.6. $\boldsymbol{\Sigma}_{RMT:trun}$ truncation by random matrix theory [27]. Plerou et al. [27] apply random matrix theory (RMT) [23] which shows that the eigenvalues of a random correlation matrix have a distribution within an interval determined by the ratio of N and T. Let Corr_{random} be a random correlation matrix

$$\mathrm{Corr}_{random} = \frac{1}{T}\boldsymbol{A}\boldsymbol{A}^T, \tag{5.11}$$

where $\boldsymbol{A} \in \mathbb{R}^{N\times T}$ contains mutually uncorrelated random elements $a_{i,t}$ with zero-mean and unit variance. When $Q = T/N \ge 1$ is fixed, the eigenvalues λ of Corr_{random} have a limiting distribution (as $N \to \infty$)

$$f(\lambda) = \begin{cases} \dfrac{Q}{2\pi\sigma^2}\dfrac{\sqrt{(\lambda_{\max}-\lambda)(\lambda_{\min}-\lambda)}}{\lambda}, & \lambda_{\min} \le \lambda \le \lambda_{\max}, \\ 0, & \text{otherwise,} \end{cases} \tag{5.12}$$

where σ^2 is the variance of the elements of $\boldsymbol{A}$, $\lambda_{\min} \le \lambda \le \lambda_{\max}$, and $\lambda^{\max}_{\min} = \sigma^2\left(1+\frac{1}{Q}\pm 2\sqrt{\frac{1}{Q}}\right)$. By comparing the eigenvalue distribution of $\mathrm{Corr}[\boldsymbol{r}(t)]$ with $f(\lambda)$, Plerou et al. show that most eigenvalues are within $[\lambda_{\min}, \lambda_{\max}]$. They conclude that only a few large eigenvalues deviating from $[\lambda_{\min}, \lambda_{\max}]$ correspond to eigenvalues of the real correlation matrix, so the other eigen-components should be removed from $\mathrm{Corr}[\boldsymbol{r}(t)]$. Thus, the filtering function $\phi^2_{RMT:trun}(\lambda_i)$ for the eigenvalue λ_i of $\mathrm{Corr}[\boldsymbol{r}(t)]$ is

$$\phi^2_{RMT:trun}(\lambda_i) = \begin{cases} 1, & \text{if } \lambda_i \ge \lambda_{\max}\ , \\ 0, & \text{otherwise.} \end{cases} \tag{5.13}$$

Therefore, the estimator of Plerou et al. is a truncation method with a criterion to determine the number of principal components. The covariance matrix after truncation is rank deficient, like $\boldsymbol{\Sigma}_{trun}$. They set the diagonal of $\mathrm{Cov}[\widetilde{\boldsymbol{z}}(t)]$ to all ones after truncating the smallest eigenvalues [27, p.14]. Thus, this modification makes $\boldsymbol{\Sigma}_{RMT:trun}$ have the same diagonal elements as $\boldsymbol{\Sigma}_{sample}$, which is the same strategy as Sharpe [32], Ledoit et al. [20], and Bengtsson and Holst [2] used.

