SIAM J. Financial Math.
Vol. 1, pp. 932–961

Portfolio Selection Using Tikhonov Filtering to Estimate the Covariance Matrix*

Sungwoo Park† and Dianne P. O'Leary‡

Abstract. Markowitz's portfolio selection problem chooses weights for stocks in a portfolio based on an estimated covariance matrix of stock returns. Our study proposes reducing noise in the estimation by using a Tikhonov filter function. In addition, we prevent rank deficiency of the estimated covariance matrix and propose a method for effectively choosing the Tikhonov parameter, which determines the filtering intensity. We put previous estimators into a common framework and compare their filtering functions for eigenvalues of the correlation matrix. We demonstrate the effectiveness of our estimator using stock return data from 1958 through 2007.

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

AMS subject classifications. 91G10, 91G60, 65F22, 62J07, 62J10

DOI. 10.1137/090749372

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, a mean-variance (MV) problem [33] finds a stock weight vector $\boldsymbol{w}$ to determine a portfolio that minimizes the 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 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. On the other hand, a global minimum variance (GMV) problem finds a portfolio that minimizes the variances of the portfolio returns without the return constraint:

$$\min_{\boldsymbol{w}} \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w} \text{ subject to } \quad \boldsymbol{w}^T \mathbf{1} = 1. \tag{1.2}$$

Even though these optimization problems play a central role in a modern portfolio theory, it has been observed that the solutions are very sensitive to their input parameters [3, 5, 6, 7]. Thus, in order to construct a good portfolio using these formulations, the covariance matrix $\boldsymbol{\Sigma}$ must be well estimated. We let $\widetilde{\boldsymbol{\Sigma}}$ denote an estimate of $\boldsymbol{\Sigma}$ and $\widetilde{\boldsymbol{\Sigma}}_{method}$ denote a resulting estimate by a particular method.

*Received by the editors February 11, 2009; accepted for publication (in revised form) September 27, 2010; published electronically December 14, 2010.
http://www.siam.org/journals/sifin/1/74937.html
†Department of Computer Science, University of Maryland, College Park, MD 20742 (swpark81@gmail.com).
‡Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742 (oleary@cs.umd.edu). This author's work was supported in part by NSF CCF 05 14213.

Let $\boldsymbol{R} = [\boldsymbol{r}(1), \ldots, \boldsymbol{r}(T)]$ be an $N \times T$ matrix containing observations on N stocks' returns for each of T times. A conventional estimator—a sample covariance matrix $\widetilde{\boldsymbol{\Sigma}}_{sample}$—can be computed from the stock return matrix $\boldsymbol{R}$ as

$$\widetilde{\boldsymbol{\Sigma}}_{sample} = \frac{1}{T}\boldsymbol{R}\left(\boldsymbol{I}_T - \frac{1}{T}\mathbf{1}\mathbf{1}^T\right)\boldsymbol{R}^T, \tag{1.3}$$

where $\boldsymbol{I}_T$ denotes a $T \times T$ identity matrix. From classical statistics, $\widetilde{\boldsymbol{\Sigma}}_{sample}$ is a consistent estimate for fixed N; in our case, since T is fixed and of the same order as N, this result is not so useful. Moreover, since the stock return matrix $\boldsymbol{R}$ contains noise, the sample covariance matrix $\widetilde{\boldsymbol{\Sigma}}_{sample}$ 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, as summarized in Table 1. We demonstrate experimentally that this improves the portfolio weight $\boldsymbol{w}$ obtained from (1.1) and (1.2).

Our study is closely related to factor analysis and principal component analysis, which were previously applied to explain the interdependency of stock returns and classify the securities into appropriate subgroups. Sharpe [43] first proposed a single-factor model in this context using market returns. King [26] analyzed stock behaviors with both multiple factors and multiple principal components. These factor models established a basis for the asset pricing models CAPM [32, 35, 44, 47] and APT [39, 40].

There have been previous efforts to improve the estimate of $\boldsymbol{\Sigma}$. Sharpe [43] proposed a market-index covariance matrix $\widetilde{\boldsymbol{\Sigma}}_{market}$ derived from a single-factor model of market returns. Ledoit and Wolf [30] introduced a shrinkage method that averages $\widetilde{\boldsymbol{\Sigma}}_{sample}$ and $\widetilde{\boldsymbol{\Sigma}}_{market}$. They in [31] also applied the shrinkage method with a different target, an identity matrix. Later, it was shown by DeMiguel et al. [10] that their shrinkage methods have the same effect as adding the constraint $||\boldsymbol{w}||_{\boldsymbol{A}} \leq \delta$ to the GMV problem (1.2), where $\boldsymbol{A}$ is the shrinkage target matrix ($\widetilde{\boldsymbol{\Sigma}}_{market}$ or $\boldsymbol{I}_N$) and δ is a given threshold. Elton and Gruber [13] estimated $\boldsymbol{\Sigma}$ using a few principal components from a correlation matrix. More recently, Plerou et al. [38], Laloux et al. [28], Conlon, Ruskin, and Crane [8], and Kwapień, Drożdż, and Oświęcimka [27] applied random matrix theory [34] 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 and Wolf [30] by damping all but the k largest eigenvalues by a single rate. In summary, the estimator of Sharpe [43] uses $\widetilde{\boldsymbol{\Sigma}}_{market}$, the estimator of Ledoit and Wolf [30, 31] takes the weighted average of $\widetilde{\boldsymbol{\Sigma}}_{sample}$ and different target matrices, the estimator of Elton and Gruber [13] truncates the smallest eigenvalues, the estimators of Plerou et al. [38], Laloux et al. [28], Conlon, Ruskin, and Crane [8], and Kwapień, Drożdż, and Oświęcimka [27] adjust principal components in some interval, and the estimator of Bengtsson and Holst [2] attenuates the smallest eigenvalues by a single rate.

Jagannathan and Ma [25] showed that a short-sale constraint ($w \geq 0$) is equivalent to shrinking the input covariance matrix $\boldsymbol{\Sigma}$ by subtracting $(\boldsymbol{\lambda}\mathbf{1}^T + \mathbf{1}^T\boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers for the constraints. DeMiguel et al. [10] showed that adding the short-sale constraint to GMV is equivalent to adding a 1-norm constraint $||\boldsymbol{w}||_1 \leq 1$, and they generalized this constraint to $||\boldsymbol{w}||_1 \leq \delta$ for a certain threshold δ which determines a short-sale budget.

Our study focuses on estimating a good covariance matrix. We propose decreasing the contribution of the smaller eigenvalues of a correlation matrix gradually by using a *Tikhonov filtering function*. To derive the Tikhonov filtering, we construct a linear model based on principal component analysis and formulate an optimization problem that finds appropriately noise-filtered factors. Using the filtered factor data, we estimate a Tikhonov covariance matrix.

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. 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 factor-based 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 comparing the covariance estimators for portfolio construction using monthly return data of 100 randomly chosen stocks from the Center for Research in Security Prices. In section 7, we highlight the differences between Tikhonov filtering and the other methods.

2. Tikhonov filtering.

To estimate the covariance matrix, we apply 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 downweighting, 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

$$||\boldsymbol{a}||^2 = \boldsymbol{a}^T\boldsymbol{a} \quad \text{and} \quad ||\boldsymbol{A}||_F^2 = \left(\sum_{i=1}^{M}\sum_{j=1}^{N} a_{ij}^2\right) \tag{2.1}$$

for a given vector $\boldsymbol{a}$ and a given matrix $\boldsymbol{A} = (a_{ij}) \in \mathbb{R}^{M\times N}$.

For a random process $\boldsymbol{x}(t)$, let $\mathbb{E}[\boldsymbol{x}(t)] \in \mathbb{R}^{N\times 1}$, $\mathrm{Var}[\boldsymbol{x}(t)] \in \mathbb{R}^{N\times 1}$, $\mathrm{Cov}[\boldsymbol{x}(t)] \in \mathbb{R}^{N\times N}$, and $\mathrm{Corr}[\boldsymbol{x}(t)] \in \mathbb{R}^{N\times N}$ denote a mean, a variance, a covariance matrix, and a correlation matrix. For a given collection of observations $\boldsymbol{X} = [\boldsymbol{x}(1),\dots,\boldsymbol{x}(T)]$ for N objects during T times, let $\mathbb{E}_s[\boldsymbol{x}(t)] \in \mathbb{R}^{N\times 1}$, $\mathrm{Var}_s[\boldsymbol{x}(t)] \in \mathbb{R}^{N\times 1}$, $\mathrm{Cov}_s[\boldsymbol{x}(t)] \in \mathbb{R}^{N\times N}$, and $\mathrm{Corr}_s[\boldsymbol{x}(t)] \in \mathbb{R}^{N\times N}$ denote the corresponding sample statistics, defined, for example, in [21, section 3.3]. An $N\times N$ identity matrix will be denoted by $\boldsymbol{I}_N$.

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

$$\mathbb{E}_s[\boldsymbol{z}(t)] = \boldsymbol{0}, \quad \mathrm{Var}_s[\boldsymbol{z}(t)] = \boldsymbol{1}, \tag{2.2}$$

[1]In this paper, the term PCA always refers to applying PCA to the matrix $\boldsymbol{R}$ of sample stock returns. For convergence properties of the sample PCA toward its population PCA, refer to [24, Chapter 4].

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

$$\boldsymbol{Z} = \boldsymbol{D}_V^{-\frac{1}{2}} \left(\boldsymbol{R} - \frac{1}{T} \boldsymbol{R} \mathbf{1} \mathbf{1}^T \right), \tag{2.3}$$

where $\boldsymbol{D}_V = \mathrm{diag}(\mathrm{Var}_s[\boldsymbol{r}(t)]) \in \mathbb{R}^{N \times N}$ is a diagonal matrix containing the N sample variances for the N stock returns. 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 [24, pp. 64–66].

PCA finds an orthogonal basis $\boldsymbol{U} = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_k] \in \mathbb{R}^{N \times k}$ 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 $\boldsymbol{Z} = [\boldsymbol{z}(1), \dots, \boldsymbol{z}(T)]$ as

$$\boldsymbol{Z} = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_k] \; \boldsymbol{F} = \boldsymbol{U}\boldsymbol{F}, \tag{2.4}$$

$$\boldsymbol{z}(t) = \boldsymbol{U} \boldsymbol{f}(t) = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_k] \; \boldsymbol{f}(t) = \sum_{i=1}^{k} f_i(t) \boldsymbol{u}_i, \tag{2.5}$$

where $\boldsymbol{f}(t) = [f_1(t), \dots, f_k(t)]^T$, a column of $\boldsymbol{F}$, is the projected data at time t, and $\mathrm{Var}_s[f_1(t)] \geq \mathrm{Var}_s[f_2(t)] \geq \cdots \geq \mathrm{Var}_s[f_k(t)]$. The projected data $f_i(t)$ is called the ith principal component in PCA or the ith factor in the factor analysis. Larger $\mathrm{Var}_s[f_i(t)]$ implies that the corresponding $\boldsymbol{f}_i(t)$ plays a more important role in representing $\boldsymbol{Z}$. The orthogonal basis $\boldsymbol{U}$ and the projected data $\boldsymbol{F}$ 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,

$\boldsymbol{U}_k = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_k] \in \mathbb{R}^{N \times k}$ is a matrix of left singular vectors,

$\boldsymbol{S}_k = \mathrm{diag}(s_1, \dots, s_k) \in \mathbb{R}^{k \times k}$ is a diagonal matrix of singular values s_i,

and $\boldsymbol{V}_k = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_k] \in \mathbb{R}^{T \times k}$ is a matrix of right singular vectors.

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

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

In PCA, the orthogonal basis matrix $\boldsymbol{U}$ corresponds to $\boldsymbol{U}_k$, and the projected data $\boldsymbol{F}$ corresponds to $(\boldsymbol{S}_k \boldsymbol{V}_k^T)$ [24, p. 193]. Moreover, the variance of the projected data $f_i(t)$ is proportional to the square of singular value s_i^2 as we now show. $\mathbb{E}_s[\boldsymbol{z}(t)] = \mathbf{0}$ means that $\boldsymbol{Z}\mathbf{1} = \mathbf{0}$. Therefore, since $\boldsymbol{z}(t) = \boldsymbol{U}\boldsymbol{f}(t)$,

$$\mathbb{E}_s[\boldsymbol{f}(t)] = \boldsymbol{U}^T \boldsymbol{Z}\mathbf{1} = \mathbf{0}, \tag{2.8}$$

so $\boldsymbol{f}(t)$ also has zero-mean. Therefore,

$$\mathrm{Var}_s[f_i(t)] = \frac{1}{T} \sum_{t=1}^{T} (f_i(t) - \mathbb{E}_s[f_i(t)])^2 = \frac{1}{T} \sum_{t=1}^{T} f_i^2(t).$$

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

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

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

$$\mathrm{Var}_s[f_i(t)] = \frac{1}{T}\sum_{t=1}^{T}(s_i v_i(t))^2 = \frac{1}{T}s_i^2(\boldsymbol{v}_i^T \boldsymbol{v}_i) = \frac{s_i^2}{T} \tag{2.10}$$

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

2.2. Tikhonov regularization.

$\boldsymbol{U}$ and $\boldsymbol{f}(t)$ 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 $f_i(t)$ 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 [36, 46, 49], sometimes called ridge regression [22, 23], to reduce the contribution of unimportant principal components to the normalized stock return $\boldsymbol{Z}$. Eventually, we construct a filtered principal component $\widetilde{\boldsymbol{f}}(t)$ 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 $\boldsymbol{b} \approx \boldsymbol{Ax}$, where the $M \times N$ matrix $\boldsymbol{A}$ has some singular values close to 0 [18, pp. 71–86]. If we write the SVD of $\boldsymbol{A}$ as

