Journal of Machine Learning Research 21 (2020) 1-37 Submitted 11/19; Revised 4/20; Published 6/20

Loss Control with Rank-one Covariance Estimate for Short-term Portfolio Optimization

Zhao-Rong Lai LAIZHR@JNU.EDU.CN
Department of Mathematics
Jinan University
Guangzhou 510632, China

Liming Tan TANLIMING@STU2014.JNU.EDU.CN
College of Economics
Jinan University
Guangzhou 510632, China

Xiaotian Wu WXIAOTIAN@JNU.EDU.CN
Liangda Fang FANGLD@JNU.EDU.CN
Department of Computer Science
Jinan University
Guangzhou 510632, China

Editor: Csaba Szepesvari

Abstract

In short-term portfolio optimization (SPO), some financial characteristics like the expected return and the true covariance might be dynamic. Then there are only a small window size w of observations that are sufficiently close to the current moment and reliable to make estimations. w is usually much smaller than the number of assets d, which leads to a typical undersampled problem. Worse still, the asset price relatives are not likely subject to any proper distributions. These facts violate the statistical assumptions of the traditional covariance estimates and invalidate their statistical efficiency and consistency in risk measurement. In this paper, we propose to reconsider the function of covariance estimates in the perspective of operators, and establish a rank-one covariance estimate in the principal rank-one tangent space at the observation matrix. Moreover, we propose a loss control scheme with this estimate, which effectively catches the instantaneous risk structure and avoids extreme losses. We conduct extensive experiments on 7 real-world benchmark daily or monthly data sets with stocks, funds and portfolios from diverse regional markets to show that the proposed method achieves state-of-the-art performance in comprehensive downside risk metrics and gains good investing incomes as well. It offers a novel perspective of rank-related approaches for undersampled estimations in SPO.

Keywords: rank-one covariance estimate, short-term portfolio optimization, undersampled condition, loss control, downside risk

1. Introduction

Short-term portfolio optimization (SPO) has emerged recently as a new topic in the machine learning community (Li and Hoi, 2012; Li et al., 2015; Huang et al., 2016; Li and Hoi, 2014; Li et al., 2016; Lai et al., 2018c). It manages the portfolio over a set of assets

in the short term and tries to achieve some realistic financial aims, such as gaining excess returns, diversifying individual risks or reducing extreme losses.

Originating from the classic mean-variance theory (Markowitz, 1952), SPO shares some basic concepts with the traditional long-term portfolio optimization (LPO), such as getting more incomes with less risk if possible. However, SPO faces much more complicated realistic difficulties and challenges that LPO does not have. One of the main difficulties is to catch the rapidly-changing financial circumstance to make timely and reliable decisions. Specifically, some financial characteristics like the expected return and the true covariance might be dynamic $\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t$ rather than static $\boldsymbol{\mu}, \boldsymbol{\Sigma}$. In this case, we usually have only a small window size w of observations that are sufficiently close to the current moment t and thus reliable to make estimations. w is usually much smaller than the number of assets d, which immediately invalidates the statistical efficiency and consistency of the traditional covariance estimates (e.g., the unbiased estimate and the maximum-likelihood estimate). Besides, other strict statistical assumptions regarding asset price relatives or probability distributions seldom hold in the SPO scenario, since the financial characteristics are dynamic. Therefore, it has been realized that machine learning methods without strict assumptions are preferable, if they are more productive and effective in a well-defined technical sense (Das et al., 2013; Shen et al., 2014; Ho et al., 2015; Lai et al., 2018c).

Without assumptions of good statistical properties, some existing machine learning methods for SPO turn to discoveries of empirical and behavioral finance, like the mean reversion criterion (Jegadeesh, 1990, 1991) and irrational investing behaviors (Kahneman and Tversky, 1979; Shiller, 2003, 2000). For example, the moving average reversion (OLMAR, Li et al. 2015) and the robust median reversion (RMR, Huang et al. 2016) are two trend-reversing systems, while the peak price tracking (PPT, Lai et al. 2018b) is a trend-following one. There is also a composite trend representation system (AICTR, Lai et al. 2018a) that combines both trends. These systems adaptively catch different kinds of instantaneous trend information and achieve favorable investing incomes. However, they do not have effective risk terms in their optimization models, thus they are actually vulnerable to downside risk and extreme losses (see Section 4.3).

In quantitative finance, risk can be interpreted by the difference between reality and expectation of the investing return. Hence covariance is a suitable risk metric that has been used for many years since it is first proposed by Markowitz (1952). However, it is adopted mainly in LPO since it is more probable to satisfy the above-mentioned strict statistical assumptions. For example, Brodie et al. (2009) propose to establish a kind of sparse and stable Markowitz portfolio (SSMP) via ℓ^1 regularization. As a trade-off between ℓ^1 and ℓ^2 regularizations, Ho et al. (2015) propose to penalize portfolio via the elastic net (Zou and Hastie, 2005). These mean-variance-related methods do establish sparse portfolios in LPO, but have not addressed the failure of traditional covariance estimates in SPO. In order to establish sparse portfolios in SPO, Lai et al. (2018c) propose to maximize the increasing factor as well as regularizing the ℓ^1-norm of the portfolio, and solve the model by the alternating direction method of multipliers (ADMM, Boyd et al. 2010). But the ℓ^1 regularization does not have an effective scheme of risk measurement and may affect the performance in loss control.

Lacking a suitable covariance estimate in SPO may cause high downside risk and extreme losses, which may lead to serious consequences, such as large-scale withdrawals of

investors and mandatory liquidation. Then the corresponding mutual fund may not recover even though it has a good profiting strategy in the long run. Since the traditional covariance estimate is neither sufficiently-sampled nor subject to good statistical assumptions in SPO, we have to reconsider its functions from a very different perspective. In this paper, we see it as a symmetric quadratic form operator composed by an orthogonal projection and a Gram matrix. Then we give a detailed investigation of the underlying operator space $\mathbb{S}^d$ ($d \times d$-dimensional symmetric matrix space) in its rank-related structure, including determinantal varieties, tangent spaces, and the corresponding orthogonal projections. We further extract the mutually orthogonal rank-one tangent spaces at the spectral components of the observation matrix from $\mathbb{S}^d$. In these tangent spaces, the principal rank-one tangent space reveals the most important risk structure of the instantaneous financial circumstance, while other tangent spaces contain trivial signals, market noise or even misleading information. We further find that the traditional covariance estimate disperses its spectral energy into other tangent spaces rather than concentrates it on the principal one (see Figure 1), thus leading to deterioration in SPO.

To better catch the instantaneous risk structure and improve robustness to downside risk in the rapidly-changing financial circumstance, we propose a loss control strategy with a rank-one covariance estimate for SPO (SPOLC). Our main contributions are:

1. Reform the covariance estimation problem from a sufficiently-sampled statistical estimation to a symmetric quadratic form operator.

2. Investigate the rank-related structure of $\mathbb{S}^d$, including the determinantal varieties, their tangent spaces at the spectral components of the observation matrix, and the corresponding orthogonal projections.

3. Propose a novel rank-one covariance estimate lying in the principal tangent space.

4. Propose a novel loss control scheme in the SPO model with the rank-one covariance estimate.

The rest of this paper is organized as follows. Section 2 gives the problem setting of SPO and some related works. Section 3 investigates the rank-related structure and the relevant tangent spaces of $\mathbb{S}^d$ in detail and establishes the loss control scheme with the rank-one covariance estimate. Section 4 conducts extensive experiments with comprehensive downside risk metrics and investing scores to assess SPOLC. Section 5 makes conclusions and discussions in the end.

2. Problem Setting and Related Works

In this section, we give the problem setting of SPO, and then show some main related works in this field.

2.1. Problem Setting of Short-term Portfolio Optimization

We adopt the standard SPO framework taken by many related works in machine learning (Cover, 1991; Li et al., 2015; Huang et al., 2016; Li et al., 2016; Lai et al., 2018c). Assume there are d assets in a financial market and their prices **at the end** of the t-th trading

period are represented by a vector $\mathbf{p}_t \in \mathbb{R}_+^d$, where $\mathbb{R}_+^d$ is the d-dimensional nonnegative real cone. As time passes $t = 0, 1, 2, \cdots$, the prices form a sequence $\mathbf{p}_0$, $\mathbf{p}_1$, $\mathbf{p}_2$, $\cdots$. Instead of the price, the price relative $\mathbf{x}_t \triangleq \frac{\mathbf{p}_t}{\mathbf{p}_{t-1}}$ is the assessment for the investing performance of assets, where the division dominates each element. $\mathbf{x}_t$ is closely related to the simple return $\mathbf{r}_t = \mathbf{x}_t - \mathbf{1}_d$ and they can be mutually converted ($\mathbf{1}_d$ denotes the d-dimensional vector with all elements equaling 1). In this paper, none of the $\mathbf{p}_t$, $\mathbf{x}_t$, and $\mathbf{r}_t$ are subject to any statistical distributions. They are collected from real-world financial markets.

Accordingly, we define a portfolio with the same dimensionality in the simplex as

$$\mathbf{b}_t \in \Delta_d := \{\mathbf{b} \in \mathbb{R}_+^d : \mathbf{1}_d^\top \mathbf{b} = 1\}. \tag{1}$$

This portfolio is self-financing ($\mathbf{1}_d^\top \mathbf{b} = 1$: no borrowing-money and full re-investment) and does not allow for short positions ($\mathbf{b}^{(i)} < 0$ means that the i-th asset has a short position), as is the most common case in real-world investments. Note that $\mathbf{b}_t$ is the proportion of the total wealth invested in different assets **throughout** the t-th period.