Estimator	Filtering function $\phi^2(\lambda_i)$
$\mathbf{\Sigma}_{sample}$	$\phi_s^2(\lambda_i) = 1$
$\mathbf{\Sigma}_{market}$[32]	$\phi_m^2(\lambda_i) \simeq \begin{cases} 1, & \text{if } i = 1, \\ 0, & \text{otherwise.} \end{cases}$
$\mathbf{\Sigma}_{s \to m}$[20]	$\phi_{s \to m}^2(\lambda_i) \simeq \begin{cases} 1, & \text{if } i = 1, \\ 1 - \beta, & \text{otherwise.} \end{cases}$
$\mathbf{\Sigma}_{trun}$[6]	$\phi_{trun}^2(\lambda_i) = \begin{cases} 1, & \text{if } i = 1, \ldots, p, \\ 0, & \text{otherwise.} \end{cases}$
$\mathbf{\Sigma}_{s \to trun}$[2]	$\phi_{s \to trun}^2(\lambda_i) = \begin{cases} 1, & \text{if } i = 1, \ldots, p, \\ 1 - \beta, & \text{otherwise.} \end{cases}$
$\mathbf{\Sigma}_{RMT:trun}$[27]	$\phi_{RMT:trun}^2(\lambda_i) = \begin{cases} 1, & \text{if } \lambda_i \geq \lambda_{\max}\ , \\ 0, & \text{otherwise.} \end{cases}$
$\mathbf{\Sigma}_{RMT:repl}$[19]	$\phi_{RMT:repl}^2(\lambda_i) = \begin{cases} 1, & \text{if } \lambda_i \geq \lambda_{\max}\ , \\ \frac{C}{\lambda_i}, & \text{otherwise.} \end{cases}$
$\mathbf{\Sigma}_{tikh}$	$\phi_{tikh}^2(\lambda_i) = \left(\frac{\lambda_i}{\lambda_i + \alpha^2} \right)^2$

Table 5.1
Definition of the filter function $\phi^2(\lambda_i)$ for each covariance estimator where $i = 1, \ldots, \mathrm{rank}(\mathbf{\Sigma}_{sample})$.

5.7. $\mathbf{\Sigma}_{RMT:repl}$ replacing the RMT eigenvalues [19]. Laloux et al. apply RMT to this problem in a way somewhat different from Plerou et al. First, they find the best fitting σ^2 in (5.12) to the eigenvalue distribution of the observed correlation matrix rather than taking a value of 1. Second, they replace each eigenvalue in the RMT interval with a constant value C, chosen so that the trace of the matrix is unchanged. Thus, the filtering function $\phi_{RMT:repl}^2(\lambda_i)$ for eigenvalues is

$$\phi_{RMT:repl}^2(\lambda_i) = \begin{cases} 1, & \text{if } \lambda_i \geq \lambda_{\max}\ , \\ \frac{C}{\lambda_i}, & \text{otherwise.} \end{cases} \tag{5.14}$$

By this replacement of small eigenvalues with a positive constant, the estimator $\mathbf{\Sigma}_{RMT:repl}$ avoids the rank-deficiency problem.

5.8. Tikhonov covariance matrix $\mathbf{\Sigma}_{tikh}$. As mentioned at Section 3.1, the Tikhonov covariance matrix $\mathbf{\Sigma}_{tikh}$ has the filtering function $\phi_{tikh}^2(\lambda_i)$ for the eigenvalues λ_i of $\mathrm{Cov}[\boldsymbol{z}(t)]$, where

$$\phi_{tikh}^2(\lambda_i) = \left(\frac{\lambda_i}{\lambda_i + \alpha^2} \right)^2 , \tag{5.15}$$

where the parameter α is determined as described in Section 4.

5.9. Comparison. The derivations in Section 5.1–5.8 provide the proof of the following theorem.

THEOREM 5.1 (Filtering functions). *The eight covariance estimators are characterized by the choice of filtering functions specified in Table 5.1.*

Tikhonov filtering preserves potential information from unimportant principal components corresponding to small eigenvalues, rather than truncating them all like $\mathbf{\Sigma}_{market}$, $\mathbf{\Sigma}_{trun}$, and $\mathbf{\Sigma}_{RMT:trun}$. In contrast to the single-rate attenuation of $\mathbf{\Sigma}_{s\to m}$ and $\mathbf{\Sigma}_{s\to trun}$ and the constant value replacement of $\mathbf{\Sigma}_{RMT:repl}$, Tikhonov filtering reduces the effect of the smallest eigenvalues more intensely. This gradual down-weighting with respect to the magnitude of eigenvalues is the key difference between the Tikhonov method and other estimators.

All the estimators proposed by [2, 20, 27, 32] overcome the rank deficiency of their covariance estimator by replacing the diagonal elements with the diagonal elements of $\mathbf{\Sigma}_{sample}$. In contrast to these estimators, our Tikhonov estimator uses the novel strategy described in Section 3.2.