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

then the minimum norm least square solution $\boldsymbol{x}_{LS}$ to $\boldsymbol{b} \approx \boldsymbol{Af}$ 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 (TSVD) method [14, 20] and a Tikhonov method [46]. 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

$$\min_{\boldsymbol{f}} ||\boldsymbol{b} - \boldsymbol{Ax}||^2 + \alpha^2||\boldsymbol{Px}||^2, \tag{2.12}$$

where α and $\boldsymbol{P}$ are predetermined. The penalty term $||\boldsymbol{Px}||^2$ 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 $\boldsymbol{f}(t)$. We formulate the linear problem to find a filtered principal component $\widetilde{\boldsymbol{f}}(t)$ as

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

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

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

$$\boldsymbol{f}(t) = \begin{bmatrix} f_1(t) \\ \vdots \\ f_k(t) \end{bmatrix} = \begin{bmatrix} s_1\ v_1(t) \\ \vdots \\ s_k v_k(t) \end{bmatrix} = \sum_{i=1}^{k} (s_i v_i(t))\boldsymbol{e}_i,$$

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

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

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

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

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

We introduce two different filtering matrices, $\boldsymbol{\Phi}_{trun}(\widehat{k})$ and $\boldsymbol{\Phi}_{tikh}(\alpha)$, which correspond to TSVD and Tikhonov regularization. First, we can simply truncate all but $\widehat{k}$ most important components as Elton and Gruber [13] did by using a filtering matrix of $\boldsymbol{\Phi}_{trun}(\widehat{k}) = \mathrm{diag}(\underbrace{1, \ldots, 1}_{\widehat{k}}, \underbrace{0, \ldots, 0}_{k-\widehat{k}})$, so the truncated principal component $\widetilde{\boldsymbol{f}}_{trun}(t)$ is

$$\widetilde{\boldsymbol{f}}_{trun}(t) = \boldsymbol{\Phi}_{trun}(\widehat{k})\boldsymbol{f}(t).$$

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

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

From (2.17), we can see that this truncation method corresponds to the TSVD [14, 20].

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

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

with

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

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

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

Let $\widetilde{f}_i(t)$ denote the ith element of $\widetilde{\boldsymbol{f}}(t)$. The matrix $\boldsymbol{P}$ scales each $\widetilde{f}_i(t)$ 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 $f_i(t)$ are more contaminated by the noise. Thus, the penalty term prevents $\widetilde{\boldsymbol{f}}(t)$ from containing large amounts of unimportant principal components. As we showed before, s_i^2 is proportional to the variance of the ith principal component $f_i(t)$. Therefore, this penalty matrix $\boldsymbol{P}$ is statistically meaningful considering that the values of $\widetilde{f}_i(t)/s_i$ in $\boldsymbol{P}\widetilde{\boldsymbol{f}}(t)$ are in proportion to the normalized principal components $\widetilde{f}_i(t)/\sqrt{\mathrm{Var}_s[f_i(t)]}$.

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

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

and thus

$$(\boldsymbol{U}^T\boldsymbol{U} + \alpha^2\boldsymbol{P}^T\boldsymbol{P})\widetilde{\boldsymbol{f}}(t) = \boldsymbol{U}^T\boldsymbol{z}(t).$$

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

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

Therefore,

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

and

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

So, our Tikhonov estimate is

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

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 $\widetilde{\boldsymbol{f}}_{tikh}(t)$ is the result after filtering the original principal component $\boldsymbol{f}(t)$ with the diagonal matrix $\boldsymbol{\Phi}_{tikh}(\alpha)$, whose diagonal elements $\phi_i^{tikh}(\alpha) = \frac{s_i^2}{s_i^2+\alpha^2}$ lie in $[0, 1]$. By (2.15) and (2.16), the resulting filtered data become

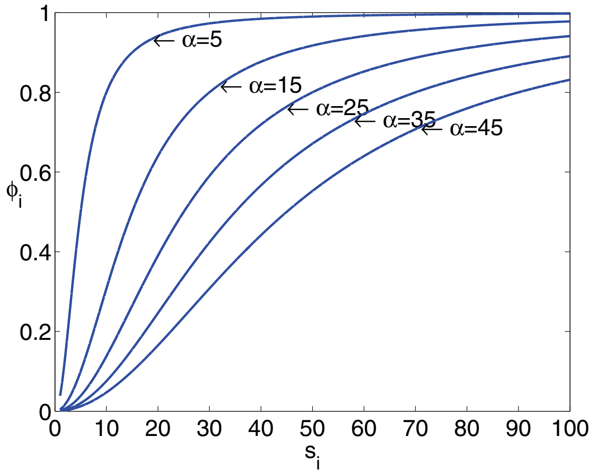


Figure 1. *Tikhonov filtering as a function of s_i for various values of α.*

$\widetilde{\boldsymbol{z}}_{tikh}(t) = \boldsymbol{U}\boldsymbol{\Phi}_{tikh}(\alpha)\,\boldsymbol{f}(t)$ and $\widetilde{\boldsymbol{Z}}_{tikh} = \boldsymbol{U}\,\boldsymbol{\Phi}_{tikh}(\alpha)\,\boldsymbol{F}$. 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 Figure 1. This is reasonable since α balances the error term and the penalty term. Later, in section 4, we will propose how we can determine an appropriate parameter α. 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.

2.3. The relation between filtered PCA and a factor model. Some asset pricing models (see, e.g., [40, 44]) model asset returns with a factor model:

$$\boldsymbol{r}(t) = \mathbb{E}[\boldsymbol{r}(t)] + \mathcal{B}\boldsymbol{\varphi}(t) + \boldsymbol{\epsilon}(t). \tag{2.19}$$

The assumptions are that

$$\mathbb{E}[\boldsymbol{\varphi}(t)] = \mathbb{E}[\boldsymbol{\epsilon}(t)] = \mathbf{0}, \tag{2.20}$$

$$\mathbb{E}(\epsilon_i(t)\epsilon_j(t)) = \mathbb{E}(\epsilon_i(t)\varphi_\ell(t)) = \mathbb{E}(\varphi_i(t)\varphi_j(t)) = 0 \quad \text{for all } i \neq j, \tag{2.21}$$

where $\boldsymbol{\varphi}(t) = [\varphi_1(t), \ldots, \varphi_\ell(t)]^T$ and $\boldsymbol{\epsilon}(t) = [\epsilon_1(t), \ldots, \epsilon_N(t)]^T$. The common factors $\varphi_i(t)$ are referred to as systematic factors, and $\epsilon_i(t)$ is called an unsystematic (idiosyncratic) factor. The matrix $\mathcal{B} = (\beta_{ik})$ is called a factor-loading matrix, and β_{ik} represents the sensitivity of the ith asset to the kth factor.

We can interpret our linear model (2.14) as a factor model. By (2.3) and (2.14), we have a linear equation for $\boldsymbol{r}(t)$ as

$$\boldsymbol{r}(t) = \mathbb{E}_s[\boldsymbol{r}(t)] + \boldsymbol{D}_V^{\frac{1}{2}}\left(\boldsymbol{U}\widetilde{\boldsymbol{f}}(t) + \boldsymbol{\epsilon}_z(t)\right) \tag{2.22}$$

$$= \mathbb{E}_s[\boldsymbol{r}(t)] + \boldsymbol{B}\widetilde{\boldsymbol{f}}(t) + \boldsymbol{\epsilon}_r(t), \tag{2.23}$$

where

$$\boldsymbol{B} = \boldsymbol{D}_V^{\frac{1}{2}}\boldsymbol{U} \quad \text{and} \quad \boldsymbol{\epsilon}_r(t) = \boldsymbol{D}_V^{\frac{1}{2}}\boldsymbol{\epsilon}_z(t). \tag{2.24}$$

Comparing (2.19) and (2.23), if we assume that $\widetilde{\boldsymbol{f}}(t)$ represents the systematic factors $\boldsymbol{\varphi}(t)$ well, we can interpret $\boldsymbol{B}$ and $\boldsymbol{\epsilon}_r(t)$ as estimates of the loading matrix $\mathcal{B}$ and the unsystematic factor $\boldsymbol{\epsilon}(t)$ in (2.19). Since $\boldsymbol{\epsilon}_z(t) = \boldsymbol{z}(t) - \boldsymbol{U}\widetilde{\boldsymbol{f}}(t)$, $\boldsymbol{\epsilon}_r(t)$ becomes

$$\boldsymbol{\epsilon}_r(t) = \boldsymbol{D}_V^{\frac{1}{2}}\boldsymbol{\epsilon}_z(t) = \boldsymbol{D}_V^{\frac{1}{2}}(\boldsymbol{z}(t) - \boldsymbol{U}\widetilde{\boldsymbol{f}}(t)). \tag{2.25}$$

Because $\boldsymbol{z}(t) = \boldsymbol{U}\boldsymbol{f}(t)$ and $\widetilde{\boldsymbol{f}}(t) = \boldsymbol{\Phi}\boldsymbol{f}(t)$, the factor models result in the estimate

$$\boldsymbol{\epsilon}_r(t) = \boldsymbol{D}_V^{\frac{1}{2}}(\boldsymbol{U}\boldsymbol{f}(t) - \boldsymbol{U}\boldsymbol{\Phi}\boldsymbol{f}(t)) = \left(\boldsymbol{D}_V^{\frac{1}{2}}\boldsymbol{U}\right)(\boldsymbol{I}_k - \boldsymbol{\Phi})\boldsymbol{f}(t) = \boldsymbol{B}(\boldsymbol{I}_k - \boldsymbol{\Phi})\boldsymbol{f}(t). \tag{2.26}$$

3. Estimate of the covariance matrix $\boldsymbol{\Sigma}$.

In this section we study how filtering changes the covariance and correlation estimates and the estimate of risk exposure, and how to ensure that the estimated covariance matrix has full rank.

3.1. A covariance estimate.

Now we derive a covariance matrix estimate $\widetilde{\boldsymbol{\Sigma}}$ from (2.23), respecting the structure of the factor model (2.19). By (2.21), the covariance matrix $\boldsymbol{\Sigma}$ is

$$\boldsymbol{\Sigma} = \mathcal{B}\mathrm{Cov}[\boldsymbol{\varphi}(t)]\mathcal{B}^T + \mathrm{Cov}[\boldsymbol{\epsilon}(t)] = \boldsymbol{\Sigma}_s + \boldsymbol{D}_\epsilon, \tag{3.1}$$

where $\boldsymbol{\Sigma}_s$ denotes the systematic component $\mathcal{B}\mathrm{Cov}[\boldsymbol{\varphi}(t)]\mathcal{B}^T$ and $\boldsymbol{D}_\epsilon$ denotes the unsystematic component $\mathrm{Cov}[\boldsymbol{\epsilon}(t)]$. We estimate the systematic part $\boldsymbol{\Sigma}_s$ by $\widetilde{\boldsymbol{\Sigma}}_s = \boldsymbol{B}\mathrm{Cov}_s[\widetilde{\boldsymbol{f}}(t)]\boldsymbol{B}^T$. Because $\boldsymbol{f}(t)$ has zero-mean, $\widetilde{\boldsymbol{f}}(t) = \boldsymbol{\Phi}\boldsymbol{f}(t)$ also has zero-mean, so

$$\mathrm{Cov}_s[\widetilde{\boldsymbol{f}}(t)] = \frac{1}{T}(\boldsymbol{\Phi}\boldsymbol{F})(\boldsymbol{\Phi}\boldsymbol{F})^T = \frac{1}{T}(\boldsymbol{\Phi}^2\boldsymbol{S}_k^2). \tag{3.2}$$

Therefore, the estimate of $\boldsymbol{\Sigma}_s$ becomes

$$\widetilde{\boldsymbol{\Sigma}}_s = \boldsymbol{B}\mathrm{Cov}_s[\widetilde{\boldsymbol{f}}(t)]\boldsymbol{B}^T = \frac{1}{T}\boldsymbol{B}(\boldsymbol{\Phi}^2\boldsymbol{S}_k^2)\boldsymbol{B}^T. \tag{3.3}$$

The unsystematic part $\boldsymbol{D}_\epsilon$ in (3.1) is diagonal since the unsystematic factors $\epsilon_i(t)$ are mutually uncorrelated. Thus, we estimate $\mathrm{Cov}[\boldsymbol{\epsilon}(t)]$ by the diagonal part of the difference $\widetilde{\boldsymbol{D}}_\epsilon$ between

$$\widetilde{\boldsymbol{\Sigma}}_{sample} = \mathrm{Cov}_s[\boldsymbol{r}(t)] = \frac{1}{T}\boldsymbol{B}\boldsymbol{S}_k^2\boldsymbol{B}^T \tag{3.4}$$

and $\widetilde{\boldsymbol{\Sigma}}_s$. Hence,

$$\widetilde{\boldsymbol{D}}_\epsilon = \mathrm{diag}\left(\widetilde{\boldsymbol{\Sigma}}_{sample} - \widetilde{\boldsymbol{\Sigma}}_s\right) = \mathrm{diag}\left(\frac{1}{T}(\boldsymbol{B}(\boldsymbol{I}_k - \boldsymbol{\Phi}^2)\boldsymbol{S}_k^2\boldsymbol{B}^T)\right). \tag{3.5}$$

Finally, the filtered covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ will be

$$\widetilde{\boldsymbol{\Sigma}} = \widetilde{\boldsymbol{\Sigma}}_s + \widetilde{\boldsymbol{D}}_\epsilon, \tag{3.6}$$

where $\widetilde{\boldsymbol{\Sigma}}_s$ and $\widetilde{\boldsymbol{D}}_\epsilon$ are defined by (3.3) and (3.5). By the definition of $\widetilde{\boldsymbol{D}}_\epsilon$, the diagonal of $\widetilde{\boldsymbol{\Sigma}}$ equals $\mathrm{Var}_s[\boldsymbol{r}(t)]$.