At the end of the t-th period, the cumulative wealth (CW) S_t increases or decreases with a factor $\mathbf{b}_t^\top \mathbf{x}_t$: $S_t = S_{t-1} \cdot (\mathbf{b}_t^\top \mathbf{x}_t)$. For simplicity, we can assume the initial wealth $S_0 = 1$ and the whole investment lasts for n periods. An SPO system aims to learn an appropriate sequence of portfolios $\{\hat{\mathbf{b}}_t\}_{t=1}^n$ such that the corresponding CW and return sequences

$$\hat{S}_t = \hat{S}_{t-1} \cdot (\hat{\mathbf{b}}_t^\top \mathbf{x}_t), \quad r_t = \hat{\mathbf{b}}_t^\top \mathbf{x}_t - 1, \quad t = 1, 2, \cdots, n, \tag{2}$$

achieve some particular financial objectives. To explain the above equations, suppose $\mathbf{x}_t = [0.9, 1, 1.1]^\top$, which means that on the t-th period the first stock decreases by 10%, the second stock remains the same, and the third stock increases by 10%. If we hold a portfolio $\hat{\mathbf{b}}_t = [1/3, 1/3, 1/3]^\top$, then the corresponding increasing factor would be $\hat{\mathbf{b}}_t^\top \mathbf{x}_t = 1$ and the return of the portfolio is $r_t = 0$. Thus the updated CW would be $\hat{S}_t = \hat{S}_{t-1} \cdot 1$, which means that there is no gain and no loss.

In order to control extreme losses and downside risk, we intend to make $\{\hat{S}_t\}_{t=1}^n$ and $\{r_t\}_{t=1}^n$ achieve low maximum drawdowns (Magdon-Ismail and Atiya, 2004) and high VaRs (Jorion, 1997), respectively. The definitions of maximum drawdown and VaR are given in (61) and (56). Furthermore, we aim to achieve better performance than some trivial strategies in gaining incomes. The first one is the Market strategy (Li et al., 2015) which disperses the wealth equally into all the assets at the very beginning and remains unchanged

$$\hat{S}_n^{Market} = \frac{1}{d} \sum_{i=1}^{d} \prod_{t=1}^{n} \mathbf{x}_t^{(i)}. \tag{3}$$

It allocates $1/d$ of the initial wealth to each asset i, then the wealth of each asset i grows along with its own price relatives $\prod_{t=1}^n \mathbf{x}_t^{(i)}$. At the end the wealths of all the assets are summed by $\sum_{i=1}^d$. Another trivial strategy is the $1/d$ strategy (DeMiguel et al., 2009), which keeps a constant portfolio with equal weights on all the assets at each rebalancing time:

$$\hat{\mathbf{b}}_t \equiv \frac{1}{d} \mathbf{1}_d, \quad \forall\, 1 \leqslant t \leqslant n. \tag{4}$$

2.2. Related Works on Mean-variance Methods

The mean-variance methods have been the most popular PO models since their first proposal by Markowitz (1952). Without loss of generality, we can formulate the original mean-variance (OMV) model as

$$\hat{\mathbf{b}} = \underset{\mathbf{b}\in\Delta_d}{\operatorname{argmin}} -\boldsymbol{\mu}^\top \mathbf{b} + \gamma \mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}, \tag{5}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the expected return and the covariance of the assets, respectively. Accordingly, $\boldsymbol{\mu}^\top \mathbf{b}$ and $\mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}$ are the expected return and the risk of the portfolio $\mathbf{b}$. The objective is to maximize the expected return and minimize the risk simultaneously with a mixing parameter $\gamma \geqslant 0$. It can be further modified for additional tasks.

For example, Sparse and Stable Markowitz Portfolio (SSMP, Brodie et al. 2009) introduces an ℓ^1 regularization

$$\begin{aligned}\hat{\mathbf{b}}_{SSMP} &= \underset{\mathbf{b}}{\operatorname{argmin}} \|\epsilon \mathbf{1}_n - \mathcal{R}\mathbf{b}\|^2 + \tau \|\mathbf{b}\|_1 \\ &= \underset{\mathbf{b}}{\operatorname{argmin}} -2\epsilon \mathbf{1}_n^\top \mathcal{R}\mathbf{b} + \mathbf{b}^\top \mathcal{R}^\top \mathcal{R}\mathbf{b} + n\epsilon^2 + \tau\|\mathbf{b}\|_1 \quad \text{s.t. } \mathbf{b}^\top \boldsymbol{\mu} = \epsilon, \mathbf{b}^\top \mathbf{1}_d = 1,\end{aligned} \tag{6}$$

where ϵ is a predefined expected growth rate, $\mathcal{R}$ is an $n \times d$-dimensional asset return matrix, $\|\cdot\|$ denotes the ℓ^2-norm, $\|\cdot\|_1$ denotes the ℓ^1-norm, and τ is the regularization strength. We notice that $2\epsilon \mathbf{1}_n^\top \mathcal{R}\mathbf{b}$ and $\mathbf{b}^\top \mathcal{R}^\top \mathcal{R}\mathbf{b}$ are essentially the expected return and the risk, respectively. Additionally, it relaxes the nonnegativity constraint of $\mathbf{b}$, forces $\mathbf{b}$ to be sparse and constrains the short position (Brodie et al., 2009).

Ho et al. (2015) further proposes a weighted elastic net penalized portfolio (WENPP) that removes the self-financing constraint $\mathbf{b}^\top \mathbf{1}_d = 1$ and adopts the elastic net penalization instead of the ℓ^1 regularization

$$\hat{\mathbf{b}}_{WENPP} = \underset{\mathbf{b}}{\operatorname{argmin}} -\mathbf{b}^\top \hat{\boldsymbol{\mu}} + \mathbf{b}^\top \hat{\boldsymbol{\Sigma}} \mathbf{b} + \sum_{i=1}^{d} \tau_i |\mathbf{b}^{(i)}| + \sum_{i=1}^{d} \iota_i |\mathbf{b}^{(i)}|^2, \tag{7}$$

where $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\mu}}$ are the estimated covariance and expectation of asset returns, respectively.

Both (6) and (7) take the form of (5), which makes a trade-off between return and risk. It cannot do without reliable estimations for $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ in order to achieve good investing performance. An intuitive approach is to introduce traditional statistical methods to make a single static estimation for $\hat{\boldsymbol{\mu}}$ or $\hat{\boldsymbol{\Sigma}}$. But it requires a large number of observations, which is nearly impossible in the SPO scenario. Neither is it adaptive to the rapidly-changing financial environment where the true $\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t$ may be dynamic.

Since the covariance term is critical in the Markowitz portfolio, it has been extensively investigated in the finance communities. For example, Ledoit and Wolf (2017) propose a Nonlinear Shrinkage of the Covariance Matrix (NSCM) $\hat{\boldsymbol{\Sigma}}_{NS}$ as follows:

$$\begin{aligned}&\hat{\boldsymbol{\Sigma}}_{UB} = \mathbf{Q}\boldsymbol{\Psi}\mathbf{Q}^\top, \quad \boldsymbol{\Psi} = \operatorname{diag}(\psi_1, \cdots, \psi_d). \\ &\forall\, 1 \leqslant i \leqslant d, \quad \tilde{\psi}_i \triangleq \begin{cases} \frac{1}{\psi_i |\tilde{\mathfrak{s}}(\psi_i)|^2} & \text{if } \psi_i > 0 \\ \frac{1}{\left(\frac{d}{w}-1\right)\tilde{\mathfrak{s}}(0)} & \text{if } \psi_i = 0, d > w \end{cases}. \\ &\hat{\boldsymbol{\Sigma}}_{NS} \triangleq \mathbf{Q}\tilde{\boldsymbol{\Psi}}\mathbf{Q}^\top, \quad \tilde{\boldsymbol{\Psi}} = \operatorname{diag}(\tilde{\psi}_1, \cdots, \tilde{\psi}_d).\end{aligned} \tag{8}$$

First, a spectral decomposition is performed on the traditional unbiased covariance $\hat{\mathbf{\Sigma}}_{UB}$ (see Eq. 17). Then each eigenvalue ψ_i is shrunk with an estimator of the (complex-valued) Stieltjes transform $\tilde{\mathfrak{s}}(\cdot)$ (Ledoit and Wolf, 2015). The resulting $\hat{\mathbf{\Sigma}}_{NS}$ is constructed by the same eigenvectors $\mathbf{Q}$ and the shrunk eigenvalues $\tilde{\mathbf{\Psi}}$. An advantage of $\hat{\mathbf{\Sigma}}_{NS}$ is that it allows for the singular case $d > w$ (more variables than observations). When combined with our loss control framework (51), NSCM shows good performance in the LPO scenario (Table 3). But it deteriorates in the SPO experiments, which indicates that a different kind of covariance estimate is needed in SPO.

2.3. Related Works on Short-term Portfolio Optimization

Since SPO systems have to adapt to the rapid change of financial environment, they turn to exploit some heuristic principles of empirical financial studies (Jegadeesh, 1990, 1991) and investing behaviors (Kahneman and Tversky, 1979; Bondt and Thaler, 1985; Jegadeesh and Titman, 1993). One main approach is the trend representation in the most recent observation window with size w.