6. Experiment. In this section, we evaluate the covariance estimators using the daily return data of the most frequently traded 112 stocks in the NYSE, AMEX, and NASDAQ. We collect the daily return data from 2006 January 3 to 2007 December 31 from the CRSP database (the Center for Research in Security Prices). (There are 502 trading days in the 2 years.) In order to test the performance of the Markowitz portfolio derived from each covariance estimator, we simulate portfolio construction under the following scenario. Using daily return data from the previous 112 days, which is the *in-sample* period, we estimate the covariance matrix. Thus, we use the 112×112 stock return data $\boldsymbol{R}$ to estimate a covariance matrix. With the estimated covariance matrix, we solve the Markowitz portfolio selection problem in (5.1) to construct a portfolio, and we hold the portfolio for 10 days, which is the *out-of-sample* period. We repeat this process, starting the experiments at days $1, 11, 21, \ldots, 381$, so we re-balance the portfolio 39 times. We compute the portfolio returns for each *out-of-sample* period, and calculate the standard deviation of these portfolio returns. A smaller standard deviation indicates a better covariance estimator which decreases the risk of the portfolio. In this experiment, we set the target portfolio return q in (5.1) to zero because the standard deviation of the portfolio return is generally minimized for $q = 0$.

6.1. Covariance estimators in experiments. We repeat the experiment above for each covariance estimator in Table 5.1 plus two diagonal matrices, $\mathbf{\Sigma}_{var}$ and $\mathbf{\Sigma}_{\boldsymbol{I}}$, for a total of 10 estimators. $\mathbf{\Sigma}_{var}$ is a diagonal matrix whose diagonal elements are equal to $\mathrm{Var}[\boldsymbol{r}(t)]$. Thus, $\mathbf{\Sigma}_{var}$ only contains variance information for each stock return, and assumes that the stock returns are uncorrelated with each other. $\mathbf{\Sigma}_{\boldsymbol{I}}$ is an $N \times N$ identity matrix. Since $\mathbf{\Sigma}_{sample}$ is rank deficient, we modify it by adding a small diagonal matrix $\boldsymbol{D}$, as in (3.9). To compute $\mathbf{\Sigma}_{market}$ and $\mathbf{\Sigma}_{s\to m}$, we need the daily market return data $r_m(t)$ in (5.1). In this experiment, we adopt equally-weighted market portfolio returns including distributions from CRSP database as $r_m(t)$. According to Ledoit et al. [20, p.607], an equally-weighted market portfolio is better than a value-weighted market portfolio for explaining stock market variances.

The parameters of p for $\mathbf{\Sigma}_{trun}$ and $\mathbf{\Sigma}_{s\to trun}$ are static, which means the parameter value remains constant over all time periods. However, in order to find a statically optimal value p, we perform the experiments varying these p in (5.8) and (5.10) from 1 to 20. In contrast, the parameters of β for $\mathbf{\Sigma}_{s\to m}$ and $\mathbf{\Sigma}_{s\to trun}$, p for $\mathbf{\Sigma}_{RMT:trun}$ and $\mathbf{\Sigma}_{RMT:repl}$, and α for $\mathbf{\Sigma}_{tikh}$ have their own parameter-choosing methods as described

Estimators	Standard deviation	Annual %	optimal p
$\mathbf{\Sigma}_{sample}$	1.147×10^{-2}	18.17%	–
$\mathbf{\Sigma}_{\boldsymbol{I}}$	2.494×10^{-3}	3.95%	–
$\mathbf{\Sigma}_{var}$	1.949×10^{-3}	3.09%	–
$\mathbf{\Sigma}_{market}$	1.793×10^{-3}	2.84%	–
$\mathbf{\Sigma}_{RMT:repl}$	1.710×10^{-3}	2.71%	$p = 2$ or 3
$\mathbf{\Sigma}_{RMT:trun}$	1.701×10^{-3}	2.69%	$p = 2$ or 3
$\mathbf{\Sigma}_{trun}$	1.656×10^{-3}	2.62%	$p = 9$
$\mathbf{\Sigma}_{s\to m}$	1.604×10^{-3}	2.54%	–
$\mathbf{\Sigma}_{s\to trun}$	1.580×10^{-3}	2.50%	$p = 1$
$\mathbf{\Sigma}_{tikh}$	1.564×10^{-3}	2.48%	–

TABLE 6.1
The standard deviation of portfolio returns and its annualized value for each covariance matrix estimator.