Now we analyze how the filtering function $\boldsymbol{\Phi}$ affects the sample correlation matrix $\mathrm{Corr}_s[\boldsymbol{r}(t)]$. By (3.6), the filtered correlation matrix $\widetilde{\boldsymbol{\Omega}}$ can be calculated as

$$\widetilde{\boldsymbol{\Omega}} = \boldsymbol{D}_V^{-\frac{1}{2}} \widetilde{\boldsymbol{\Sigma}} \boldsymbol{D}_V^{-\frac{1}{2}} = \frac{1}{T} \boldsymbol{U} \boldsymbol{\Phi}^2 \boldsymbol{S}_k^2 \boldsymbol{U}^T + \boldsymbol{D}_V^{-\frac{1}{2}} \tilde{\boldsymbol{D}}_\epsilon \boldsymbol{D}_V^{-\frac{1}{2}}, \tag{3.7}$$

where the second term makes the diagonal elements of $\widetilde{\boldsymbol{\Omega}}$ equal one. On the other hand, the sample correlation matrix $\mathrm{Corr}_s[\boldsymbol{r}(t)]$ can be calculated as

$$\mathrm{Corr}_s[\boldsymbol{r}(t)] = \boldsymbol{D}_V^{-\frac{1}{2}} \widetilde{\boldsymbol{\Sigma}}_{sample} \boldsymbol{D}_V^{-\frac{1}{2}}.$$

By (2.24) and (3.4), this becomes

$$\mathrm{Corr}_s[\boldsymbol{r}(t)] = \boldsymbol{D}_V^{-\frac{1}{2}} \left(\frac{1}{T} \boldsymbol{B} \boldsymbol{S}_k^2 \boldsymbol{B}^T \right) \boldsymbol{D}_V^{-\frac{1}{2}} = \frac{1}{T} \boldsymbol{U} \boldsymbol{S}_k^2 \boldsymbol{U}^T. \tag{3.8}$$

Comparing $\widetilde{\boldsymbol{\Omega}}$ in (3.7) and $\mathrm{Corr}_s[\boldsymbol{r}(t)]$ in (3.8), we can see that $\widetilde{\boldsymbol{\Omega}}$ is the result of applying the filtering matrix $\boldsymbol{\Phi}^2$ to $\boldsymbol{S}_k^2$ in $\mathrm{Corr}_s[\boldsymbol{r}(t)]$ and replacing the diagonal elements with one. Since each diagonal element of $\boldsymbol{S}_k^2$ corresponds to an eigenvalue of $\mathrm{Corr}_s[\boldsymbol{r}(t)]$, the filtering matrix $\boldsymbol{\Phi}^2$ attenuates the eigenvalues of $\mathrm{Corr}_s[\boldsymbol{r}(t)]$. In the previous section, we introduced two filtering matrices:

$$\boldsymbol{\Phi}_{trun}(\widehat{k}) = \mathrm{diag}(\underbrace{1, \dots, 1}_{\widehat{k}}, \underbrace{0, \dots, 0}_{k-\widehat{k}}) \tag{3.9}$$

and

$$\boldsymbol{\Phi}_{tikh}(\alpha) = \mathrm{diag}\left(\frac{s_1^2}{s_1^2 + \alpha^2}, \dots, \frac{s_k^2}{s_k^2 + \alpha^2} \right). \tag{3.10}$$

Therefore, $\boldsymbol{\Phi}_{trun}^2(\widehat{k})$ truncates the eigencomponents corresponding to the $(k - \widehat{k})$ smallest eigenvalues, and $\boldsymbol{\Phi}_{tikh}^2(\alpha)$ downweights all the eigenvalues at a rate $(\frac{s_i^2}{s_i^2+\alpha^2})^2 = (\frac{\lambda_i}{\lambda_i+\alpha^2})^2$, where λ_i is the ith largest eigenvalue of $\mathrm{Cov}_s[\boldsymbol{z}(t)]$. Hence, the TSVD filtering functions $\phi_{trun}^2(\lambda_i)$ for eigenvalues λ_i become

$$\phi_{trun}^2(\lambda_i) = \begin{cases} 1 & \text{if } i \le \widehat{k}, \\ 0 & \text{otherwise,} \end{cases}$$

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

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

We let $\widetilde{\boldsymbol{\Sigma}}_{trun}$ and $\widetilde{\boldsymbol{\Sigma}}_{tikh}$ denote the estimates resulting from applying $\boldsymbol{\Phi}_{trun}^2(\widehat{k})$ and $\boldsymbol{\Phi}_{tikh}^2(\alpha)$ to (3.6). Finally, we can summarize the process of estimating the covariance matrix as Table 1.

Table 1

The algorithm to compute the covariance estimate $\widetilde{\boldsymbol{\Sigma}}$. *For Tikhonov, the filter factors are* $\boldsymbol{\Phi}_{tikh} = \mathrm{diag}(\frac{s_1^2}{s_1^2+\alpha^2}, \ldots, \frac{s_k^2}{s_k^2+\alpha^2})$.

Step 1.	Estimate the systematic component of the covariance $\frac{1}{T}\boldsymbol{B}(\boldsymbol{\Phi}^2\boldsymbol{S}_k^2)\boldsymbol{B}^T$, where $\boldsymbol{\Phi}$ is the diagonal matrix of filter factors.
Step 2.	Change the main diagonal to be the sample variances.

3.2. Risk exposure to factors. By (3.1), the variance of a portfolio return can be expressed as

$$\boldsymbol{w}^T\boldsymbol{\Sigma}\boldsymbol{w} = \boldsymbol{w}^T\left(\boldsymbol{\Sigma}_s + \boldsymbol{D}_\epsilon\right)\boldsymbol{w} = \boldsymbol{w}^T\boldsymbol{\Sigma}_s\boldsymbol{w} + \boldsymbol{w}^T\boldsymbol{D}_\epsilon\boldsymbol{w}. \tag{3.11}$$

The systematic risk is

$$\boldsymbol{w}^T\boldsymbol{\Sigma}_s\boldsymbol{w} = \boldsymbol{w}^T\left(\mathcal{B}\mathrm{Cov}[\boldsymbol{\varphi}(t)]\mathcal{B}^T\right)\boldsymbol{w} = \boldsymbol{w}^T\left(\mathcal{B}\,\mathrm{diag}(\mathrm{Var}[\boldsymbol{\varphi}(t)])\mathcal{B}^T\right)\boldsymbol{w} \tag{3.12}$$

because $\varphi_i(t)$ are mutually uncorrelated by (2.21). This can be expanded as

$$\boldsymbol{w}^T\boldsymbol{\Sigma}_s\boldsymbol{w} = \sum_{i=1}^{k}\mathrm{Var}[\varphi_i(t)](\boldsymbol{w}^T\boldsymbol{\beta}_i)^2, \tag{3.13}$$

where $\boldsymbol{\beta}_i$ is the ith column of $\mathcal{B}$. The ith term in (3.13) represents the risk exposure of the portfolio to the ith factor.

On the other hand, the estimated matrix $\widetilde{\boldsymbol{\Sigma}}_s$ in (3.3) can be rewritten as

$$\widetilde{\boldsymbol{\Sigma}}_s = \frac{1}{T}\boldsymbol{B}\boldsymbol{\Phi}^2\boldsymbol{S}_k^2\boldsymbol{B}^T = \boldsymbol{B}\boldsymbol{\Phi}^2\left(\frac{\boldsymbol{S}_k^2}{T}\right)\boldsymbol{B}^T = \boldsymbol{B}\boldsymbol{\Phi}^2\,\mathrm{diag}(\mathrm{Var}_s[\boldsymbol{f}(t)])\;\boldsymbol{B}^T \tag{3.14}$$

because $\mathrm{Var}_s[\boldsymbol{f}(t)] = \mathrm{diag}(\boldsymbol{S}_k^2/T)$ by (2.10). Hence, we can calculate the estimated systematic risk as

$$\boldsymbol{w}^T\widetilde{\boldsymbol{\Sigma}}_s\boldsymbol{w} = \sum_{i=1}^{k}\phi_i^2\left(\mathrm{Var}_s[\boldsymbol{f}_i(t)](\boldsymbol{w}^T\boldsymbol{b}_i)^2\right), \tag{3.15}$$

where $\boldsymbol{b}_i$ is the ith column of $\boldsymbol{B}$. Therefore, we can see that our estimate of the risk exposure to the ith factor is reduced by ϕ_i^2. This equation explains how the estimated covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ affects the estimated risk measure of a portfolio, downweighting risk factors corresponding to small values of $\phi_i(\alpha)$.

3.3. Rank deficiency of the covariance matrix. Since the covariance matrix is positive semidefinite, the MV problem (1.1) and the GMV problem (1.2) always have 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 $\widetilde{\boldsymbol{\Sigma}}_{sample}$ from (1.3) has rank $(T-1)$ at most. Therefore, whenever the number of observations T is less than or equal to the number of stocks N, $\widetilde{\boldsymbol{\Sigma}}_{sample}$ is rank deficient. To ensure full rank and a 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.

Recall that the covariance matrix estimate $\widetilde{\boldsymbol{\Sigma}}$ is the sum of the systematic part $\widetilde{\boldsymbol{\Sigma}}_s$ and the unsystematic part $\widetilde{\boldsymbol{D}}_\epsilon$. By (3.3), we can see that $\widetilde{\boldsymbol{\Sigma}}_s$ has nonnegative eigenvalues. On the other hand, by (3.5),

$$\text{(The } i\text{th diagonal element of } \widetilde{\boldsymbol{D}}_\epsilon) = \boldsymbol{e}_i^T \left(\frac{1}{T} \boldsymbol{B}(\boldsymbol{I}_k - \boldsymbol{\Phi}^2)\boldsymbol{S}_k^2 \boldsymbol{B}^T \right) \boldsymbol{e}_i. \tag{3.16}$$

It is reasonable to assume that $\boldsymbol{e}_i^T\boldsymbol{B}$ is not zero for any i since it becomes zero only when the ith stock has zero variance of returns by (2.24). Thus, the diagonal matrix $\widetilde{\boldsymbol{D}}_\epsilon$ is positive definite whenever all $\phi_i < 1$. In the case of Tikhonov filtering, whenever $\alpha > 0$,

$$\phi_i^{tikh}(\alpha) = \frac{s_i^2}{s_i^2 + \alpha^2} < 1,$$

so $\widetilde{\boldsymbol{D}}_\epsilon$ is positive definite. Therefore, since $\widetilde{\boldsymbol{\Sigma}}_s$ is positive semidefinite, adding a positive definite matrix ensures that Tikhonov covariance matrix $\widetilde{\boldsymbol{\Sigma}}_{tikh}$ is positive definite and therefore full rank.

Sharpe [43], Ledoit and Wolf [30], Bengtsson and Holst [2], and Plerou et al. [38] also overcome the rank-deficiency problem by replacing the diagonals of their estimate with the sample variances as in Step 2 in Table 1. However, some of their filtering values ϕ_i could have a value of 1 as we will see in section 5. This implies that the resulting estimate $\widetilde{\boldsymbol{\Sigma}}$ could be rank-deficient or very ill conditioned even after adding $\widetilde{\boldsymbol{D}}_\epsilon$ because $\widetilde{\boldsymbol{D}}_\epsilon$ is positive semidefinite. In the case that the estimate still has a large condition number even after Step 2, we can fix the problem by a small modification as follows:

$$\widetilde{\boldsymbol{\Sigma}}_{ii} \leftarrow \widetilde{\boldsymbol{\Sigma}}_{ii} + \delta_i \quad \text{for } i = 1, \ldots, N, \tag{3.17}$$

where δ_i is a small positive number.

Theorem 3.1 (condition number modification). *Replacing the main diagonal of the covariance estimate $\widetilde{\boldsymbol{\Sigma}}$ as specified in (3.17) guarantees that*

$$\text{cond}(\widetilde{\boldsymbol{\Sigma}}) \leq \frac{\lambda_{\max}(\widetilde{\boldsymbol{\Sigma}}) + \max(\delta_i)}{\min(\delta_i)},$$

where $\lambda_{\max}(\cdot)$ is the maximum eigenvalue of the matrix.

Proof. This is a direct consequence of the eigenvalue interlacing theorem [45, p. 203] and the positive semidefiniteness of $\widetilde{\boldsymbol{\Sigma}}$. ■

This modification is useful especially for the sample covariance matrix $\widetilde{\boldsymbol{\Sigma}}_{sample}$ when $T \leq N$, and for the truncation-based estimators whose filtering factors ϕ_i equal 1 for some i.

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

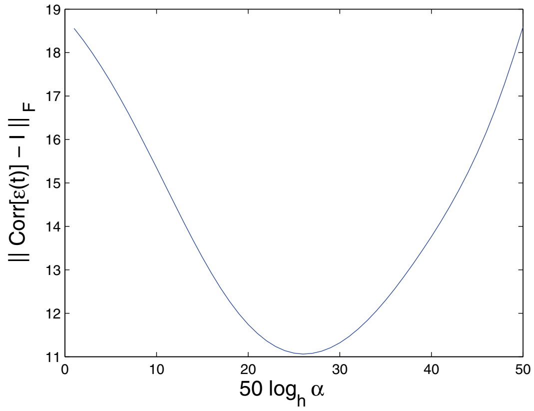


Figure 2. *The difference* $||Corr_s[\boldsymbol{\epsilon}_r(t)] - \mathbf{I}_N||_F$ *as a function of log-scaled* α *where* $h = \max(s_i)$.

parameter α. In regularization methods for discrete ill-posed problems, there are intensive studies about choosing α using methods such as generalized cross-validation [15], L-curves [17, 19], and residual periodograms [41, 42].