OLMAR (Li et al., 2015) and RMR (Huang et al., 2016) are two typical mean-reversing strategies. The former uses the popular tool of moving average (MA) while the latter uses the ℓ^1-median (Vardi and Zhang, 2000) as future trends.

$$\hat{\mathbf{x}}_{t+1}^{OLMAR}(w) = \frac{MA_t(w)}{\mathbf{p}_t} = \frac{\sum_{k=0}^{w-1} \mathbf{p}_{t-k}}{w\mathbf{p}_t} = \frac{1}{w}\left(\mathbf{1}_d + \frac{\mathbf{1}_d}{\mathbf{x}_t} + \cdots + \frac{\mathbf{1}_d}{\bigotimes_{k=0}^{w-2} \mathbf{x}_{t-k}}\right). \tag{9}$$

$$\hat{\mathbf{x}}_{t+1}^{RMR} = \frac{\hat{\mathbf{p}}_{t+1}}{\mathbf{p}_t}, \quad \hat{\mathbf{p}}_{t+1} = \operatorname*{argmin}_{\mathbf{p}\in\mathbb{R}_+^d} \sum_{k=0}^{w-1} \|\mathbf{p}_{t-k} - \mathbf{p}\|. \tag{10}$$

The notation $\bigotimes$ is the element-wise chain-multiplication. The reason why (10) is called the ℓ^1-median is that when $\hat{\mathbf{p}}_{t+1}$, $\mathbf{p}_{t-k}$ and $\mathbf{p}$ are one-dimensional, the ℓ^2-norm in the argmin becomes an absolute value, then the sum in the argmin has a form of ℓ^1-norm with respect to the dimension k. Both OLMAR and RMR use the following model to optimize the portfolio

$$\mathbf{b}_{t+1} = \operatorname*{argmin}_{\mathbf{b}} \frac{1}{2}\|\mathbf{b} - \hat{\mathbf{b}}_t\|^2 \quad \text{s.t. } \mathbf{b}^\top \hat{\mathbf{x}}_{t+1} \geqslant \epsilon > 0. \quad \hat{\mathbf{b}}_{t+1} = \operatorname*{argmin}_{\mathbf{b}\in\Delta_d} \|\mathbf{b} - \mathbf{b}_{t+1}\|^2. \tag{11}$$

Since the $\mathbf{b}_{t+1}$ computed by the former equation in (11) could violate the portfolio constraint (1), it should be further normalized to the simplex (the latter equation) with the method of Duchi et al. (2008). This strategy tries to ensure an expected return $\mathbf{b}^\top \hat{\mathbf{x}}_{t+1} \geqslant \epsilon$ with a portfolio closest to the current one $\hat{\mathbf{b}}_t$.

On the other hand, the peak price tracking (PPT, Lai et al. 2018b) system is a trend-following strategy

$$\begin{aligned} \hat{\mathbf{x}}_{t+1}^{PPT} &= \frac{\hat{\mathbf{p}}_{t+1}}{\mathbf{p}_t}, \quad \hat{\mathbf{p}}_{t+1}^{(i)} = \max_{0\leqslant k\leqslant w-1} \mathbf{p}_{t-k}^{(i)}, \quad i = 1, 2, \cdots, d. \\ \hat{\mathbf{b}}_{t+1} &= \operatorname*{argmax}_{\mathbf{b}\in\Delta_d} \mathbf{b}^\top \hat{\mathbf{x}}_{t+1}^{PPT}, \quad \text{s.t. } \|\mathbf{b} - \hat{\mathbf{b}}_t\| \leqslant \epsilon, \epsilon > 0. \end{aligned} \tag{12}$$

It tries to maximize the increasing factor $\mathbf{b}^\top \hat{\mathbf{x}}_{t+1}^{PPT}$ with a portfolio in a neighborhood of $\hat{\mathbf{b}}_t$.

The adaptive input and composite trend representation system (AICTR, Lai et al. 2018a) further combines different trend types with a set of radius basis functions (RBF)

$$\hat{\mathbf{b}}_{t+1} = \underset{\mathbf{b}}{\operatorname{argmax}} \sum_{l=1}^{L} \phi_l \mathbf{b}^\top \hat{\mathbf{x}}_{l,t+1}, \quad \phi_l(\hat{\mathbf{x}}_{l,t+1}) = \exp\left(\frac{-\|\hat{\mathbf{x}}_{l,t+1} - \boldsymbol{\mu}_l\|^2}{2\sigma_l^2}\right),$$
$$\text{s.t.} \quad \mathbf{b} \in \Delta_d, \|\mathbf{b} - \hat{\mathbf{b}}_t\| \leqslant \epsilon, \epsilon > 0, \tag{13}$$

where $\{\hat{\mathbf{x}}_{l,t+1}\}_{l=1}^{L}$ denotes the L different trend representations. AICTR uses a generalized increasing factor with the RBFs adjusting the influence of different trend representations.

To establish a sparse portfolio in SPO, Lai et al. (2018c) simultaneously adopts an ℓ^1 regularization and a self-financing constraint in the increasing factor model

$$\hat{\mathbf{b}}_{t+1} = \min_{\mathbf{b}} \mathbf{b}^\top \boldsymbol{\varphi}_t + \lambda \|\mathbf{b}\|_1, \quad \text{s.t.}\, \mathbf{1}^\top \mathbf{b} = 1, \tag{14}$$

where $\lambda > 0$ controls the regularization strength and $\mathbf{b}^\top \boldsymbol{\varphi}_t$ is a negative increasing factor. (14) can be further formulated as an unconstrained augmented Lagrangian that has a saddle point, which can be solved by the ADMM.

Although the above SPO systems achieve state-of-the-art performance in gaining investing incomes, they are vulnerable to downside risk and extreme losses (see Section 4.3). The main reason is that the risk measurements in their models are ineffective. It is very difficult and challenging to estimate an appropriate $d \times d$-dimensional covariance matrix with a small window size w in the rapidly-changing financial environment, which will be explained and handled in the next section.

3. Loss Control with Rank-one Covariance Estimate

Suppose we have observed the price relatives for d assets in the nearest w trading periods at the current moment t. There are only a small window size w of observations that are sufficiently close to the current moment t and reliable to make estimations of the dynamic financial characteristics. These price relatives constitute an observation matrix

$$\mathbf{X} = [\mathbf{x}_{t-w+1}, \mathbf{x}_{t-w+2}, \cdots, \mathbf{x}_t]^\top \in \mathbb{R}^{w \times d}. \tag{15}$$

This matrix is almost all the information we could exploit to establish an SPO model in this section. We use four stages to find a suitable covariance estimate and design the loss control scheme.

3.1. Reform of Covariance Estimate as Operator

The first stage is to reform the function of a covariance estimate in the perspective of operators, and then find an adaptive operator in a suitable operator space.

3.1.1. Sufficiently-sampled Statistical Estimation

There are two traditional covariance estimates $\hat{\mathbf{\Sigma}}_{ML}$ and $\hat{\mathbf{\Sigma}}_{UB}$ in the practical statistical estimation

$$\hat{\mathbf{\Sigma}}_{MP} \triangleq \mathbf{X}^\top(\mathbf{I}_w - \frac{1}{w}\mathbf{1}_w\mathbf{1}_w^\top)\mathbf{X} \in \mathbb{S}^d, \tag{16}$$

$$\hat{\mathbf{\Sigma}}_{ML} \triangleq \frac{1}{w}\hat{\mathbf{\Sigma}}_{MP}, \quad \hat{\mathbf{\Sigma}}_{UB} \triangleq \frac{1}{w-1}\hat{\mathbf{\Sigma}}_{MP}, \tag{17}$$

where $\mathbf{I}_w$ and $\mathbf{1}_w$ denote the $w \times w$-dimensional identity matrix and the w-dimensional vector with all elements equaling 1, respectively. The two estimates have the same main part $\hat{\mathbf{\Sigma}}_{MP}$ but have different coefficients. Both estimates lie in the space of $d \times d$-dimensional symmetric matrices $\mathbb{S}^d$. $\hat{\mathbf{\Sigma}}_{ML}$ is the maximum-likelihood estimate while $\hat{\mathbf{\Sigma}}_{UB}$ is the unbiased estimate, if the rows of $\mathbf{X}$ are independent and identically distributed (i.i.d.) with the same d-dimensional normal distribution and $w > d$ (or else $\hat{\mathbf{\Sigma}}_{MP}$ would be singular since its rank is at most $\min\{w-1, d\}$). To summarize, $\hat{\mathbf{\Sigma}}_{ML}$ and $\hat{\mathbf{\Sigma}}_{UB}$ have statistical meaning only if $\mathbf{X}$ is sufficiently-sampled.

Unfortunately, the rows of $\mathbf{X}$ are not likely subject to a normal distribution (or other proper distributions) in the SPO scenario. Moreover, $w \ll d$ since we have to observe a considerable number of assets in a rather small window size that is sufficiently close to the current moment, in order to better catch the true dynamic covariance. Thus the assumptions of $\hat{\mathbf{\Sigma}}_{ML}$ and $\hat{\mathbf{\Sigma}}_{UB}$ are easily violated. Though (16) and (17) can still be used to compute $\hat{\mathbf{\Sigma}}_{ML}$ and $\hat{\mathbf{\Sigma}}_{UB}$ in this case, they may deteriorate in the OMV model (5).