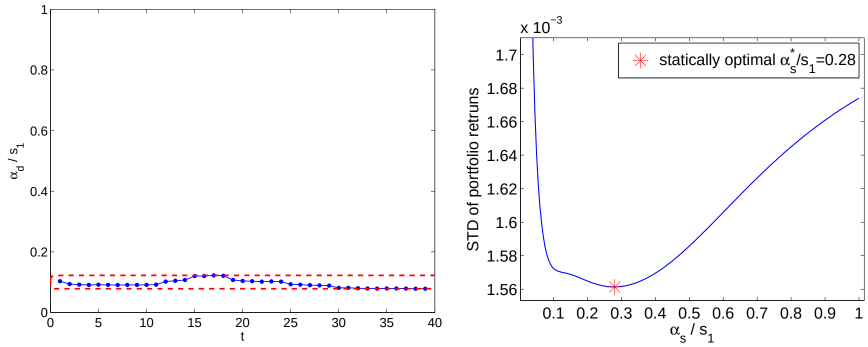
in Section 5, so we dynamically determine these parameters each time the portfolio is re-balanced.

6.2. Results.

6.2.1. Standard deviation of portfolio returns. The standard deviation from each covariance estimator is summarized in Table 6.1. We annualize the standard deviations by multiplying by $\sqrt{201}$, since there were 201 trading days per year.

Surprisingly, $\mathbf{\Sigma}_{\boldsymbol{I}}$ and $\mathbf{\Sigma}_{var}$ work better than $\mathbf{\Sigma}_{sample}$ in our experiment, which shows the seriousness of noise contamination in $\mathbf{\Sigma}_{sample}$. $\mathbf{\Sigma}_{trun}$ is best when the parameter $p = 9$. The estimators of $\mathbf{\Sigma}_{RMT:trun}$ and $\mathbf{\Sigma}_{RMT:repl}$ choose $p = 2$ or $p = 3$, which shows that these estimators based on RMT trust a relatively small number of principal components compared to the statically optimal $p = 9$ for $\mathbf{\Sigma}_{trun}$. In contrast, $\mathbf{\Sigma}_{s\to trun}$ is best when the parameter $p = 1$, which is the same result as Bengtsson and Holst [2] obtained using Swedish stock return data. The shrinkage estimators $\mathbf{\Sigma}_{s\to m}$ and $\mathbf{\Sigma}_{s\to trun}$ outperform the corresponding target estimators $\mathbf{\Sigma}_{market}$ and $\mathbf{\Sigma}_{trun}$, which implies that down-weighting less important principal components is better than completely truncating them. Our Tikhonov estimator $\mathbf{\Sigma}_{tikh}$ outperforms the shrinkage estimators, which shows that a gradual down-weighting reduces noise better than a single-rate down-weighting. Experiments with other time periods showed similar results, except that the RMT estimators cannot be used when the number of time periods is less than the number of stocks.

6.2.2. Stability of Tikhonov parameter choice. In this section, we evaluate the stability of our parameter choice method from Section 4. Fig. 6.1(a) illustrates the change of the ratio of the dynamically chosen Tikhonov parameter α_D to the largest singular value s_1 of $\mathrm{Cov}[\boldsymbol{z}(t)]$. The ratio of α_D/s_1 is bounded in a range of $[0.0781, 0.1222]$ during the experiment, and has a standard deviation of 0.0127. Thus, the parameter choice method was quite stable during the whole experiment.