In factor analysis and PCA, there are analogous studies to determine the number of factors such as Bartlett's test [1], SCREE test [4], average root [16], partial correlation procedure [50], and cross-validation [51]. More recently, Plerou et al. [37, 38] 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 [48] studied the behavior of eigenvalues as the number of assets increases, and Connor and Korajczyk [9] studied the probabilistic behavior of noise factors.

The use of these methods requires various statistical properties for $\boldsymbol{\epsilon}_r(t)$ in the linear model (2.23). We note that since $\mathbb{E}_s[\boldsymbol{f}(t)] = \mathbf{0}$ by (2.8), the noise $\boldsymbol{\epsilon}_r(t)$ in (2.23) has zero-mean: By (2.26),

$$\mathbb{E}_s[\boldsymbol{\epsilon}_r(t)] = \boldsymbol{B}(\boldsymbol{I}_k - \boldsymbol{\Phi})\ \mathbb{E}_s[\boldsymbol{f}(t)] = \mathbf{0}. \tag{4.1}$$

For our Tikhonov estimation, we propose a new method adopting a mutually uncorrelated noise assumption in a factor model (2.21), so $\mathrm{Corr}_s[\boldsymbol{\epsilon}_r(t)] \simeq \boldsymbol{I}_N$. Hence, as a criterion to determine an appropriate parameter α, we formulate an optimization problem minimizing the correlations among the noise,

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

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 [50] to determine the number of principal components. Figure 2 illustrates an example of $||\mathrm{Corr}_s[\boldsymbol{\epsilon}_r(t)] - \boldsymbol{I}_N||_F$ as a function of α in the range $[s_k, s_1]$. The parameter might alternatively be determined by an asymptotic analysis proposed by Ledoit and Wolf [30, 31] or a cross-validation used by DeMiguel et al. [10].

5. Comparison to other estimators.

In this section, we compare other covariance estimators to our Tikhonov estimator and put them all in a common framework. We summarize how they filter the eigenvalues of the sample correlation matrix with filtering functions $\phi^2(\lambda_i)$. Most of these methods use a two-step procedure as shown in Table 1: filter the eigenvalues and then adjust the main diagonal. We note any exceptions in our descriptions.

5.1. $\widetilde{\boldsymbol{\Sigma}}_{sample}$: Sample covariance matrix.

A sample covariance matrix is the filtering target of most covariance estimators, including our Tikhonov estimator. Thus, the sample covariance matrix $\widetilde{\boldsymbol{\Sigma}}_{sample}$ can be thought of as an unfiltered covariance matrix, so the filtering function $\phi_s^2(\lambda_i)$ for eigenvalues of $\mathrm{Cov}_s[\boldsymbol{z}(t)]$ is

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

5.2. $\widetilde{\boldsymbol{\Sigma}}_{market}$ from the single market index model [43].

Sharpe [43] proposed a single index market model

$$\boldsymbol{r}(t) = \mathbb{E}[\boldsymbol{r}(t)] + \boldsymbol{b}\, r_m(t) + \boldsymbol{\epsilon}(t), \tag{5.1}$$

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,
$\boldsymbol{\epsilon}(t)$ is zero-mean uncorrelated error at time t,
and $\boldsymbol{b} \in \mathbb{R}^{N\times 1}$.

Unlike the factor model (2.19), this model assumes that the stock returns $\boldsymbol{r}(t)$ have only one common factor, the market return $r_m(t)$. Interestingly, Plerou et al. [38, p. 8] observed that the principal component corresponding to the largest eigenvalue of the correlation matrix $\mathrm{Corr}_s[\boldsymbol{r}(t)](= \mathrm{Cov}_s[\boldsymbol{z}(t)])$ is highly correlated with $r_m(t)$. 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 $f_1(t)$ in (2.5) represents the market return $r_m(t)$. Thus, we can represent the relation between $\tilde{\boldsymbol{f}}(t) = [\tilde{f}_1(t), \dots, \tilde{f}_k(t)]$ in (2.23) and $r_m(t)$ as

$$\tilde{f}_i(t) \simeq \begin{cases} C\, r_m(t) & \text{when } i = 1, \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

for some constant C. Hence, the corresponding filtering function $\phi_m^2(\lambda_i)$ for $\widetilde{\boldsymbol{\Sigma}}_{market}$ becomes

$$\phi_m^2(\lambda_i) \simeq \begin{cases} 1 & \text{if } i = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{5.3}$$

Therefore, the filter function implicitly truncates all but the largest eigencomponent of $\mathrm{Corr}_s[\boldsymbol{r}(t)]$.

5.3. $\widetilde{\boldsymbol{\Sigma}}_{s\to m}$: Shrinkage toward $\widetilde{\boldsymbol{\Sigma}}_{market}$ [30].

Ledoit and Wolf propose a shrinkage method from $\widetilde{\boldsymbol{\Sigma}}_{sample}$ to $\widetilde{\boldsymbol{\Sigma}}_{market}$ as

$$\widetilde{\boldsymbol{\Sigma}}_{s\to m} = \gamma\, \widetilde{\boldsymbol{\Sigma}}_{market} + (1-\gamma)\widetilde{\boldsymbol{\Sigma}}_{sample}, \tag{5.4}$$

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

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

Since the true covariance matrix $\boldsymbol{\Sigma}$ is unknown, they use an asymptotic variance to determine an optimal γ. (Refer to [30, sections 2.5–2.6] for a detailed description.) Considering that $\widetilde{\boldsymbol{\Sigma}}_{market}$ is the result of the implicit truncation method, we can think of this shrinkage method as implicitly downweighting all eigenvalues but the largest at a rate $(1-\gamma)$. Therefore, we can represent the filtering function $\phi^2_{s\to m}(\lambda_i)$ as

$$\phi^2_{s\to m}(\lambda_i) \simeq \begin{cases} 1 & \text{if } i = 1, \\ 1-\gamma, \text{ where } 0 \le \gamma \le 1 & \text{otherwise.} \end{cases} \tag{5.5}$$

5.4. Truncated covariance matrix $\widetilde{\boldsymbol{\Sigma}}_{trun}$ [13]. As mentioned in section 3.1, the truncated covariance matrix $\widetilde{\boldsymbol{\Sigma}}_{trun}$ has the filtering function $\phi^2_{trun}(\lambda_i)$ for the eigenvalues λ_i of $\mathrm{Cov}_s[\boldsymbol{z}(t)]$, where

$$\phi^2_{trun}(\lambda_i) = \begin{cases} 1 & \text{if } i = 1, \dots, \widehat{k}, \\ 0 & \text{otherwise.} \end{cases} \tag{5.6}$$

Thus, the model of Elton and Gruber [13] truncates all but the $\widehat{k}$ largest eigencomponents of $\mathrm{Cov}_s[\boldsymbol{z}(t)]$.

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

$$\widetilde{\boldsymbol{\Sigma}}_{s\to trun} = \gamma\, \widetilde{\boldsymbol{\Sigma}}_{trun} + (1-\gamma)\widetilde{\boldsymbol{\Sigma}}_{sample}, \tag{5.7}$$

where $0 \le \gamma \le 1$. They determine the parameter γ in a way similar to [30]. (Refer to [2, sections 4.1–4.2] for a detailed description.) Therefore, $\widetilde{\boldsymbol{\Sigma}}_{s\to trun}$ is a variant of the shrinkage method toward $\widetilde{\boldsymbol{\Sigma}}_{trun}$. Because $\widetilde{\boldsymbol{\Sigma}}_{trun}$ is the truncated covariance matrix containing the $\widehat{k}$ most significant eigencomponents of $\mathrm{Cov}_s[\boldsymbol{z}(t)]$, we can regard $\widetilde{\boldsymbol{\Sigma}}_{s\to trun}$ as damping the smallest eigenvalues by $(1-\gamma)$. Thus, the filtering function corresponding to this approach is

$$\phi^2_{s\to trun}(\lambda_i) = \begin{cases} 1 & \text{if } i = 1, \dots, \widehat{k}, \\ 1-\gamma, \text{ where } 0 \le \gamma \le 1 & \text{otherwise.} \end{cases} \tag{5.8}$$

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. Bengtsson and Holst conclude that their shrinkage matrix $\widetilde{\boldsymbol{\Sigma}}_{s\to trun}$ performed best in the Swedish stock market when the shrinkage target $\widetilde{\boldsymbol{\Sigma}}_{trun}$ took only the most significant principal component ($\widehat{k} = 1$). They also mention that the result is consistent with random matrix theory because only the largest eigenvalue deviates far from the range of $[\lambda_{\min}, \lambda_{\max}]$, explained in the next section.

5.6. $\widetilde{\boldsymbol{\Sigma}}_{RMT:trun}$ truncation by random matrix theory [38]. Plerou et al. apply random matrix theory (RMT) [34], 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,

$$\mathrm{Corr}_{random} = \frac{1}{T}\boldsymbol{A}\boldsymbol{A}^T, \tag{5.9}$$

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

$$f(\lambda) = \begin{cases} \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{\max} - \lambda)(\lambda_{\min} - \lambda)}}{\lambda} & \lambda_{\min} \leq \lambda \leq \lambda_{\max}, \\ 0 & \text{otherwise,} \end{cases} \tag{5.10}$$

where σ^2 is the variance of the elements of $\boldsymbol{A}$, $\lambda_{\min} \leq \lambda \leq \lambda_{\max}$, and $\lambda_{\min}^{\max} = \sigma^2(1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}})$. By comparing the eigenvalue distribution of $\mathrm{Corr}_s[\boldsymbol{r}(t)]$ 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 eigencomponents should be removed from $\mathrm{Corr}_s[\boldsymbol{r}(t)]$. Thus, the filtering function $\phi^2_{RMT:trun}(\lambda_i)$ for the eigenvalue λ_i of $\mathrm{Corr}_s[\boldsymbol{r}(t)]$ is

$$\phi^2_{RMT:trun}(\lambda_i) = \begin{cases} 1 & \text{if } \lambda_i \geq \lambda_{\max}, \\ 0 & \text{otherwise.} \end{cases} \tag{5.11}$$

5.7. $\widetilde{\boldsymbol{\Sigma}}_{RMT:repl}$ replacing the RMT eigenvalues [28]. 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 (5.10) to the eigenvalue distribution of the observed correlation matrix rather than assuming that $\sigma^2 = 1$. 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 $\phi^2_{RMT:repl}(\lambda_i)$ for eigenvalues is

$$\phi^2_{RMT:repl}(\lambda_i) = \begin{cases} 1 & \text{if } \lambda_i \geq \lambda_{\max}, \\ \frac{C}{\lambda_i} & \text{otherwise.} \end{cases} \tag{5.12}$$

This approach does not require the application of Step 2 in Table 1 since it replaces the smallest eigenvalues with a positive constant. The resulting covariance matrix does not preserve the original variances.

5.8. $\widetilde{\boldsymbol{\Sigma}}_{s\to I}$: Shrinkage toward I [31]. Ledoit and Wolf also introduced a shrinkage method from $\widetilde{\boldsymbol{\Sigma}}_{sample}$ to the identity matrix $\boldsymbol{I}_N$ as

$$\widetilde{\boldsymbol{\Sigma}}_{s\to I} = \gamma\ (m\boldsymbol{I}_N) + (1-\gamma)\widetilde{\boldsymbol{\Sigma}}_{sample}, \tag{5.13}$$

where $m = \frac{\mathrm{trace}(\widetilde{\boldsymbol{\Sigma}}_{sample})}{N}$ and $0 \leq \gamma \leq 1$. They provide a method to estimate an optimal γ. (Refer to [31, section 3] for a detailed description.) There is no simple expression for the filter factors. In addition, this method does not use Step 2 in Table 1 since its shrinkage target $\boldsymbol{I}_N$ has full rank.