3.1.2. Implicit Orthogonal Projection

These facts motivate us to redefine the function of such "covariance estimates". Dropping the statistical properties, we notice that the main part $\hat{\mathbf{\Sigma}}_{MP}$ is actually a composed operator with a projection followed by a Gram matrix.

$$\mathrm{P}_{\mathcal{K}} : \mathbb{R}^{w\times d} \mapsto \mathbb{R}^{w\times d}, \quad \mathrm{P}_{\mathcal{K}}(\mathbf{X}) \triangleq (\mathbf{I}_w - \frac{1}{w}\mathbf{1}_w\mathbf{1}_w^\top)\mathbf{X}; \tag{18}$$

$$\mathrm{g} : \mathbb{R}^{w\times d} \mapsto \mathbb{S}^d, \quad \mathrm{g}(\mathbf{Y}) \triangleq \mathbf{Y}^\top\mathbf{Y}; \tag{19}$$

$$\mathrm{g} \circ \mathrm{P}_{\mathcal{K}} : \mathbb{R}^{w\times d} \mapsto \mathbb{R}^{w\times d} \mapsto \mathbb{S}^d,$$

$$\hat{\mathbf{\Sigma}}_{MP} \triangleq \mathrm{g} \circ \mathrm{P}_{\mathcal{K}}(\mathbf{X}) = \mathbf{X}^\top(\mathbf{I}_w - \frac{1}{w}\mathbf{1}_w\mathbf{1}_w^\top)^2\mathbf{X} = \mathbf{X}^\top(\mathbf{I}_w - \frac{1}{w}\mathbf{1}_w\mathbf{1}_w^\top)\mathbf{X}. \tag{20}$$

The last equation holds since the projection $\mathrm{P}_{\mathcal{K}} = \mathbf{I}_w - \frac{1}{w}\mathbf{1}_w\mathbf{1}_w^\top$ is symmetric and idempotent.

Next, we have a closer look into $\mathrm{P}_{\mathcal{K}}$. It is an orthogonal projection because

$$\begin{aligned}
\mathrm{P}_{\mathcal{K}} \cdot (\mathbf{I}_w - \mathrm{P}_{\mathcal{K}}) &= (\mathbf{I}_w - \frac{1}{w}\mathbf{1}_w\mathbf{1}_w^\top) \cdot \frac{1}{w}\mathbf{1}_w\mathbf{1}_w^\top \\
&= \frac{1}{w}\mathbf{1}_w\mathbf{1}_w^\top - \frac{1}{w^2}\mathbf{1}_w(\mathbf{1}_w^\top\mathbf{1}_w)\mathbf{1}_w^\top \\
&= \frac{1}{w}\mathbf{1}_w\mathbf{1}_w^\top - \frac{1}{w}\mathbf{1}_w\mathbf{1}_w^\top = \boldsymbol{O}_{w\times w},
\end{aligned} \tag{21}$$

where $\boldsymbol{O}_{w\times w}$ denotes the $w \times w$ zero matrix. $\mathrm{P}_{\mathcal{K}}$ projects the elements of $\mathbb{R}^{w\times d}$ onto the linear subspace with centralized matrices

$$\mathcal{K} = \{\mathbf{Y} \in \mathbb{R}^{w\times d} : \mathbf{1}_w^\top \mathbf{Y} = \mathbf{0}_d^\top\}. \tag{22}$$

Given any $\mathbf{Y}_1$, $\mathbf{Y}_2 \in \mathcal{K}$ and any α_1, $\alpha_2 \in \mathbb{R}$, we have

$$\begin{aligned}\mathbf{1}_w^\top(\alpha_1\mathbf{Y}_1 + \alpha_2\mathbf{Y}_2) &= \alpha_1\mathbf{1}_w^\top\mathbf{Y}_1 + \alpha_2\mathbf{1}_w^\top\mathbf{Y}_2 \\ &= \alpha_1\mathbf{0}_d^\top + \alpha_2\mathbf{0}_d^\top = \mathbf{0}_d^\top.\end{aligned} \tag{23}$$

Hence $(\alpha_1\mathbf{Y}_1 + \alpha_2\mathbf{Y}_2) \in \mathcal{K}$ and $\mathcal{K}$ is a linear subspace of $\mathbb{R}^{w\times d}$. Since $\mathrm{P}_{\mathcal{K}}$ is orthogonal, it induces the following direct sum decomposition

$$\mathbb{R}^{w\times d} = \mathcal{K} \oplus \mathcal{K}^\perp, \tag{24}$$

where $\mathcal{K}^\perp$ is the range of $(\mathbf{I}_w - \mathrm{P}_{\mathcal{K}})$.

In the literature of probability, two random variables are uncorrelated if and only if their inner product in the integration (expectation) form equals zero. If we use the standard inner product for $\mathbb{R}^{w\times d}$

$$< \mathbf{X}, \mathbf{Y} > \triangleq \mathrm{tr}(\mathbf{X}^\top\mathbf{Y}), \tag{25}$$

then (24) indicates that the two subspaces $\mathcal{K}$ and $\mathcal{K}^\perp$ of the whole space $\mathbb{R}^{w\times d}$ are uncorrelated. To verify this statement, assume $\mathbf{Y}_1 \in \mathcal{K}$ and $\mathbf{Y}_2 \in \mathcal{K}^\perp$. According to the definitions of $\mathcal{K}$ and $\mathcal{K}^\perp$, there exist $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^{w\times d}$ such that $\mathbf{Y}_1 = \mathrm{P}_{\mathcal{K}}\mathbf{X}_1$ and $\mathbf{Y}_2 = (\mathbf{I}_w - \mathrm{P}_{\mathcal{K}})\mathbf{X}_2$. Then according to (21),

$$\mathrm{tr}(\mathbf{Y}_1^\top\mathbf{Y}_2) = \mathrm{tr}(\mathbf{X}_1^\top\mathrm{P}_{\mathcal{K}}(\mathbf{I}_w - \mathrm{P}_{\mathcal{K}})\mathbf{X}_2) = \mathrm{tr}(\mathbf{X}_1^\top\boldsymbol{O}_{w\times w}\mathbf{X}_2) = \mathrm{tr}(\boldsymbol{O}_{d\times d}) = 0. \tag{26}$$

The decomposition $\mathcal{K} \oplus \mathcal{K}^\perp$ helps to separate structural information. In this sense, $\hat{\boldsymbol{\Sigma}}_{MP}$ actually takes the structural information lying in $\mathcal{K}$ and drops that in $\mathcal{K}^\perp$.

However, the projection $\mathrm{P}_{\mathcal{K}}$ and the decomposition $\mathcal{K} \oplus \mathcal{K}^\perp$ do not exploit any specific information in the observation $\mathbf{X}$. Therefore, when the assumptions of covariance and the law of large numbers easily fail in SPO, $\mathrm{P}_{\mathcal{K}}$ and $\mathcal{K}\oplus\mathcal{K}^\perp$ cannot catch the instantaneous risk structure embedded in $\mathbf{X}$. It motivates us to find another suitable operator and another adaptive decomposition.

3.1.3. Symmetric Quadratic Form Operator

We notice that a covariance estimate $\hat{\boldsymbol{\Sigma}}$ can be seen as a symmetric quadratic form in $\mathbb{S}^d$, which measures the portfolio risk $\mathbf{b}^\top\hat{\boldsymbol{\Sigma}}\mathbf{b}$ when we manage the portfolio $\mathbf{b}$. Hence $\mathbb{S}^d$ is the very space we should investigate to explore the instantaneous risk structure.

1. $\mathbb{S}^d$ is a Hilbert space with the standard inner product $< \mathbf{X}, \mathbf{Y} > \triangleq \mathrm{tr}(\mathbf{XY})$, thus it is self-conjugate.
2. $\mathbb{S}^d$ is a ring with the standard matrix addition and multiplication.
3. Each element in $\mathbb{S}^d$ is a self-adjoint linear operator acting on $\mathbb{R}^d$.

No. 1 and 3 are obvious and we give more details about No. 2. First, $\mathbb{S}^d$ is an Abelian group with the standard matrix addition $+$ and the zero element is $\boldsymbol{O}_{d\times d}$. Second, $\mathbb{S}^d$ is a semi-group with the standard matrix multiplication $\cdot$, because the elements in $\mathbb{S}^d$ can be multiplied directly due to the same dimensionality of the rows and the columns, and $\mathbf{S}_1\cdot\mathbf{S}_2\cdot\mathbf{S}_3 = \mathbf{S}_1\cdot(\mathbf{S}_2\cdot\mathbf{S}_3), \forall \mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3 \in \mathbb{S}^d$. Third, the law of distribution holds: $\mathbf{S}_1\cdot(\mathbf{S}_2+\mathbf{S}_3) = \mathbf{S}_1\cdot\mathbf{S}_2+\mathbf{S}_1\cdot\mathbf{S}_3$. Thus $\mathbb{S}^d$ is a ring. These algebraic properties make it possible to investigate the structure of $\mathbb{S}^d$, especially with the specific information in $\mathbf{X}$.

In comparison, the space $\mathbb{R}^{w\times d}$ does not have the advantages of $\mathbb{S}^d$. It motivates us to find a new orthogonal decomposition of $\mathbb{S}^d$ instead of $\mathbb{R}^{w\times d}$. To do this, we switch the composing order of operators in (20):

$$\mathrm{P}\circ\mathrm{g}:\mathbb{R}^{w\times d}\mapsto\mathbb{S}^d\mapsto\mathbb{S}^d,\quad \mathrm{P}\circ\mathrm{g}(\mathbf{X})=\mathrm{P}(\mathbf{X}^\top\mathbf{X}). \tag{27}$$

The Gram matrix g goes first to map $\mathbf{X}$ to $\mathbb{S}^d$, then an orthogonal projection P acts on $\mathbb{S}^d$ to extract the risk structure.