(a) Variation in the dynamic α_D over the course of the experiment.

(b) Standard deviation of portfolio returns for various choices of α_S

FIG. 6.1. *Performance of static α_S and dynamically chosen α_D.*

We repeated our numerical experiment keeping the ratio α/s_1 constant over all time periods. (We use the notation α_S for this statically determined parameter.) This statical parameter choice may not be practical in real market trading, since we cannot access the future return information when we construct a portfolio. However, we can find a statically optimal ratio from this experiment for a comparison to α_D/s_1. Fig. 6.1(b) shows how the standard deviation of portfolio returns changes as the ratio increases. The optimal ratio α_S^*/s_1 was 0.28 whose resulting standard deviation of portfolio returns was 1.561×10^{-3} (annual 2.47%). The dynamically chosen ratios α_D/s_1 are relatively small compared to the statically optimal ratio α_S^*/s_1, but the result of α_D/s_1 is quite close to the result of the statically optimal α_S^*/s_1 in minimizing the risk of Markowitz portfolio. Therefore, we can see the reasonableness of our parameter choice method from this comparison.

7. Conclusion. In this study, we applied Tikhonov regularization to improve the covariance matrix estimate used in the Markowitz portfolio selection problem. We put the previous covariance estimators in a common framework based on the filtering function $\phi^2(\lambda_i)$ for the eigenvalues of Cov$[\boldsymbol{z}(t)]$. The Tikhonov estimator $\boldsymbol{\Sigma}_{tikh}$ attenuates smaller eigenvalues more intensely, which is a key difference between it and the other filter functions. We also proposed a new approach to overcome the rank deficiency of the covariance matrix estimate which better accounts for noise.

In order to choose an appropriate Tikhonov parameter α that determines the intensity of attenuation, we formulated an optimization problem minimizing the difference between Corr$[\boldsymbol{\epsilon}(t)]$ and $\boldsymbol{I}$ based on the assumption that noise in stock return data is uncorrelated. Finally, we demonstrated the superior performance of our estimator using the most frequently traded 112 stocks in NYSE, AMEX, and NASDAQ.

REFERENCES

[1] Bartlett. Tests of significnance in factor analysis. *Br. J. Psych. Stat. Sec.*, 3:77–85, 1950.
[2] Christoffer Bengtsson and Jan Holst. On portfolio selection: Improved covariance matrix estimation for Swedish asset returns. In *31st Meeting, Euro Working Group on Financial Modeling*. Hermes Center, Nov 2002. available as `http://www.hermes.ucy.ac.cy/conferences/ewgfm/papers/bengtsson_holst_20021004.pdf`.
[3] R. B. Cattell. The scree test for the number of factors. *Mult. Behav. Res.*, 1:245–276, 1966.

[4] T. Conlon, H. J. Ruskin, and M. Crane. Random matrix theory and fund of funds portfolio optimisation. *Physica*, 382:565–576, August 2007.

[5] Gregory Connor and Robert A. Korajczyk. A test for the number of factors in an approximate factor model. *The Journal of Finance*, 48(4):1263–1291, Sep 1993.

[6] Edwin J. Elton and Martin J. Gruber. Estimating the dependence structure of share prices – implications for portfolio selection. *The Journal of Finance*, 28(5):1203–1232, 1973.

[7] G. Golub and W. Kahan. Calculating the singular values and pseudo-inverse of a matrix. *J. Soc. Indust. Appl. Math.: Ser. B, Numer. Anal.*, 2:205–224, 1965.

[8] Gene H. Golub, Michael Heath, and Grace Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.

[9] L. Guttman. Some necessary conditions for common-factor analysis. *Psychometrika*, 19:149–161, 1954.

[10] Per Christian Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34:561–580, 1992.

[11] Per Christian Hansen, James G. Nagy, and Dianne P. O'Leary. *Deblurring Images: Matrices, Spectra, and Filtering*. SIAM, 2006.