Table 2

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

Estimator	Filtering function $\phi^2(\lambda_i)$
$\widetilde{\mathbf{\Sigma}}_{sample}$	$\phi_s^2(\lambda_i) = 1$
$\widetilde{\mathbf{\Sigma}}_{market}$[43]	$\phi_m^2(\lambda_i) \simeq \begin{cases} 1 & \text{if } i = 1, \\ 0 & \text{otherwise.} \end{cases}$
$\widetilde{\mathbf{\Sigma}}_{s\to m}$[30]	$\phi_{s\to m}^2(\lambda_i) \simeq \begin{cases} 1 & \text{if } i = 1, \\ 1-\gamma & \text{otherwise.} \end{cases}$
$\widetilde{\mathbf{\Sigma}}_{trun}$[13]	$\phi_{trun}^2(\lambda_i) = \begin{cases} 1 & \text{if } i = 1, \ldots, \widehat{k}, \\ 0 & \text{otherwise.} \end{cases}$
$\widetilde{\mathbf{\Sigma}}_{s\to trun}$[2]	$\phi_{s\to trun}^2(\lambda_i) = \begin{cases} 1 & \text{if } i = 1, \ldots, \widehat{k}, \\ 1-\gamma & \text{otherwise.} \end{cases}$
$\widetilde{\mathbf{\Sigma}}_{RMT:trun}$[38]	$\phi_{RMT:trun}^2(\lambda_i) = \begin{cases} 1 & \text{if } \lambda_i \geq \lambda_{\max}, \\ 0 & \text{otherwise.} \end{cases}$
$\widetilde{\mathbf{\Sigma}}_{RMT:repl}$[28]	$\phi_{RMT:repl}^2(\lambda_i) = \begin{cases} 1 & \text{if } \lambda_i \geq \lambda_{\max}, \\ \frac{C}{\lambda_i} & \text{otherwise.} \end{cases}$
$\widetilde{\mathbf{\Sigma}}_{tikh}$	$\phi_{tikh}^2(\lambda_i) = \left(\frac{\lambda_i}{\lambda_i + \alpha^2}\right)^2$

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

$$\phi_{tikh}^2(\lambda_i) = \left(\frac{\lambda_i}{\lambda_i + \alpha^2}\right)^2, \tag{5.14}$$

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

5.10. Comparison. The derivations in section 5 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 2.*

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

In addition, all the estimators except $\widetilde{\mathbf{\Sigma}}_{s\to I}$ and $\widetilde{\mathbf{\Sigma}}_{RMT:repl}$ overcome the rank deficiency of the covariance matrix by replacing the diagonal elements with the corresponding variances after filtering. This is what we did by preserving $\widetilde{\boldsymbol{D}}_\epsilon$ in Step 2 in Table 1. However, most estimators have $\phi^2(\lambda_i) = 1$ for the largest eigenvalues as Table 2 shows, so the resulting covariance matrix can be still rank deficient as we discussed in section 3.3. During experiments in section 6, we actually observed the rank deficiency for some estimators even after preserving diagonal parts. This implies that an extra modification like (3.17) is necessary to overcome rank deficiency.

6. Experiments.

In this section, we evaluate the covariance estimators using return data from the NYSE, AMEX, and NASDAQ. We collected the monthly data from January 1958 to December 2007 from the CRSP (Center for Research in Security Prices) database. There are 600 months over 50 years, and we randomly chose 100 stocks among those traded throughout this period.

Chopra and Ziemba [7] have noted that the MV problem is much more sensitive to errors in $\boldsymbol{\mu}$ than to errors in $\boldsymbol{\Sigma}$, and our experience confirms this observation. In fact, uncertainty in the estimates of $\boldsymbol{\mu}$ made the true return quite different from the target return. In addition, recently DeMiguel, Garlappi, and Uppal [11] showed that some common portfolio strategies do not yield consistently better Sharpe ratios, certainty-equivalent returns, or turnovers, compared to a naive $1/N$ portfolio. The instability of the MV portfolio tends to increase turnover costs, so recent studies strengthen the stability by formulating new optimization problems [12]. However, since our study focuses on estimating the covariance matrix $\boldsymbol{\Sigma}$, we evaluated the estimators based on how well they minimize the risk variances in the MV and GMV portfolios.

First, in section 6.1, we evaluate the risk of the GMV portfolio using the covariance estimators of Table 2 with various *in-sample* periods. We then compare the stability and performance of the Tikhonov estimator to those of the shrinkage estimate $\widetilde{\boldsymbol{\Sigma}}_{s\to m}$. Next, in section 6.2, we perform similar experiments for the MV portfolio, varying the *in-sample* and *out-of-sample* periods as well as the required portfolio returns. We bypass the difficulties of estimating μ by assuming that it is known so that we can focus just on the effects of the different covariance estimators. Finally, in section 6.3, we compare the GMV and MV portfolio returns, and in section 6.4 we compare their predictions of risk.

6.1. GMV portfolio.

We simulate portfolio construction under the following scenario. We solve the GMV problem to construct a portfolio to hold for 1 month, the *out-of-sample* period T_o. We repeat this process for every month until we reach December 2007. Finally, we evaluate the variance of the *out-of-sample* returns from the GMV portfolio for each covariance estimator.

When performing this experiment, the choice of *in-sample* window size T_w is important. If T_w is too long, the data may include out-of-date information. On the other hand, if T_w is too short, the resulting covariance estimate could suffer from lack of information. We vary T_w from 1 year to 10 years. Later, in section 6.2, we will consider the change of the *out-of-sample* period T_o as well. We start each experiment at January 1968, giving 480 rebalancing steps for all values of T_w. For each covariance estimator, we perform the simulation for 20 different choices of 100 stocks.

6.1.1. Covariance estimators in experiments.

We perform the experiment above for all the covariance estimators from section 5.1 to section 5.9 plus two diagonal matrices, $\widetilde{\boldsymbol{\Sigma}}_V$ and

$\widetilde{\mathbf{\Sigma}}_{\boldsymbol{I}}$, for a total of 11 estimators. $\widetilde{\mathbf{\Sigma}}_V$ has diagonal elements equal to $\mathrm{Var}_s[\boldsymbol{r}(t)]$, and any correlations between stocks are neglected. $\widetilde{\mathbf{\Sigma}}_{\boldsymbol{I}}$ is an $N \times N$ identity matrix, which would yield an evenly distributed portfolio as the solution for the GMV problem (1.2); thus it is a good benchmark for a well-distributed portfolio. Since $\widetilde{\mathbf{\Sigma}}_{sample}$ is rank deficient, we modify it by adding small positive constants δ_i to its diagonal elements as in (3.17). To compute $\widetilde{\mathbf{\Sigma}}_{market}$ and $\widetilde{\mathbf{\Sigma}}_{s\to m}$, we need the monthly market return data $r_m(t)$ in (5.1). In this experiment, we adopt equally weighted market portfolio returns including distributions from the CRSP database as $r_m(t)$. According to Ledoit and Wolf [30, p. 607], an equally weighted market portfolio is better than a value-weighted market portfolio for explaining stock market variances.

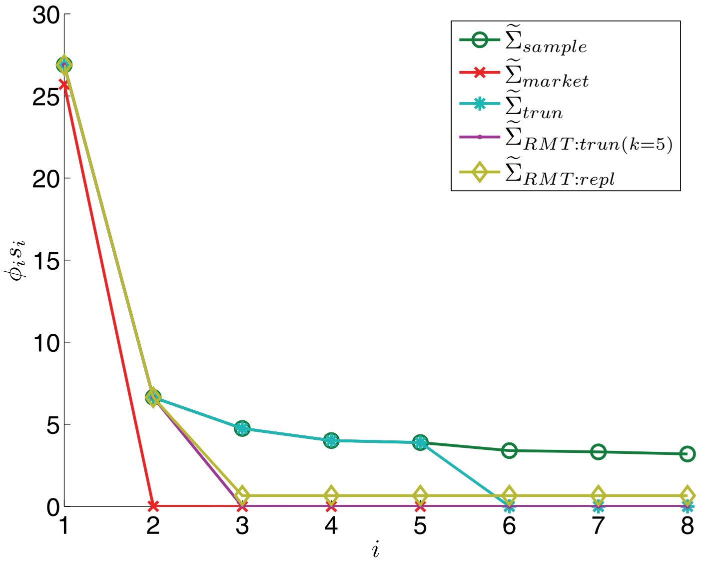
The parameter $\widehat{k}$ for $\widetilde{\mathbf{\Sigma}}_{trun}$ and $\widetilde{\mathbf{\Sigma}}_{s\to trun}$ is static, constant over all time periods. In our experiment, we perform the experiments with $\widehat{k} = 1, 5, 9$ for $\widetilde{\mathbf{\Sigma}}_{trun}$ and $\widehat{k} = 1, 2, 3$ for $\widetilde{\mathbf{\Sigma}}_{s\to trun}$. In contrast, the parameters of γ for $\widetilde{\mathbf{\Sigma}}_{s\to m}$ and $\widetilde{\mathbf{\Sigma}}_{s\to trun}$, $\widehat{k}$ for $\widetilde{\mathbf{\Sigma}}_{RMT:trun}$ and $\widetilde{\mathbf{\Sigma}}_{RMT:repl}$, and α for $\widetilde{\mathbf{\Sigma}}_{tikh}$ have their own parameter choice methods as described in section 5, so we dynamically determine these parameters each time the portfolio is rebalanced.

Figure 3 shows singular value plots from each estimator, which illustrates the filtering characteristics for the first *in-sample* period of $T_w = 4$ years with a particular set of 100 stocks.

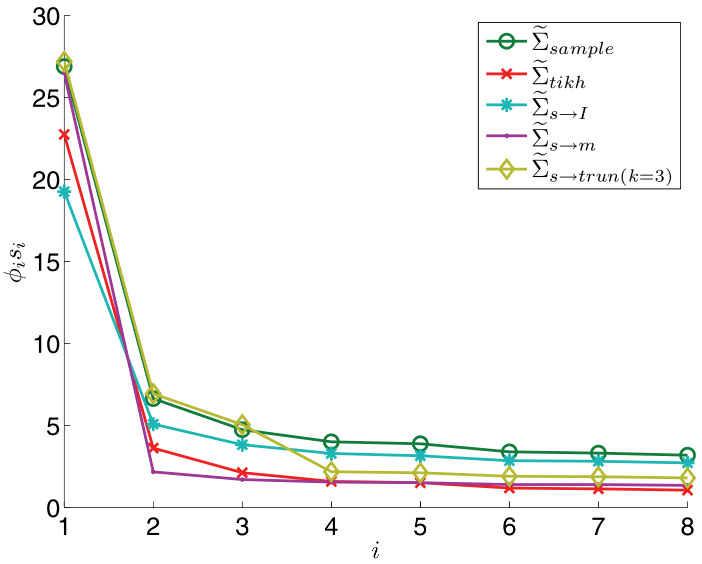
6.1.2. Effect of in-sample period T_w. For each randomly chosen data set ($i = 1, \ldots, 20$), we calculate $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}}$, the annualized standard deviation of the sample portfolio return, by multiplying the monthly standard deviation by $\sqrt{12}$. The subscript $\widetilde{\mathbf{\Sigma}}$ denotes the specific choice of covariance estimator. Figures 4(a) and 4(b) show the means of $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}}$ for the static estimators and the dynamic estimators. The standard deviations of the $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}}$ from each estimator were at most 0.56 for all time periods, except for the occurrence of values up to 3.38 for $\widetilde{\mathbf{\Sigma}}_{sample}$ and up to 6.50 for $\widetilde{\mathbf{\Sigma}}_{s\to trun(\widehat{k}=3)}$, so the results did not seem sensitive to the particular choice of 100 stocks.

For most estimators, the $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}}$ decrease until a particular T_w and increase after that point, showing the advantage of using a sufficient amount of history but not too much out-of-date information. This is particularly evident for $\widetilde{\mathbf{\Sigma}}_{sample}$ since it assumes that all of its data are reliable. At the opposite extreme, $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}_{market}}$ from $\widetilde{\mathbf{\Sigma}}_{market}$ increases with T_w, which implies that the correlation among stocks cannot be fully explained by a single market index. For small values of k, $\widetilde{\mathbf{\Sigma}}_{trun}$ behaves like $\widetilde{\mathbf{\Sigma}}_{market}$, but performance can be improved by taking $k \approx 5$, making the estimator less sensitive to out-of-date information. The diagonal $\widetilde{\mathbf{\Sigma}}_V$ shows a better tolerance to out-of-date information than $\widetilde{\mathbf{\Sigma}}_{sample}$, which may imply that the sample variance estimation is less sensitive to the choice of T_w than the sample covariance estimation. The estimators that dynamically determine the filtering parameters ($\widetilde{\mathbf{\Sigma}}_{tikh}$, $\widetilde{\mathbf{\Sigma}}_{s\to m}$, $\widetilde{\mathbf{\Sigma}}_{s\to I}$, $\widetilde{\mathbf{\Sigma}}_{s\to trun(\widehat{k}=1)}$, $\widetilde{\mathbf{\Sigma}}_{RMT:repl}$, and $\widetilde{\mathbf{\Sigma}}_{RMT:trun}$) also show good tolerance. Therefore, modestly filtered factor structures are better at filtering the out-of-date information than a single-factor or full-factor structure, but all estimators benefit from an appropriate choice of window size.

Compared to the truncation-based estimators like $\widetilde{\mathbf{\Sigma}}_{RMT:trun}$ and $\widetilde{\mathbf{\Sigma}}_{trun}$, Tikhonov generally performs better when the *in-sample* period is shorter than its own optimal size, which



(a) The singular values from truncation-based estimator.

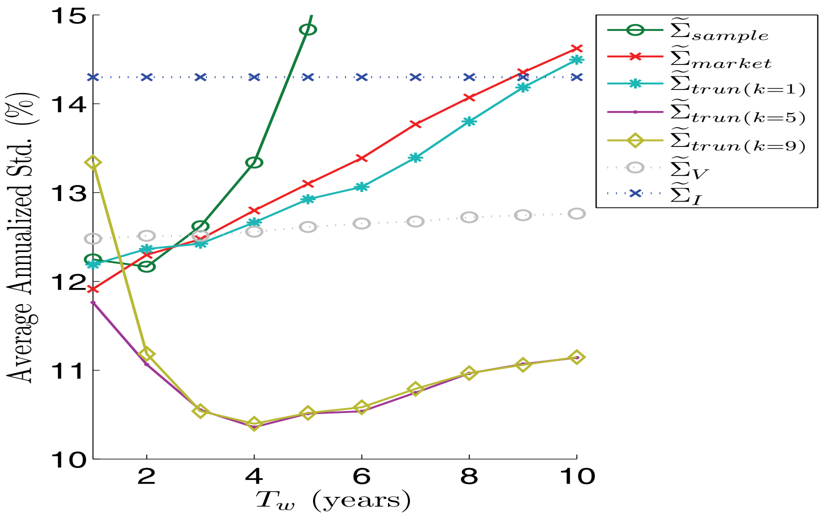


(b) The singular values from shrinkage-based estimator.