The next step is to find an appropriate P. The focus comes back to the undersampled problem in SPO, where the number of observations w is much smaller than the number of assets d. It leads to $\mathrm{rank}(\mathbf{X}) \leqslant w \ll d$, which fails the maximum-likelihood or consistent covariance estimate. It shows a clue that we should start with the rank-related structure of $\mathbb{S}^d$.

3.2. Determinantal Variety and Its Tangent Spaces at Spectral Components

The second stage is to investigate the rank-related structure of $\mathbb{S}^d$ thoroughly, including the determinantal varieties, their tangent spaces, and the spectral components regarding the observation $\mathbf{X}$.

3.2.1. Determinantal Variety for $\mathbb{R}^{w\times d}$

Without loss of generality, we assume $\mathrm{rank}(\mathbf{X}) = w \leqslant d$ in the rest of this paper, since the observation $\mathbf{X}$ has full rank in most practical situations. If not, one could just replace w with the actual $\mathrm{rank}(\mathbf{X})$ and the conclusions still hold.

The rank $\mathfrak{r}$ of a matrix $\mathbf{X}$ is determined by the largest $\mathfrak{r}$ such that one of the $\mathfrak{r}$-minors (the determinants of the $\mathfrak{r}\times\mathfrak{r}$ submatrices allowing for index permutations) is nonzero. In other words, $\mathrm{rank}(\mathbf{X}) \leqslant \mathfrak{r}$ if and only if all the $(\mathfrak{r}+1)$-minors of $\mathbf{X}$ equal zero. Since a determinant is a polynomial of the elements in $\mathbf{X}$, the set $\mathscr{N}(w\times d,\mathfrak{r}) \triangleq \{\mathbf{X}\in\mathbb{R}^{w\times d} : \mathrm{rank}(\mathbf{X}) \leqslant \mathfrak{r}\}$ are the common roots of the polynomial equations defined by the $(\mathfrak{r}+1)$-minors. Hence, $\mathscr{N}(w\times d,\mathfrak{r})$ is an algebraic variety that has some special properties regarding its determinantal ideals. It helps to study some rank-related problems in $\mathbb{R}^{w\times d}$ as well as some tangent spaces at some specific matrix $\mathbf{X}$. Interested readers could refer to Király et al. (2015); Chandrasekaran et al. (2009); Bruns and Vetter (1988) for more details in this topic.

3.2.2. Determinantal Variety and Its Tangent Spaces for $\mathbb{S}^d$

Instead of $\mathbb{R}^{w\times d}$, we mainly investigate the rank-related structure of $\mathbb{S}^d$, in order to find an appropriate orthogonal projection P. Furthermore, we could exploit the information in a specific observation $\mathbf{X}$ to characterize $\mathbb{S}^d$, which is more adaptive to the current financial circumstance. We start with the following definition:

Definition 1 (Determinantal Variety for $\mathbb{S}^d$) *The determinantal variety of rank $\mathfrak{r}$ for $\mathbb{S}^d$ is*

$$\mathscr{M}(d,\mathfrak{r}) \triangleq \{\mathbf{M} \in \mathbb{S}^d : \operatorname{rank}(\mathbf{M}) \leqslant \mathfrak{r}\}. \tag{28}$$

It is well-defined since $\operatorname{rank}(\mathbf{M}) \leqslant \mathfrak{r}$ *if and only if all the* $(\mathfrak{r}+1)$*-minors of* $\mathbf{M}$ *equal zero (if* $\mathfrak{r} = d$*,* $\mathscr{M}(d,\mathfrak{r}) = \mathbb{S}^d$*).* $\mathscr{M}(d,\mathfrak{r})$ *consists of all the symmetric roots of the polynomial equations defined by the* $(\mathfrak{r}+1)$*-minors.*

Suppose we obtain some specific information matrix $\mathbf{S} \in \mathbb{S}^d$, let the spectral decomposition of $\mathbf{S}$ be

Definition 2 (Spectral Decomposition)

$$\mathbf{S} = \mathbf{U}\boldsymbol{\Theta}\mathbf{U}^\top \quad s.t.\, \mathbf{U} \in \mathbb{R}^{d\times d}, \mathbf{U}^\top\mathbf{U} = \mathbf{I}_d, \boldsymbol{\Theta} = \operatorname{diag}(\theta_1, \theta_2, \cdots, \theta_d). \tag{29}$$

Note that even if $\operatorname{rank}(\mathbf{S}) = \mathfrak{r} \leqslant d$*, one could still find a full orthonormal basis* $\mathbf{U}$ *for* $\mathbb{R}^d$*. Without loss of generality, we assume* $\theta_1 \geqslant \theta_2 \geqslant \cdots \geqslant \theta_d$*.*

The orthonormal basis $\mathbf{U}$ is a useful tool to represent $\mathbb{R}^d$ and $\mathbb{S}^d$ in the perspective of the observed information $\mathbf{S}$. We could further look into such a representation via the tangent spaces of the determinantal variety.

Theorem 1 *Let* $J \triangleq \{j : \theta_j \neq 0\}$ *be the index set of nonzero eigenvalues of* $\mathbf{S} \in \mathbb{S}^d$ *with* $|J| = \mathfrak{r}$*. Then the tangent space of the variety* $\mathscr{M}(d,\mathfrak{r})$ *at* $\mathbf{S}$ *can be characterized as*

$$\mathscr{T}(\mathbf{S}) = \{\mathbf{U}_J\mathbf{D}\mathbf{U}_J^\top : \mathbf{D} \in \mathbb{S}^{\mathfrak{r}}\}, \tag{30}$$

where $\mathbf{U}_J \in \mathbb{R}^{d\times\mathfrak{r}}$ *is the orthonormal set corresponding to the index set* J*.*

Proof $\mathscr{T}(\mathbf{S})$ is the largest linear subspace of $\mathbb{S}^d$ containing $\mathbf{S} = \mathbf{U}_J\boldsymbol{\Theta}_J\mathbf{U}_J^\top$ such that the rank of any matrix in $\mathscr{T}(\mathbf{S})$ is no larger than $\mathfrak{r}$. Since $\{\mathbf{U}_J\mathbf{D}\mathbf{U}_J^\top : \mathbf{D} \in \mathbb{S}^{\mathfrak{r}}\}$ is a linear space containing $\mathbf{S}$ and $\operatorname{rank}(\mathbf{U}_J\mathbf{D}\mathbf{U}_J^\top) \leqslant \mathfrak{r}$ for every $\mathbf{U}_J\mathbf{D}\mathbf{U}_J^\top$ in this space, we have $\{\mathbf{U}_J\mathbf{D}\mathbf{U}_J^\top : \mathbf{D} \in \mathbb{S}^{\mathfrak{r}}\} \subseteq \mathscr{T}(\mathbf{S})$. Besides, if $\mathfrak{r} = d$, then $\mathbf{U}_J = \mathbf{U}$ and $\{\mathbf{U}\mathbf{D}\mathbf{U}^\top : \mathbf{D} \in \mathbb{S}^d\}$ fills up the whole linear space $\mathscr{T}(\mathbf{S})$. Hence we just need to prove that

$$\mathscr{T}(\mathbf{S}) \subseteq \{\mathbf{U}_J\mathbf{D}\mathbf{U}_J^\top : \mathbf{D} \in \mathbb{S}^{\mathfrak{r}}\} \quad \text{when } \mathfrak{r} < d. \tag{31}$$

Given an arbitrary matrix $\mathbf{W} \in \mathscr{T}(\mathbf{S})$, let the spectral decomposition of $\mathbf{W}$ be $\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^\top$ with $\mathbf{V}$ and $\boldsymbol{\Lambda}$ similar to Definition 2. Let $K \triangleq \{k : \lambda_k \neq 0\}$ be the index set of nonzero eigenvalues of $\mathbf{W}$, then $\mathbf{W} = \mathbf{V}_K\boldsymbol{\Lambda}_K\mathbf{V}_K^\top$ and $|K| \leqslant \mathfrak{r}$. We first explain that $\mathbf{W}$ should lie in either the same row-space or the same column-space of $\mathbf{S}$. If not, there exist at least one eigenvector $\mathbf{v}_l \in \mathbb{R}^d$ in the orthonormal set $\mathbf{V}_K$ that is linearly independent of the other orthonormal set $\mathbf{U}_J$. Combining $\mathbf{v}_l$ with $\mathbf{U}_J$ forms a linearly independent set of rank $(\mathfrak{r}+1)$, which exceeds the rank constraint of $\mathscr{T}(\mathbf{S})$. To be specific, $\mathbf{v}_l\mathbf{v}_l^\top \in \mathscr{T}(\mathbf{S})$ because $\mathbf{W} \in \mathscr{T}(\mathbf{S})$, and $\mathbf{U}_J\mathbf{U}_J^\top \in \mathscr{T}(\mathbf{S})$ according to the above paragraph. Since $\mathscr{T}(\mathbf{S})$ is a linear space,

$$\mathbf{v}_l\mathbf{v}_l^\top + \mathbf{U}_J\mathbf{U}_J^\top \in \mathscr{T}(\mathbf{S}) \quad \text{but } \operatorname{rank}(\mathbf{v}_l\mathbf{v}_l^\top + \mathbf{U}_J\mathbf{U}_J^\top) = \mathfrak{r}+1. \tag{32}$$

It violates the rank constraint of $\mathscr{T}(\mathbf{S})$ and leads to a contradiction.