[12] Per Christian Hansen and Dianne P. O'Leary. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.*, 14(6):1487–1503, 1993.

[13] Richard J. Hanson. A numerical method for solving Fredholm integral equations of the first kind using singular values. *SIAM J. Numer. Anal.*, 8:616–622, 1971.

[14] A. E. Hoerl and R. W. Kennard. Ridge regression: Applications to nonorthogonal problems. *Technometrics*, 12:69–82, 1970.

[15] A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67, 1970.

[16] J. Edward Jackson. *A User's Guide to Principal Components*. Wiley-IEEE, 2003.

[17] Benjamin F. King. Market and industry factors in stock price behavior. *The Journal of Business*, 39(1):139–190, Jan 1966. Part 2: Supplement on Security Prices.

[18] J. Kwapień, S. Drożdż, and P. Oświęcimka. The bulk of the stock market correlation matrix is not pure noise. *Physica*, 359:589–606, January 2006.

[19] Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud, and Marc Potters. Noise dressing of financial correlation matrices. *Phys. Rev. Lett.*, 83(7):1467–1470, Aug 1999.

[20] Olivier Ledoit and Michael Wolf. Improved estimation of the covariance matrix of stock returns with an application to portofolio selection. *J. of Empirical Finance*, 10:603–621, 2003.

[21] John Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, pages 13–37, Feb 1965.

[22] H. Markowitz. Portfolio selection. *J. Finance*, 7:77–91, 1952.

[23] Madan Lal Mehta. *Random Matrices*. 3rd edition, 2004. Cited by [27].

[24] Jan. Mossin. Equilibrium in a capital asset market. *Journal of Finance*, pages 768–783, Oct 1966.

[25] David L. Phillips. A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.*, 9:84–97, 1962.

[26] Vasiliki Plerou, Parameswaran Gopikrishnan, Luís A. Nunes Amaral, Martin Meyer, and H. Eugene Stanley. Scaling of the distribution of price fluctuations of individual companies. *Phys. Rev. E*, 60(6):6519–6529, Dec 1999.

[27] Vasiliki Plerou, Parameswaran Gopikrishnan, Bernd Rosenow, Luís A. Nunes Amaral, Thomas Guhr, and H. Eugene Stanley. Random matrix approach to cross correlations in financial data. *Physical Review E*, 65(6):066126.1–066126.18, June 2002.

[28] Richard Roll and Stephen A. Ross. An empirical investigation of the arbitrage pricing theory. *Journal of Finance*, 35(5):1073–1103, Dec 1980.

[29] Stephen A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13:341–360, 1976.

[30] Bert W. Rust. Parameter selection for constrained solutions to ill-posed problems. *Computing Science and Statistics*, 32:333–347, 2000.

[31] Bert W. Rust and Dianne P. O'Leary. Residual periodograms for choosing regularization parameters for ill-posed problems. *Inverse Problems*, 24, 2008.

[32] William F. Sharpe. A simplified model for portfolio analysis. *Management Science*, 9:277–293, 1963.

[33] William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19:425–442, Sep 1964.

[34] G. W. Stewart and Ji guang Sun. *Matrix Perturbation Theory*. Academic Press, 1990.

[35] A. N. Tikhonov and V. Y. Arsenin. *Solution of Ill-posed Problems*. John Wiley & Sons, 1977.

[36] Jack L. Treynor. Toward a theory of the market value of risky assets. *Unpublished manuscript*, 1961.

[37] Charles Trzcinka. On the number of factors in the arbitrage pricing model. *The Journal of Finance*, 41(2):347–368, June 1986.

[38] S. Twomey. On the numerical solution of Fredholm integral equations of the first kind by inversion of the linear system produced by quadrature. *J. Assoc. Comput. Mach.*, 10:97–101, 1963.

[39] W. F. Velicer. Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41(3):321–327, 1976.

[40] S. Wold. Cross validatory estimation of the number of components in factor and principal component analysis. *Technometrics*, 20:397–405, 1978.