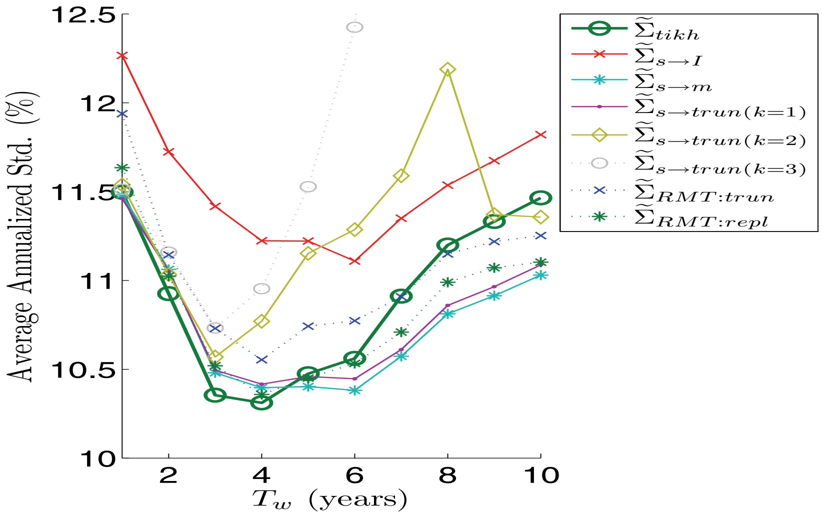
Figure 3. *GMV portfolios: the singular values from each estimator when $T_w = 4$ years.*

is $T_w = 4$. This result can be explained by the characteristics of their filtering functions. While $\phi^2_{tikh}(\lambda_i)$ preserves the relative magnitudes of eigenvalues by gradual attenuation, $\phi^2_{RMT:trun}(\lambda_i)$ or $\phi^2_{trun}(\lambda_i)$ discard them all. Thus, when the smallest eigenvalues are still important, the Tikhonov filter shows superiority empirically. However, as noise level increases with longer T_w, the performance reverses.

Compared to the other shrinkage-based estimators, Tikhonov filtering $\phi^2_{tikh}(\lambda_i)$ preserves



(a) The mean of $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}}$ from the static estimator $\widetilde{\mathbf{\Sigma}}$.

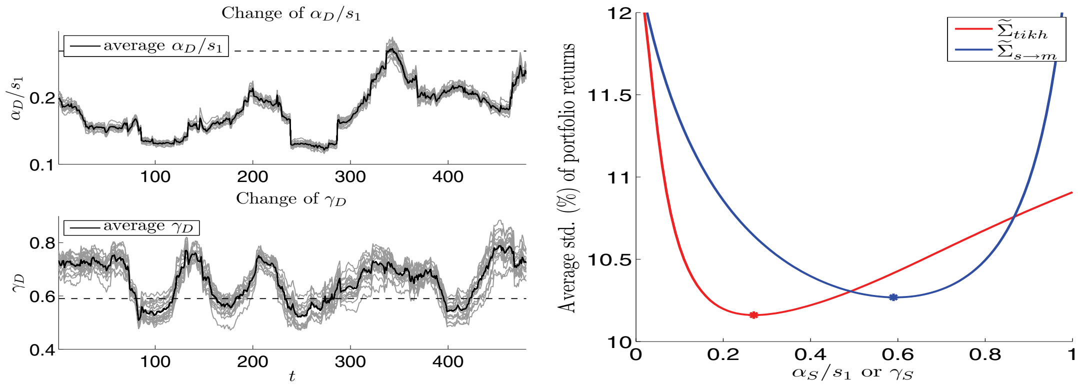


(b) The mean of $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}}$ from the dynamic estimator $\widetilde{\mathbf{\Sigma}}$.

Figure 4. *GMV portfolios: the mean of* $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}}$ *over different choices of* T_w.

the smallest but still informative factors better than a single rate reduction by $\phi^2_{s\rightarrow m}(\lambda_i)$ and $\phi^2_{s\rightarrow trun}(\lambda_i)$ or a replacement with a constant value by $\phi^2_{RMT:repl}(\lambda_i)$ when T_w is relatively short ($T_w < 4$). On the other hand, for $T_w > 7$, it becomes evident that $\widetilde{\mathbf{\Sigma}}_{s\rightarrow m}$, $\widetilde{\mathbf{\Sigma}}_{s\rightarrow trun(\widehat{k}=1)}$, and $\widetilde{\mathbf{\Sigma}}_{RMT:repl}$ show better performances than $\widetilde{\mathbf{\Sigma}}_{tikh}$. This is because $\widetilde{\mathbf{\Sigma}}_{tikh}$ has relatively weaker tolerance to the contamination by out-of-date information.

6.1.3. Stability of Tikhonov parameter choice. In this section, we evaluate the stability of our parameter choice method from section 4. For a particular choice of 100 stocks, we observe the change of the dynamic parameters α for $\widetilde{\mathbf{\Sigma}}_{tikh}$ and γ for $\widetilde{\mathbf{\Sigma}}_{s\rightarrow m}$. In this experiment,



(a) Variation in the dynamic α_D and γ_D over the course of 20 experiments.

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

Figure 5. *GMV portfolios: the performance of static and dynamic choices of* α *and* γ *in* 20 *experiments.*

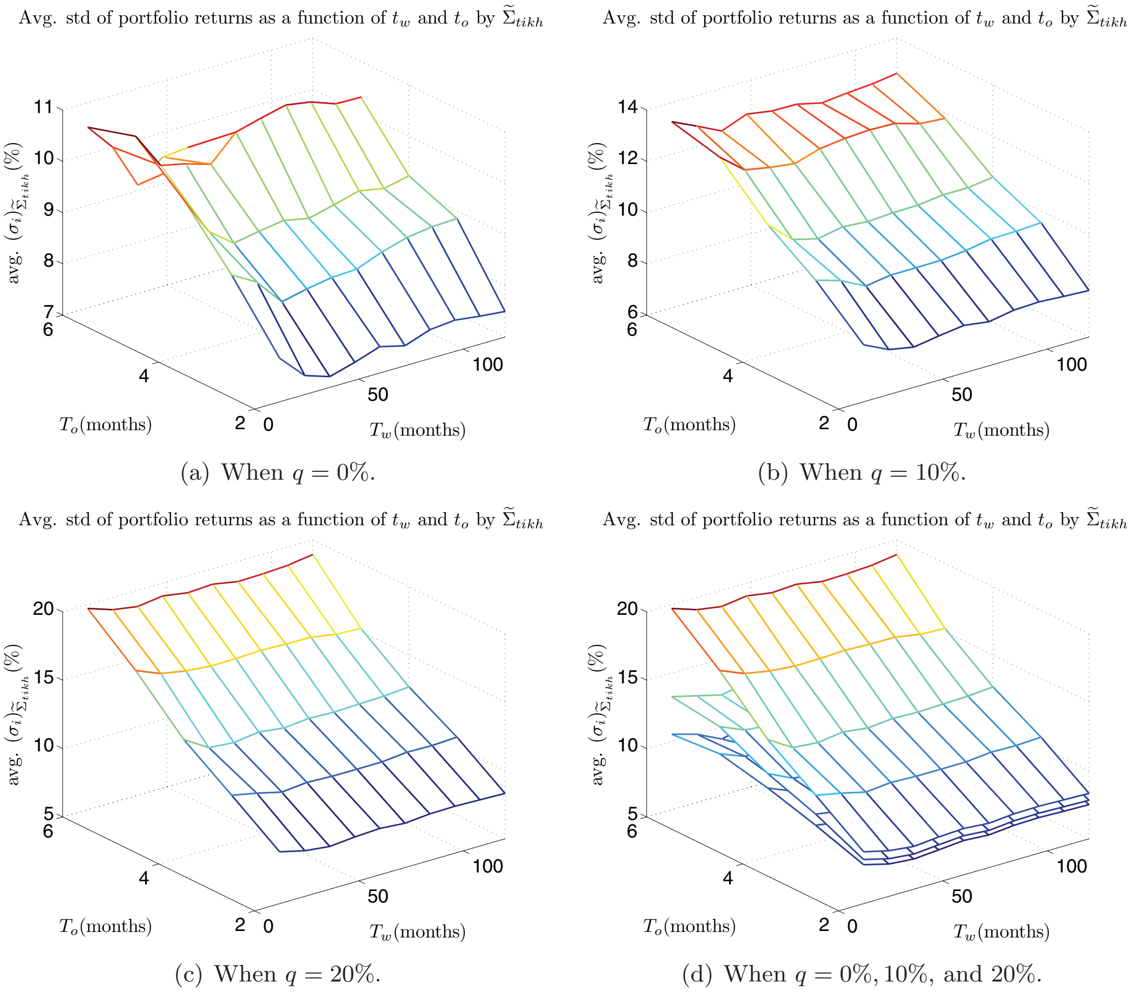
we set the window size as $T_w = 48$ because both estimators have the smallest mean value of $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}}$ for that window size.

Figure 5(a) illustrates the change of the ratio of the dynamically chosen Tikhonov parameter α_D to the largest singular value s_1 of $\mathrm{Corr}_s[\boldsymbol{r}(t)]$ and the change of γ_D for $\widetilde{\mathbf{\Sigma}}_{s\to m}$. The results for 20 choices of the 100 stocks are shown, showing that both parameter choice methods for α_D and γ_D are quite stable during the whole experiment. The resulting annualized standard deviations of $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}}$ range from 10.16% to 10.30% for $\widetilde{\mathbf{\Sigma}}_{tikh}$ and $\widetilde{\mathbf{\Sigma}}_{s\to m}$, for both the static and dynamically determined parameters.

We repeated this numerical experiment keeping the ratio α/s_1 and the parameter γ constant over all time periods. (We use the notation α_S and γ_S for this statically determined parameter.) This static 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 and γ_D. Figure 5(b) shows how the standard deviation of portfolio returns changes as α_S/s_1 and γ_S increase. The optimal ratio α_S^*/s_1 was 0.27 with resulting standard deviation of portfolio returns 10.16%, and the optimal γ_S^* was 0.59 with resulting standard deviation 10.27%. These statically optimal values are represented by dashed lines in Figure 5(a). Therefore, we can see that both α_D/s_1 and γ_D remain near their statically optimal values α_S^*/s_1 and γ_S^*. Moreover, the static and varying α values produce similar risk variance.

6.2. MV portfolio. Now, we observe the behavior of the MV portfolio resulting from each covariance estimator. In this experiment, we vary the *out-of-sample* period T_o and the required portfolio return q as well as the *in-sample* period T_w. We change T_o from 2 months to 6 months,[2] T_w from 1 year to 10 years, and q from 0% to 20%. As we mentioned before, the performance of the MV portfolio is quite sensitive to the estimation of stock returns $\boldsymbol{\mu}$.

[2] We omit the case of $T_o = 1$ month since it gives us a trivial result that the portfolio returns are equal to the required portfolio return q making $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}}$ zero for any covariance $\widehat{\mathbf{\Sigma}}$ and any window size T_w. This is because $\boldsymbol{\mu}$ equals the realized stock returns $\boldsymbol{r}(t)$ in the *out-of-sample* period.



(a) When $q = 0\%$.

(b) When $q = 10\%$.

(c) When $q = 20\%$.

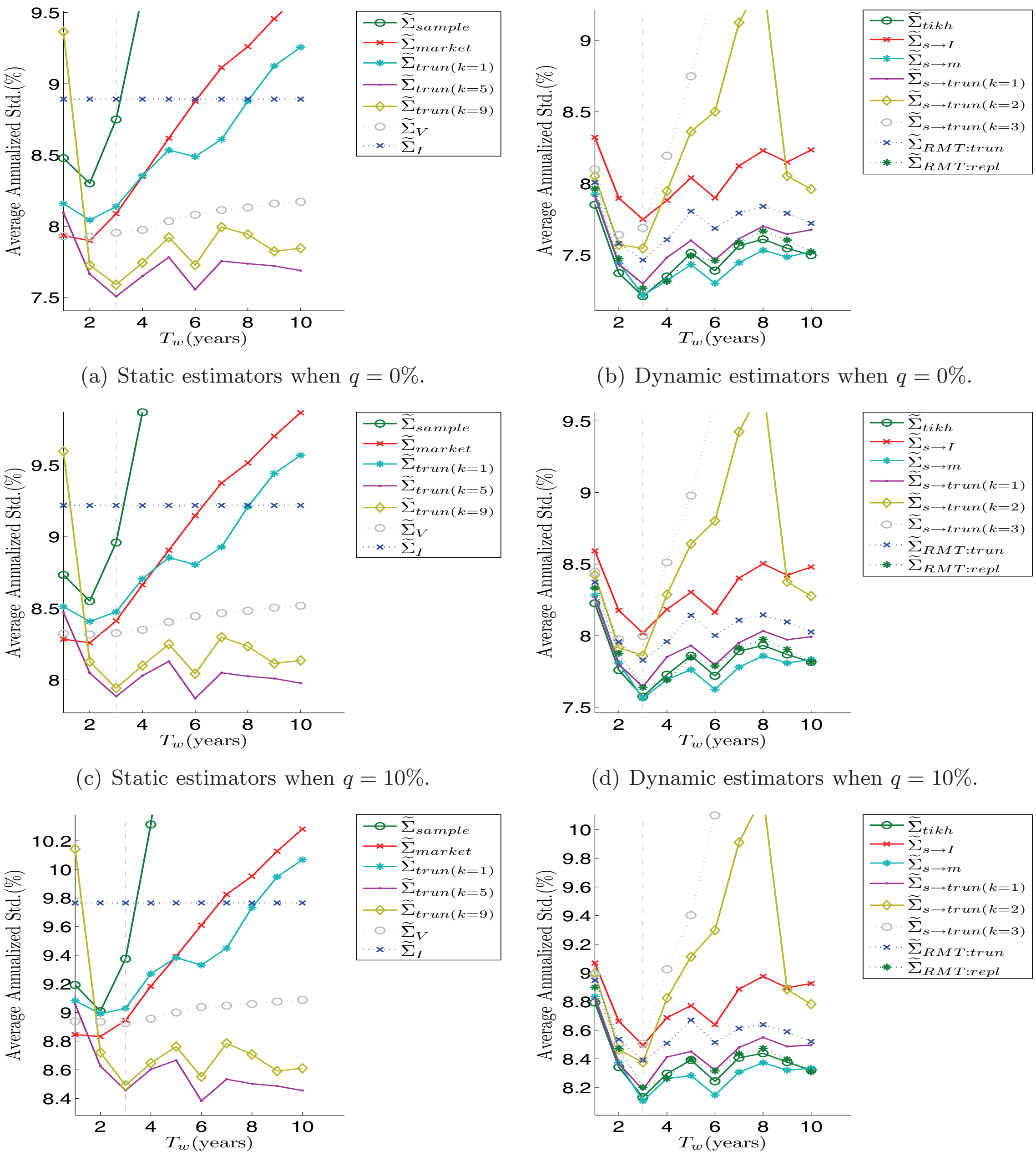
(d) When $q = 0\%, 10\%,$ and 20%.