Therefore, $\mathbf{W}$ should lie in either the same row-space or the same column-space of $\mathbf{S}$. In either case, we could let $\mathbf{V}_K = \mathbf{U}_J\mathbf{C}$ for some linear transform $\mathbf{C} \in \mathbb{R}^{\mathfrak{r}\times|K|}$ and represent $\mathbf{W}$ via $\mathbf{U}_J$

$$\mathbf{W} = \mathbf{V}_K\mathbf{\Lambda}_K\mathbf{V}_K^\top = \mathbf{U}_J(\mathbf{C}\mathbf{\Lambda}_K\mathbf{C}^\top)\mathbf{U}_J^\top, \quad \mathbf{C}\mathbf{\Lambda}_K\mathbf{C}^\top \in \mathbb{S}^{\mathfrak{r}}. \tag{33}$$

Thus $\mathbf{W} \in \{\mathbf{U}_J\mathbf{D}\mathbf{U}_J^\top : \mathbf{D} \in \mathbb{S}^{\mathfrak{r}}\}$. Now we have proven that $\forall \mathbf{W} \in \mathscr{T}(\mathbf{S}) \Rightarrow \mathbf{W} \in \{\mathbf{U}_J\mathbf{D}\mathbf{U}_J^\top : \mathbf{D} \in \mathbb{S}^{\mathfrak{r}}\}$, thus (31) has been proven. Together with the first paragraph of this proof, the whole theorem has been proven. ■

For convenience, we need not state the underlying nonsingular variety $\mathscr{M}(d, \mathfrak{r})$ for a rank$\leqslant \mathfrak{r}$ tangent space in the rest of this paper. Based on Theorem 1, the following theorem further indicates that two tangent spaces at two nonoverlapping sets of spectral components are mutually orthogonal.

Theorem 2 *Denote two arbitrary nonoverlapping index sets by $J_1, J_2 \subseteq \{1, 2, \cdots, d\}$ and $J_1 \cap J_2 = \emptyset$. Let $\mathbf{U}_{J_1}$ and $\mathbf{U}_{J_2}$ be two nonoverlapping orthonormal sets regarding the spectral decomposition of $\mathbf{S}$. Then $\mathscr{T}(\mathbf{U}_{J_1}\mathbf{U}_{J_1}^\top) \perp \mathscr{T}(\mathbf{U}_{J_2}\mathbf{U}_{J_2}^\top)$ in the following two senses:*

1. *Consider $\mathscr{T}(\mathbf{U}_{J_1}\mathbf{U}_{J_1}^\top)$ and $\mathscr{T}(\mathbf{U}_{J_2}\mathbf{U}_{J_2}^\top)$ as operator spaces of linear mappings for $\mathbb{R}^d \mapsto \mathbb{R}^d$.*

2. *Consider $\mathscr{T}(\mathbf{U}_{J_1}\mathbf{U}_{J_1}^\top)$ and $\mathscr{T}(\mathbf{U}_{J_2}\mathbf{U}_{J_2}^\top)$ as linear subspaces of the Hilbert space $\mathbb{S}^d$ with its standard inner product.*

Proof Given two arbitrary matrices $\mathbf{W}_1 \in \mathscr{T}(\mathbf{U}_{J_1}\mathbf{U}_{J_1}^\top)$ and $\mathbf{W}_2 \in \mathscr{T}(\mathbf{U}_{J_2}\mathbf{U}_{J_2}^\top)$, we could characterize them as $\mathbf{W}_1 = \mathbf{U}_{J_1}\mathbf{D}_1\mathbf{U}_{J_1}^\top$ and $\mathbf{W}_2 = \mathbf{U}_{J_2}\mathbf{D}_2\mathbf{U}_{J_2}^\top$ based on Theorem 1. Note that $\text{rank}(\mathbf{W}_1)$ and $\text{rank}(\mathbf{W}_2)$ need not be the same. Composing $\mathbf{W}_1$ and $\mathbf{W}_2$ leads to

$$\begin{aligned}\mathbf{W}_1\mathbf{W}_2 &= \mathbf{U}_{J_1}\mathbf{D}_1(\mathbf{U}_{J_1}^\top\mathbf{U}_{J_2})\mathbf{D}_2\mathbf{U}_{J_2}^\top \\ &= \mathbf{U}_{J_1}\mathbf{D}_1\boldsymbol{O}_{|J_1|\times|J_2|}\mathbf{D}_2\mathbf{U}_{J_2}^\top \\ &= \boldsymbol{O}_{d\times d}.\end{aligned} \tag{34}$$

Since $\mathbf{U}_{J_1}$ and $\mathbf{U}_{J_2}$ are mutually orthogonal, $\mathbf{U}_{J_1}^\top\mathbf{U}_{J_2}$ vanishes to the zero matrix $\boldsymbol{O}_{|J_1|\times|J_2|}$. It results in the following two conclusions.

1. The composition $\mathbf{W}_1\mathbf{W}_2 = \boldsymbol{O}_{d\times d}$ is exactly a matrix multiplication. If we consider this multiplication as an operator inner product with $\boldsymbol{O}_{d\times d}$ being the zero element, we have $\mathbf{W}_1 \perp \mathbf{W}_2$. For the arbitrariness of $\mathbf{W}_1$ and $\mathbf{W}_2$, $\mathscr{T}(\mathbf{U}_{J_1}\mathbf{U}_{J_1}^\top) \perp \mathscr{T}(\mathbf{U}_{J_2}\mathbf{U}_{J_2}^\top)$ in the sense of operator spaces.

2. Computing the standard inner product of $\mathbf{W}_1$ and $\mathbf{W}_2$ in $\mathbb{S}^d$ leads to $\text{tr}(\mathbf{W}_1\mathbf{W}_2) = \text{tr}(\boldsymbol{O}_{d\times d}) = 0$. Hence $\mathscr{T}(\mathbf{U}_{J_1}\mathbf{U}_{J_1}^\top) \perp \mathscr{T}(\mathbf{U}_{J_2}\mathbf{U}_{J_2}^\top)$ in the sense of linear subspaces of $\mathbb{S}^d$.

■

Theorem 2 can be easily extended to the case of multiple tangent spaces. This property induces an orthogonal projection from $\mathbb{S}^d$ to a tangent space $\mathscr{T}(\mathbf{U}_K\mathbf{U}_K^\top)$.

Theorem 3 *Let $K \subseteq \{1, 2, \cdots, d\}$ be an arbitrary index set and denote $\mathscr{T}_K \triangleq \mathscr{T}(\mathbf{U}_K\mathbf{U}_K^\top)$ as the tangent space at the spectral component $\mathbf{U}_K\mathbf{U}_K^\top$ of $\mathbf{S}$. Then*

$$\mathrm{P}_K : \mathbb{S}^d \mapsto \mathscr{T}_K, \qquad \mathrm{P}_K(\mathbf{W}) \triangleq \mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top \tag{35}$$

defines an orthogonal projection from $\mathbb{S}^d$ to $\mathscr{T}_K$.

Proof First, we verify that P_K is a projection:

$$\begin{aligned}
\mathrm{P}_K \circ \mathrm{P}_K(\mathbf{W}) &= \mathbf{U}_K\mathbf{U}_K^\top(\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top)\mathbf{U}_K\mathbf{U}_K^\top \\
&= \mathbf{U}_K(\mathbf{U}_K^\top\mathbf{U}_K)\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K(\mathbf{U}_K^\top\mathbf{U}_K)\mathbf{U}_K^\top \\
&= \mathbf{U}_K\mathbf{I}_{|K|}\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{I}_{|K|}\mathbf{U}_K^\top \\
&= \mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top = \mathrm{P}_K(\mathbf{W}), \quad \forall\, \mathbf{W} \in \mathbb{S}^d. \qquad (36) \\
\mathrm{P}_K(\eta_1\mathbf{W}_1 + \eta_2\mathbf{W}_2) &= \mathbf{U}_K\mathbf{U}_K^\top(\eta_1\mathbf{W}_1 + \eta_2\mathbf{W}_2)\mathbf{U}_K\mathbf{U}_K^\top \\
&= \eta_1\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}_1\mathbf{U}_K\mathbf{U}_K^\top + \eta_2\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}_2\mathbf{U}_K\mathbf{U}_K^\top \\
&= \eta_1\mathrm{P}_K(\mathbf{W}_1) + \eta_2\mathrm{P}_K(\mathbf{W}_2), \quad \forall\, \mathbf{W}_1, \mathbf{W}_2 \in \mathbb{S}^d,\ \forall\, \eta_1, \eta_2 \in \mathbb{R}. \qquad (37)
\end{aligned}$$

Hence $\mathrm{P}_K \circ \mathrm{P}_K = \mathrm{P}_K$ and P_K is a projection.

To further prove that P_K is orthogonal, we decompose $\forall\, \mathbf{W} \in \mathbb{S}^d$ into two components

$$\begin{aligned}
\mathbf{W} &= \mathrm{P}_K(\mathbf{W}) + (\mathrm{I} - \mathrm{P}_K)(\mathbf{W}) \\
&= (\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top) + (\mathbf{W} - \mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top),
\end{aligned} \tag{38}$$

where I denotes the identity mapping. If $\mathrm{P}_K(\mathbf{W}) \perp (\mathrm{I} - \mathrm{P}_K)(\mathbf{W})$, then P_K is orthogonal. It can be verified by computing their standard inner product in $\mathbb{S}^d$:

$$\begin{aligned}
&\mathrm{tr}((\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top)(\mathbf{W} - \mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top)) \\
=&\,\mathrm{tr}(\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}) - \mathrm{tr}(\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K(\mathbf{U}_K^\top\mathbf{U}_K)\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top) \\
=&\,\mathrm{tr}(\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}) - \mathrm{tr}(\underbrace{\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K}_{\mathbf{A}}\underbrace{\mathbf{U}_K^\top}_{\mathbf{B}}) \\
=&\,\mathrm{tr}(\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}) - \mathrm{tr}((\mathbf{U}_K^\top\mathbf{U}_K)\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K) \\
=&\,\mathrm{tr}(\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}) - \mathrm{tr}(\underbrace{\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}}_{\mathbf{A}}\underbrace{\mathbf{U}_K}_{\mathbf{B}}) \\
=&\,\mathrm{tr}(\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}) - \mathrm{tr}(\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}\mathbf{U}_K\mathbf{U}_K^\top\mathbf{W}) = 0.
\end{aligned} \tag{39}$$

We use the fact $\mathbf{U}_K^\top\mathbf{U}_K = \mathbf{I}_{|K|}$ in the second and the fourth equations, and the fact $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ in the third and the fifth equations. Since the above inner product is zero, P_K is orthogonal. ■

Some notes: 1. Now we have already established an orthogonal projection P_K and a direct sum decomposition $\mathscr{T}_K \oplus \mathscr{T}_K^{\perp} = \mathbb{S}^d$ for $\mathbb{S}^d$. In fact, if $J \cap K = \emptyset$, then $\mathscr{T}_J$ and $\mathscr{T}_K$ are uncorrelated in the sense of orthogonality, which is similar to (24) and (26). This property helps to separate the structural information of $\mathbb{S}^d$ into uncorrelated components and extract the most significant component. Moreover, P_K and $\mathscr{T}_K$ are adaptive to the current observation $\mathbf{X}$ since they exploit the information therein.

2. An eligible covariance estimate is not only symmetric but also positive semidefinite. Thus from (27), we could use $\mathbf{S} = \mathbf{X}^\top \mathbf{X} \in \mathbb{S}_+^d$ to characterize $\mathbb{S}^d$ and $\mathscr{T}(\mathbf{S})$ in (29) and (30), where $\mathbb{S}_+^d$ denotes the cone of positive semidefinite symmetric matrices. In this case, $\mathbf{U}_J$ simply consists of the first $\mathfrak{r} = \mathrm{rank}(\mathbf{S})$ eigenvectors and $\theta_{\mathfrak{r}+1} = \theta_{\mathfrak{r}+2} = \cdots = \theta_d = 0$.

3.3. Rank-one Covariance Estimate

The third stage is to find the most important tangent space among all the tangent spaces in $\mathbb{S}^d$, and then find a suitable covariance estimate therein.

3.3.1. Rank-one Tangent Spaces at Spectral Components

Suppose we use the Gram matrix $\mathbf{S} = \mathbf{X}^\top \mathbf{X} \in \mathbb{S}_+^d$ with $\mathrm{rank}(\mathbf{S}) = \mathrm{rank}(\mathbf{X}) = w$. Its spectral decomposition is illustrated by the above notes. We further discuss a special type of tangent spaces: the rank-one tangent space at each spectral component

$$\mathscr{T}_i \triangleq \mathscr{T}(\mathbf{u}_i \mathbf{u}_i^\top), \quad \mathbf{u}_i \text{ is the } i\text{-th eigenvector of } \mathbf{S}, \quad i = 1, 2, \cdots, d. \tag{40}$$

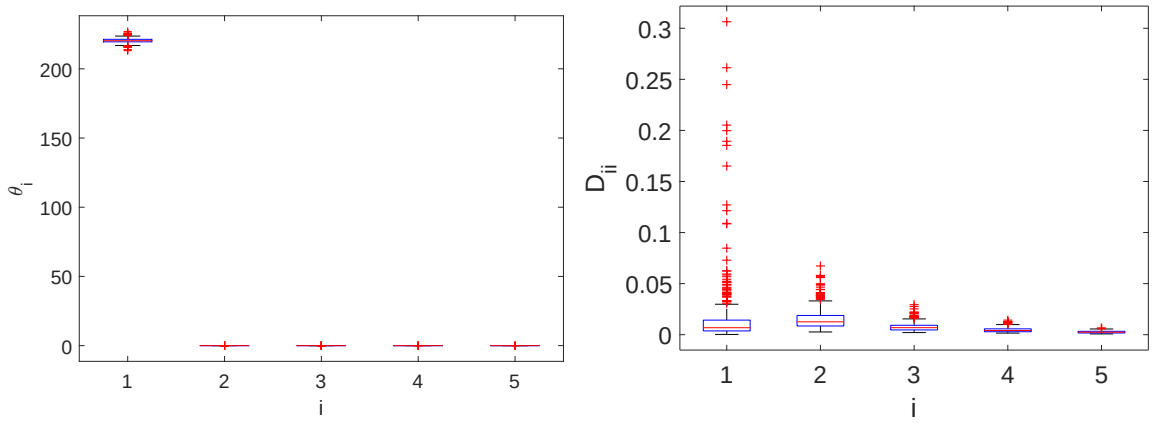
The orthogonal projection induced by $\mathscr{T}_i$ in Theorem 3 is denoted by P_i. $\{\mathscr{T}_i\}_{i=1}^d$ are mutually uncorrelated in the sense of orthogonality.

As explained in Section 3.1.3, a covariance estimate $\hat{\mathbf{\Sigma}}$ can be seen as a quadratic form in $\mathbb{S}^d$. From the perspective of operators, $\{\mathbf{u}_i \mathbf{u}_i^\top\}_{i=1}^w$ are the main components of $\mathbf{S}$ that reveal the instantaneous risk structure in $\mathbb{S}^d$ for the current financial circumstance. In these main components, $\mathbf{u}_1 \mathbf{u}_1^\top$ is the principal component that carries the largest spectral value θ_1 among all the components. Actually θ_1 is much larger than other θ_is in SPO situations. For example, Figure 1a shows the box plots of $\{\theta_i\}_{i=1}^w$ ($w = 5$) for the observations of the HS300 data set (Lai et al., 2018a). It indicates that $\theta_1 > 200$ in all the observations while $\theta_2 \sim \theta_5$ are rather small. The average spectral values are

$$[\theta_1, \theta_2, \cdots, \theta_5] = [220.3898, 0.0153, 0.0075, 0.0045, 0.0026]. \tag{41}$$

Thus $\mathscr{T}_1$ possesses the most important risk structure while other $\mathscr{T}_i$s contain trivial signals, market noise or even misleading information.

We further investigate the spectral values of the traditional covariance estimate $\hat{\mathbf{\Sigma}}_{MP}$ at $\{\mathbf{u}_i \mathbf{u}_i^\top\}_{i=1}^w$ to see their differences from $\{\theta_i\}_{i=1}^w$. We implement the singular value decomposition (SVD) on $\mathbf{X} = \mathbf{V}_J \Xi \mathbf{U}_J^\top$, where the right singular vector $\mathbf{U}_J$ is exactly the



(a) Spectral values of $\mathbf{S} = \mathbf{X}^\top \mathbf{X}$

(b) Spectral values of $\hat{\mathbf{\Sigma}}_{MP} = \mathbf{X}^\top (\mathbf{I}_w - \frac{1}{w}\mathbf{1}_w \mathbf{1}_w^\top)\mathbf{X}$

Figure 1: Box plots of spectral values at $\{\mathbf{u}_i \mathbf{u}_i^\top\}_{i=1}^w$ in an SPO experiment ($w = 5$). (a) θ_1 is much larger than other $\{\theta_i\}_{i\neq 1}$, thus $\mathscr{T}_1$ possesses the most important risk structure. (b) The spectral energy is dispersed into $\mathbf{D}_{11} \sim \mathbf{D}_{55}$, which weakens the representation via $\mathscr{T}_1$.

orthonormal set of $\mathbf{S} = \mathbf{X}^\top \mathbf{X}$. Based on Theorem 1, we observe that

$$\begin{aligned} \hat{\mathbf{\Sigma}}_{MP} = \mathrm{g} \circ \mathrm{P}_{\mathcal{K}}(\mathbf{X}) &= \mathbf{X}^\top (\mathbf{I}_w - \frac{1}{w}\mathbf{1}_w \mathbf{1}_w^\top)\mathbf{X} \\ &= \mathbf{U}_J (\Xi^2 - \frac{1}{w}\Xi \mathbf{V}_J^\top \mathbf{1}_w \mathbf{1}_w^\top \mathbf{V}_J \Xi)\mathbf{U}_J^\top \\ &\triangleq \mathbf{U}_J \mathbf{D} \mathbf{U}_J^\top \in \mathscr{T}(\mathbf{S}) \cap \mathbb{S}_+^d. \end{aligned} \tag{42}$$

Thus we could use the same spectral components w.r.t. $\mathbf{S}$ to represent $\hat{\mathbf{\Sigma}}_{MP}$. To keep a simple notation and avoid ambiguity as well, we use $\mathbf{D}$ particularly for the representation of $\hat{\mathbf{\Sigma}}_{MP}$ in the rest of this paper. The diagonal elements $\{\mathbf{D}_{ii}\}_{i=1}^w$ of $\mathbf{D}$ are the corresponding spectral values at the spectral components $\{\mathbf{u}_i \mathbf{u}_i^\top\}_{i=1}^w$. We conduct the same experiment as the above paragraph to compute $\{\mathbf{D}_{ii}\}_{i=1}^w$ and show their box plots in Figure 1b. Different from Figure 1a, the gap between $\{\mathbf{D}_{11}\}$ and other $\{\mathbf{D}_{ii}\}_{i\neq 1}$ is much smaller in Figure 1b. In fact, the average spectral values are

$$[\mathbf{D}_{11}, \mathbf{D}_{22}, \cdots, \mathbf{D}_{55}] = [0.0167, 0.0153, 0.0075, 0.0045, 0.0026], \tag{43}$$

which indicates that the spectral energy is dispersed into $\mathbf{D}_{11} \sim \mathbf{D}_{55}$. Hence $\hat{\mathbf{\Sigma}}_{MP}$ allocates a considerable proportion of spectral energy to the minor spectral components $\{\mathbf{u}_i \mathbf{u}_i^\top\}_{i\neq 1}$, which weakens the representation via $\mathscr{T}_1$. Experiments in Section 4.2.2 will further demonstrate that $\hat{\mathbf{\Sigma}}_{MP}$ fails to catch the instantaneous principal risk structure and thus deteriorates in investing performance.