Figure 6. *MV portfolios: the average annualized standard deviations* $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}_{tikh}}$ *of portfolio returns as* in-sample *period* T_w *and* out-of-sample *period* T_o *change with different settings of required portfolio return* q.

In order to evaluate covariance estimation with no influence of mean estimation, we assume a perfect prediction of stock returns $\boldsymbol{\mu}$, which means we estimate $\boldsymbol{\mu}$ by the average $\boldsymbol{r}(t)$ during the *out-of-sample* period.

6.2.1. Effect of out-of-sample period T_o. The *out-of-sample* period T_o determines how fast we react to the changes in the market. Figure 6 shows how the average $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}_{tikh}}$ changes as T_o and T_w vary for $q = 0\%$, 10%, and 20%. We can see that $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}_{tikh}}$ has a tendency to increase as we hold the portfolio for longer T_o. Similar results were obtained for all other covariance estimators.

6.2.2. Effect of in-sample period T_w. Similar to Figure 4 for the GMV experiment, we compared the mean of $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}}$ for different covariance estimators with varying T_w and q in Figure 7. Based on the result of section 6.2.1, we fixed T_o as 2 months in order to compare



(a) Static estimators when $q = 0\%$.

(b) Dynamic estimators when $q = 0\%$.

(c) Static estimators when $q = 10\%$.

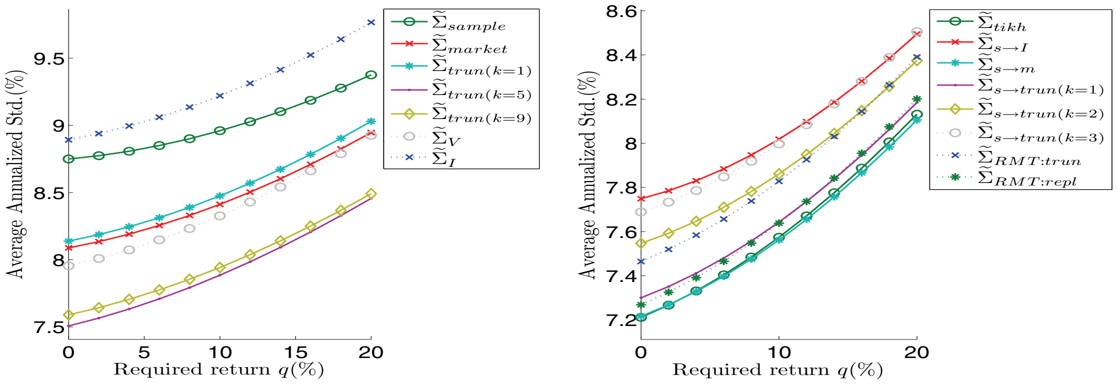
(d) Dynamic estimators when $q = 10\%$.

(e) Static estimators when $q = 20\%$.

(f) Dynamic estimators when $q = 20\%$.

Figure 7. *MV portfolios: the mean of $(\sigma_i)_{\widetilde{\Sigma}}$ over different choices of T_w and q when $T_o = 2$ months.*

the smallest standard deviations from the estimators. The behaviors of MV portfolios with respect to the change of T_w are very similar to GMV portfolios for most covariance estimators. For example, as we observed for the previous GMV experiments, the MV portfolios in Figure 7 also suffered from lack of information when T_w was too short and suffered from out-of-date



(a) Static estimators. (b) Dynamic estimators.

Figure 8. *MV portfolios: the average annualized $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}}$ versus required return q for each estimator when $T_o = 2$ months and $T_w = 3$ years.*

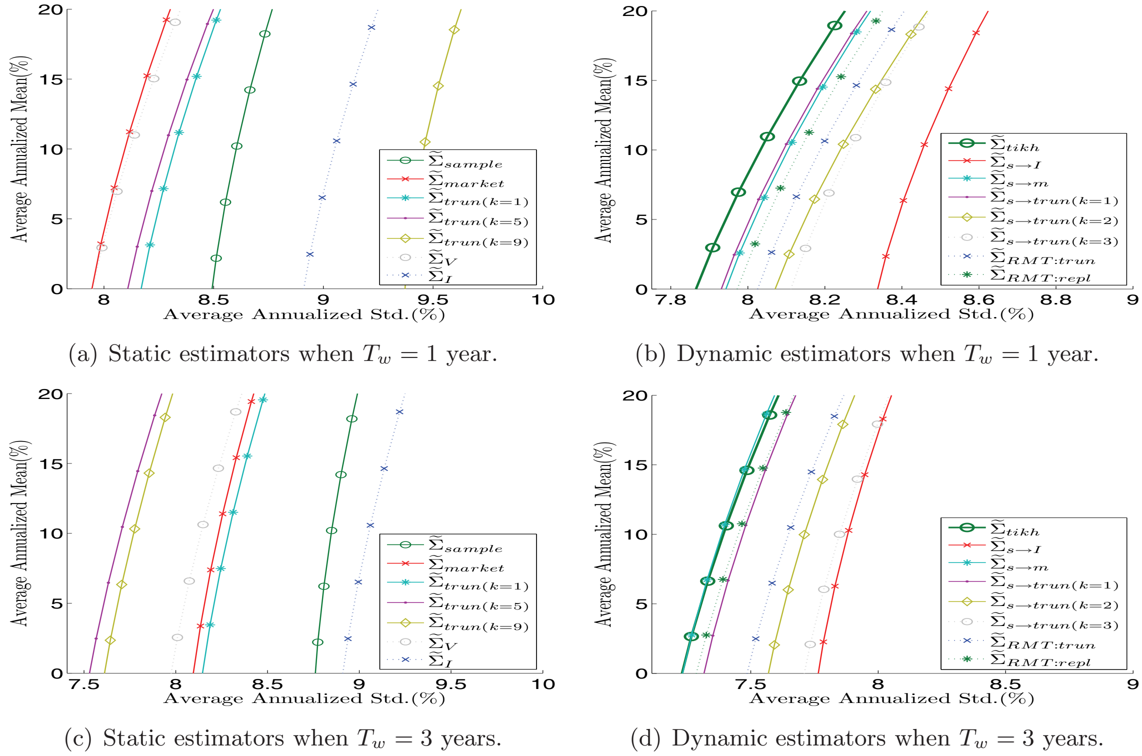
information when T_w was too long. This implies that the choice of window size T_w is very important for the MV portfolio as well as the GMV portfolio. Moreover, each estimator shows very similar shapes of curves for the GMV and the MV problems, except that the curves for the MV problems tend to shift upward as q increases.

However, in contrast to the GMV problem, where most of the competitive estimators have optimal T_w around 4 years, the optimal T_w for most estimators was around 3 years for the MV problem (gray-colored vertical dot-dash lines indicate $T_w = 3$ years in Figure 7). This may be because they have different *out-of-sample* periods: $T_o = 1$ month for the GMV problem in Figure 4 and $T_o = 2$ months for the MV problem in Figure 7.

6.2.3. Effect of required portfolio return q. Figure 6(d) summarizes the results from Figure 6(a) to Figure 6(c). As we can expect, the surfaces of $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}_{tikh}}$ move upward as q increases. For all the estimators $\widetilde{\mathbf{\Sigma}}$ with particular choices of $T_o = 2$ months and $T_w = 3$ years, Figure 8 also shows that $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}}$ gradually increases as q increases from 0% to 20%, which explains a trade-off between risk and return from the MV portfolio.

6.2.4. Efficiency of portfolio. The mean-variance plot shows the efficiency of the MV portfolios. Let $(\mu_i)_{\widetilde{\mathbf{\Sigma}}}$ denote the annualized mean of the realized portfolio returns in the ith random choice of 100 stocks ($i = 1, \ldots, 20$). In order to evaluate the portfolio efficiency by each estimator, we compare the change of average $(\mu_i)_{\widetilde{\mathbf{\Sigma}}}$ versus the change of average $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}}$ with varying the required return q from 0% to 20%. Figure 9 presents the average of realized means and standard deviations of all the estimators for the cases of $T_o = 2$ months and $T_w = 1$ year or 3 years. Curves to the left of and above the others correspond to the more efficient portfolios.

When $T_w = 1$ year, where we have insufficient historical data, $\widetilde{\mathbf{\Sigma}}_{tikh}$ generates the most efficient portfolios (see Figure 9(b)). The shrinkage estimators with a target of a single factor like $\widetilde{\mathbf{\Sigma}}_{s\to m}$ and $\widetilde{\mathbf{\Sigma}}_{s\to trun(k=1)}$ are also efficient compared to other dynamic estimators. When $T_w = 3$ years, where we have near optimal historical data, $\widetilde{\mathbf{\Sigma}}_{tikh}$, $\widetilde{\mathbf{\Sigma}}_{s\to m}$, $\widetilde{\mathbf{\Sigma}}_{RMT:repl}$, and $\widetilde{\mathbf{\Sigma}}_{s\to m}$ generate relatively efficient portfolios (see Figure 9(d)).



(a) Static estimators when $T_w = 1$ year.

(b) Dynamic estimators when $T_w = 1$ year.

(c) Static estimators when $T_w = 3$ years.

(d) Dynamic estimators when $T_w = 3$ years.

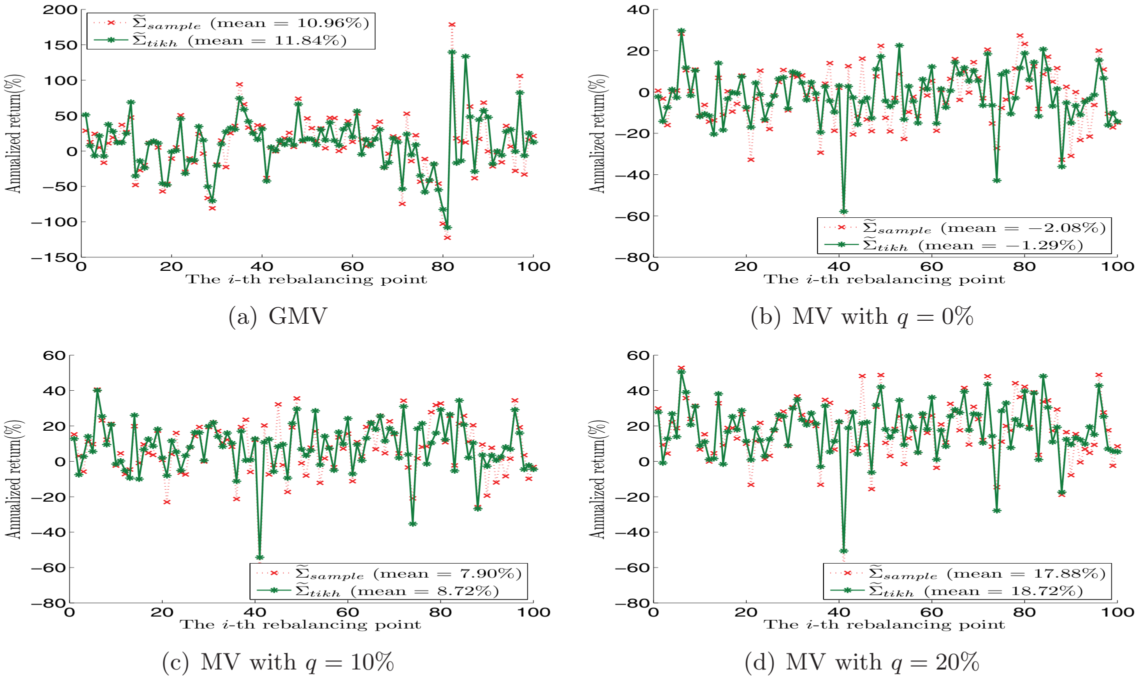
Figure 9. *MV portfolios: the average annualized* $(\mu_i)_{\widetilde{\Sigma}}$ *versus average annualized* $(\sigma_i)_{\widetilde{\Sigma}}$.

6.3. Comparison of GMV and MV portfolios. Now we observe how the covariance estimators affect the realized portfolio returns at every rebalancing point for the GMV and the MV problems. For instance, Figure 10 shows the fluctuations of the portfolio returns by $\widetilde{\mathbf{\Sigma}}_{sample}$ and $\widetilde{\mathbf{\Sigma}}_{tikh}$ at the first 100 rebalancing points when $T_w = 3$ years and $T_o = 2$ months. While the annualized returns of the GMV portfolios fluctuate around 11%, the annualized returns of the MV portfolios fluctuate around their required return q. Note that the GMV mean return is greater than that for the MV portfolio with $q = 0\%$. Similarly, the standard deviations in Figure 4 are greater than the corresponding ones in Figures 7(a) and 7(b).

On the other hand, for both GMV and MV, the $\widetilde{\mathbf{\Sigma}}_{tikh}$ portfolios have greater mean return and smaller variance than those from $\widetilde{\mathbf{\Sigma}}_{sample}$, which implies more efficient portfolios. This result is consistent with the plots of means versus standard deviations in Figure 9.

6.4. Risk prediction. Laloux et al. [29] showed empirically that their estimator $\widetilde{\mathbf{\Sigma}}_{RMT:repl}$ predicts the risk more accurately than $\widetilde{\mathbf{\Sigma}}_{sample}$. They simply divided the dataset into two equal time periods for *in-sample* and *out-of-sample* periods and compared the estimated standard deviation $(\boldsymbol{w}^T\widetilde{\mathbf{\Sigma}}\boldsymbol{w})^{\frac{1}{2}}$ from (1.1) to the realized standard deviation $(\sigma_i)_{\widetilde{\mathbf{\Sigma}}}$ for the *out-of-sample* period. They assumed perfect prediction for means of stock returns as we did in section 6.2.

We evaluate the accuracy of the risk prediction of each covariance estimator in a similar way. However, rather than following their equal division of *in-sample* and *out-of-sample* periods, we varied T_w with $T_o = 2$ months, and we simulated the rebalancing scenario as



(a) GMV (b) MV with $q = 0\%$ (c) MV with $q = 10\%$ (d) MV with $q = 20\%$

Figure 10. *GMV and MV portfolios: the annualized portfolio returns at the rebalancing points for the GMV and the MV problems with different required returns q.*

in section 6.2. Finally, we compute the relative difference between the average estimated standard deviations from (1.1) and the average realized standard deviations for the most competitive estimators.

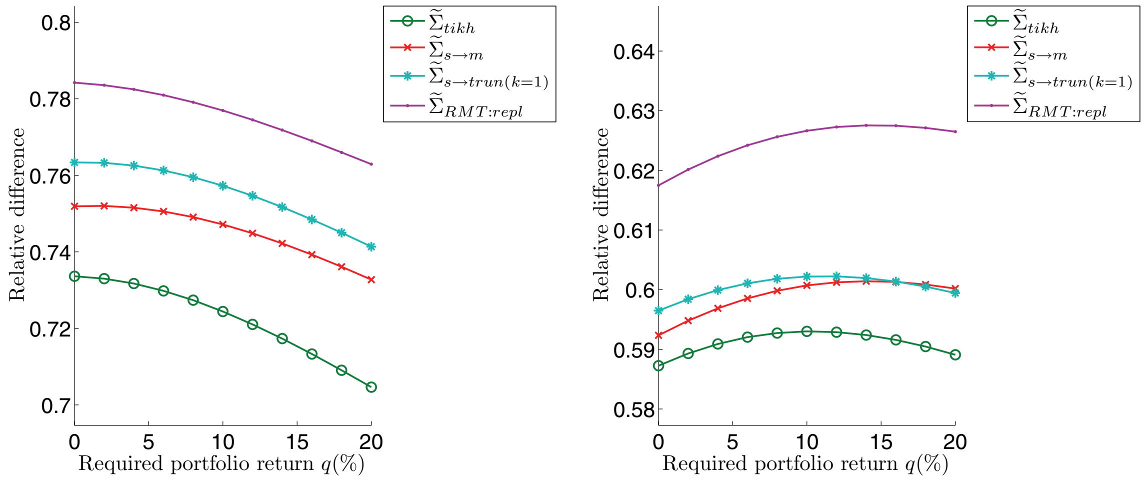
Figure 11 shows the relative difference for the cases of $T_w = 1$ and 3 years, which correspond to the case of insufficient historical data and the minimizer of average $(\sigma_i)_{\widetilde{\boldsymbol{\Sigma}}}$. The realized standard deviations were greater than the estimated standard deviations for all estimators. However, it turns out that $\widetilde{\boldsymbol{\Sigma}}_{tikh}$ has the smallest difference for both cases, giving us the best risk prediction.

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 $\mathrm{Corr}_s[\boldsymbol{r}(t)]$. The Tikhonov estimator $\widetilde{\boldsymbol{\Sigma}}_{tikh}$ attenuates smaller eigenvalues more intensely, which is a key difference between it and the other filter functions.

In order to choose an appropriate Tikhonov parameter α that determines the intensity of attenuation, we formulated an optimization problem minimizing the difference between $\mathrm{Corr}_s[\boldsymbol{\epsilon}_z(t)]$ and $\boldsymbol{I}_N$ based on the assumption that the unsystematic factors are uncorrelated.

We performed empirical experiments to evaluate covariance estimators. For the GMV portfolio selection problem, the Tikhonov choice gave the smallest average standard deviation of the return when the *out-of-sample* period was 3 or 4 years, and it was not much worse than competitors for other periods. The choice of parameter was relatively stable. For the MV



(a) When $T_w = 1$ year. (b) When $T_w = 3$ years.

Figure 11. *MV portfolios: the relative differences between average estimated risks and average realized risks by each covariance matrix with varying different required returns q.*

portfolio selection problem, the Tikhonov choice was among the most efficient portfolios and the best estimates of risk. Moreover, the Tikhonov estimator performs relatively well in the circumstance of insufficient historical data. We believe that this parameter selection method is quite promising relative to previously proposed methods.

Acknowledgments. We are grateful for the comments of the associate editor and the referees, who provided additional references and suggestions to make the presentation clearer and the experiments more convincing.

REFERENCES

[1] M. S. Bartlett, *Tests of significance in factor analysis*, Br. J. Psych. Stat. Sec., 3 (1950), pp. 77–85.
[2] C. Bengtsson and J. Holst, *On portfolio selection: Improved covariance matrix estimation for Swedish asset returns*, in 31st Meeting of the Euro Working Group on Financial Modeling, Agia Napa, Cyprus, 2002; also available online from http://hermes-mba.cc.ucy.ac.cy/conferences/ewgfm/papers/bengtsson_holst_20021004.pdf.
[3] M. J. Best and R. R. Grauer, *On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results*, Rev. Financ. Stud., 4 (1991), pp. 315–342.
[4] R. B. Cattell, *The scree test for the number of factors*, Mult. Behav. Res., 1 (1966), pp. 245–276.
[5] L. K. C. Chan, J. Karceski, and J. Lakonishok, *On portfolio optimization: Forecasting covariances and choosing the risk model*, Rev. Financ. Stud., 12 (1999), pp. 937–974.
[6] V. K. Chopra, C. R. Hensel, and A. L. Turner, *Massaging mean-variance inputs: Returns from alternative global investment strategies in the* 1980s, Management Sci., 39 (1993), pp. 845–855.
[7] V. K. Chopra and W. T. Ziemba, *The effect of errors in means, variances, and covariances on optimal portfolio choice*, J. Portfolio Management, 19:2 (1993), pp. 6–11.
[8] T. Conlon, H. J. Ruskin, and M. Crane, *Random matrix theory and fund of funds portfolio optimisation*, Phys. A, 382 (2007), pp. 565–576.
[9] G. Connor and R. A. Korajczyk, *A test for the number of factors in an approximate factor model*, J. Finance, 48 (1993), pp. 1263–1291.

[10] V. DeMiguel, L. Garlappi, F. J. Nogales, and R. Uppal, *A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms*, Management Sci., 55 (2009), pp. 798–812.
[11] V. DeMiguel, L. Garlappi, and R. Uppal, *Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy?*, Rev. Financ. Stud., 22 (2009), pp. 1915–1953.
[12] V. DeMiguel and F. J. Nogales, *Portfolio selection with robust estimation*, Oper. Res., 57 (2009), pp. 560–577.
[13] E. J. Elton and M. J. Gruber, *Estimating the dependence structure of share prices—implications for portfolio selection*, J. Finance, 28 (1973), pp. 1203–1232.
[14] 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 (1965), pp. 205–224.
[15] G. H. Golub, M. Heath, and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics, 21 (1979), pp. 215–223.
[16] L. Guttman, *Some necessary conditions for common-factor analysis*, Psychometrika, 19 (1954), pp. 149–161.
[17] P. C. Hansen, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev., 34 (1992), pp. 561–580.
[18] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images: Matrices, Spectra, and Filtering*, SIAM, Philadelphia, 2006.
[19] P. C. Hansen and D. P. O'Leary, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci. Comput., 14 (1993), pp. 1487–1503.
[20] R. J. Hanson, *A numerical method for solving Fredholm integral equations of the first kind using singular values*, SIAM J. Numer. Anal., 8 (1971), pp. 616–622.
[21] W. Härdle and L. Simar, *Applied Multivariate Statistical Analysis*, Springer, Berlin, 2003.
[22] A. E. Hoerl and R. W. Kennard, *Ridge regression: Applications to nonorthogonal problems*, Technometrics, 12 (1970), pp. 69–82.
[23] A. E. Hoerl and R. W. Kennard, *Ridge regression: Biased estimation for nonorthogonal problems*, Technometrics, 12 (1970), pp. 55–67.
[24] J. E. Jackson, *A User's Guide to Principal Components*, Wiley-Interscience, New York, 2003.
[25] R. Jagannathan and T. Ma, *Risk reduction in large portfolios: Why imposing the wrong constraints helps*, J. Finance, 58 (2003), pp. 1651–1684.
[26] B. F. King, *Market and industry factors in stock price behavior*, J. Business, 39 (1966), pp. 139–190.
[27] J. Kwapień, S. Drożdż, and P. Oświęcimka, *The bulk of the stock market correlation matrix is not pure noise*, Phys. A, 359 (2006), pp. 589–606.
[28] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, *Noise dressing of financial correlation matrices*, Phys. Rev. Lett., 83 (1999), pp. 1467–1470.
[29] L. Laloux, P. Cizeau, M. Potters, and J.-P. Bouchaud, *Random matrix theory and financial correlations*, Internat. J. Theoret. Appl. Finance, 3 (2000), pp. 391–397.
[30] O. Ledoit and M. Wolf, *Improved estimation of the covariance matrix of stock returns with an application to portfolio selection*, J. Empirical Finance, 10 (2003), pp. 603–621.
[31] O. Ledoit and M. Wolf, *A well-conditioned estimator for large-dimensional covariance matrices*, J. Multivariate Anal., 88 (2004), pp. 365–411.
[32] J. Lintner, *The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets*, Rev. Econom. Statist., 47 (1965), pp. 13–37.
[33] H. Markowitz, *Portfolio selection*, J. Finance, 7 (1952), pp. 77–91.
[34] M. L. Mehta, *Random Matrices*, 3rd ed., Academic Press, New York, 2004. Cited by [38].
[35] J. Mossin, *Equilibrium in a capital asset market*, Econometrica, 34 (1966), pp. 768–783.
[36] D. L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind*, J. Assoc. Comput. Mach., 9 (1962), pp. 84–97.
[37] V. Plerou, P. Gopikrishnan, L. A. Nunes Amaral, M. Meyer, and H. E. Stanley, *Scaling of the distribution of price fluctuations of individual companies*, Phys. Rev. E, 60 (1999), pp. 6519–6529.
[38] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. Nunes Amaral, T. Guhr, and H. E. Stanley, *Random matrix approach to cross-correlations in financial data*, Phys. Rev. E, 65 (2002), article 066126.

[39] R. Roll and S. A. Ross, *An empirical investigation of the arbitrage pricing theory*, J. Finance, 35 (1980), pp. 1073–1103.
[40] S. A. Ross, *The arbitrage theory of capital asset pricing*, J. Econom. Theory, 13 (1976), pp. 341–360.
[41] B. W. Rust, *Parameter selection for constrained solutions to ill-posed problems*, in Proceedings of the 32nd Symposium on the Interface: Computing Science and Statistics, 2000, pp. 333–347; also available online from http://www.galaxy.gmu.edu/interface/I00/I2000Proceedings/BRust/BRust.pdf.
[42] B. W. Rust and D. P. O'Leary, *Residual periodograms for choosing regularization parameters for ill-posed problems*, Inverse Problems, 24 (2008), article 034005.
[43] W. F. Sharpe, *A simplified model for portfolio analysis*, Management Sci., 9 (1963), pp. 277–293.
[44] W. F. Sharpe, *Capital asset prices: A theory of market equilibrium under conditions of risk*, J. Finance, 19 (1964), pp. 425–442.
[45] G. W. Stewart and J. Sun, *Matrix Perturbation Theory*, Academic Press, Boston, 1990.
[46] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems*, John Wiley & Sons, New York, 1977.
[47] J. L. Treynor, *Toward a theory of the market value of risky assets*, 1962 manuscript published as Chapter 2 in Asset Pricing and Portfolio Performance: Models, Strategy and Performance Metrics, R. A. Korayczyk, ed., Risk Books, London, 1999.
[48] C. Trzcinka, *On the number of factors in the arbitrage pricing model*, J. Finance, 41 (1986), pp. 347–368.
[49] 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 (1963), pp. 97–101.
[50] W. F. Velicer, *Determining the number of components from the matrix of partial correlations*, Psychometrika, 41 (1976), pp. 321–327.
[51] S. Wold, *Cross-validatory estimation of the number of components in factor and principal component analysis*, Technometrics, 20 (1978), pp. 397–405.