3.3.2. Rank-one Covariance Estimate in $\mathscr{T}_1$

Based on the above discoveries, the Gram matrix $\mathbf{S} = \mathbf{X}^\top \mathbf{X}$ seems to be a good substitute for the traditional covariance $\hat{\mathbf{\Sigma}}_{MP}$. But there are at least two drawbacks in which $\mathbf{S}$ needs further improvements:

1. It does not extract any asset volatility information from $\hat{\mathbf{\Sigma}}_{MP}$.
2. Though it allocates most spectral energy to the principal rank-one tangent space $\mathscr{T}_1$, it still has some energy in the minor tangent spaces $\{\mathscr{T}_i\}_{i\neq 1}$.

As for the first point, $\hat{\mathbf{\Sigma}}_{MP}$ has a mathematical expression that measures asset volatility, which could be useful in setting the magnitude of covariance estimate. To explain the second point, we give an example

$$\hat{\mathbf{\Sigma}} = \mathbf{u}_1 \zeta_1 \mathbf{u}_1^\top + \mathbf{u}_j \zeta_j \mathbf{u}_j^\top, \quad j \neq 1, \zeta_j > 0. \tag{44}$$

Then if the portfolio happens to be $\mathbf{b} = \eta \mathbf{u}_j$ ($\eta \neq 0$), the representation via $\mathscr{T}_1$ will vanish in the risk term

$$\mathbf{b}^\top \hat{\mathbf{\Sigma}} \mathbf{b} = \zeta_j \eta^2, \tag{45}$$

no matter how large ζ_1 is, since $\mathscr{T}_1 \perp \mathscr{T}_j$. In this case, $\mathbf{b}^\top \hat{\mathbf{\Sigma}} \mathbf{b}$ is mostly market noise that is detrimental to the mean-variance model. In fact, this problem could appear when $\mathbf{b} \approx \eta \mathbf{u}_j$. Even if $\mathbf{S}$ were useful, it cannot do without the detailed analyses of the rank-related structure of $\mathbb{S}^d$ in this whole section. Experiments in Section 4.2.3 would further show that $\mathbf{S}$ is not so effective as the rank-one covariance estimate $\hat{\mathbf{\Sigma}}_{RO}$ to be proposed below.

Therefore, we eliminate all the minor spectrums in $\{\mathscr{T}_i\}_{i\neq 1}$ and only keep the spectrum in $\mathscr{T}_1$. It leads to the rank-one covariance estimate

$$\hat{\mathbf{\Sigma}}_{RO} = \mathbf{u}_1 \zeta_1 \mathbf{u}_1^\top \in \mathscr{T}_1 \cap \mathbb{S}_+^d. \tag{46}$$

$\mathscr{T}_1$ has already determined almost the whole $\hat{\mathbf{\Sigma}}_{RO}$. The only issue left to be addressed is the magnitude of the spectral energy ζ_1, which could be set with the asset volatility information from $\hat{\mathbf{\Sigma}}_{MP}$. Since no more assumptions on the price relatives or the covariance estimate are made, we provide a heuristic and empirical way to find ζ_1, which shows good performance in reducing downside risk and improving investing return for SPO.

First, we define a magnitude function of a symmetric matrix $\mathbf{\Sigma} \in \mathbb{S}^d$ as follows

$$\rho(\mathbf{\Sigma}) = \frac{\|\mathbf{\Sigma}\|_*}{d \cdot \operatorname{rank}(\mathbf{\Sigma})}. \tag{47}$$

We use the nuclear norm since it has a simple calculation $\|\cdot\|_* = \operatorname{tr}(\cdot)$ in $\mathbb{S}^d$. Besides, we divide $\|\mathbf{\Sigma}\|_*$ by $d \cdot \operatorname{rank}(\mathbf{\Sigma})$ as a normalization regarding the dimensionality and the rank of the matrix. Then we set ζ_1 as a trade-off between the magnitude functions of the two covariance estimates $\mathbf{u}_1 \theta_1 \mathbf{u}_1^\top = \mathrm{P}_1 \circ \mathrm{g}(\mathbf{X})$ and $\hat{\mathbf{\Sigma}}_{MP} = \mathrm{g} \circ \mathrm{P}_{\mathcal{K}}(\mathbf{X})$

$$\zeta_1^* = \operatorname*{argmin}_{\zeta_1 > 0} \frac{\rho^2(\mathbf{u}_1 \theta_1 \mathbf{u}_1^\top)}{\rho(\hat{\mathbf{\Sigma}}_{RO})} + \rho(\hat{\mathbf{\Sigma}}_{RO}) \cdot \rho(\hat{\mathbf{\Sigma}}_{MP}). \tag{48}$$

When $\rho(\hat{\mathbf{\Sigma}}_{RO})$ increases, the first term decreases while the second term increases, and vice versa. We square $\rho(\mathbf{u}_1\theta_1\mathbf{u}_1^\top)$ to amplify its influence because $\mathbf{u}_1\theta_1\mathbf{u}_1^\top$ carries the spectrum regarding $\mathbf{X}$ in the principal tangent space $\mathscr{T}_1$. Then (48) reaches its minimum when

$$\begin{aligned}
\frac{\rho^2(\mathbf{u}_1\theta_1\mathbf{u}_1^\top)}{\rho(\hat{\mathbf{\Sigma}}_{RO})} &= \rho(\hat{\mathbf{\Sigma}}_{RO})\cdot\rho(\hat{\mathbf{\Sigma}}_{MP}),\\
\rho(\hat{\mathbf{\Sigma}}_{RO}) &= \frac{\rho(\mathbf{u}_1\theta_1\mathbf{u}_1^\top)}{\sqrt{\rho(\hat{\mathbf{\Sigma}}_{MP})}},\\
\frac{\|\mathbf{u}_1\zeta_1^*\mathbf{u}_1^\top\|_*}{d\cdot 1} &= \frac{\|\mathbf{u}_1\theta_1\mathbf{u}_1^\top\|_*}{d\cdot 1}\Big/\sqrt{\frac{\|\hat{\mathbf{\Sigma}}_{MP}\|_*}{d\cdot(w-1)}},\\
\mathrm{tr}(\mathbf{u}_1\zeta_1^*\mathbf{u}_1^\top) &= \mathrm{tr}(\mathbf{u}_1\theta_1\mathbf{u}_1^\top)\Big/\sqrt{\frac{\mathrm{tr}(\mathbf{U}_J\mathbf{D}\mathbf{U}_J^\top)}{d\cdot(w-1)}},\\
\mathrm{tr}(\mathbf{u}_1^\top\mathbf{u}_1\zeta_1^*) &= \mathrm{tr}(\mathbf{u}_1^\top\mathbf{u}_1\theta_1)\Big/\sqrt{\frac{\mathrm{tr}(\mathbf{U}_J^\top\mathbf{U}_J\mathbf{D})}{d\cdot(w-1)}},\\
\zeta_1^* &= \theta_1\left(\frac{\mathrm{tr}(\mathbf{D})}{d(w-1)}\right)^{-\frac{1}{2}}.
\end{aligned} \tag{49}$$

In the rest of this paper, we use $\hat{\mathbf{\Sigma}}_{RO} = \mathbf{u}_1\zeta_1^*\mathbf{u}_1^\top$ as the covariance estimate for our method.

Figure 2 shows how to construct $\hat{\mathbf{\Sigma}}_{RO}$ via tangent spaces. $\mathscr{T}_J$ contains $\hat{\mathbf{\Sigma}}_{MP}$ and the principal rank-one tangent space $\mathscr{T}_1$, where $\mathscr{T}_1$ reveals the main instantaneous risk structure of the current financial circumstance. At first, $\mathbf{S}$ is orthogonally projected onto $\mathscr{T}_1$ as $\mathbf{u}_1\theta_1\mathbf{u}_1^\top$. Then a point in $\mathscr{T}_1$ is found as a trade-off between $\mathbf{u}_1\theta_1\mathbf{u}_1^\top$ and $\hat{\mathbf{\Sigma}}_{MP}$ according to (49), which is the desired $\hat{\mathbf{\Sigma}}_{RO}$.

3.4. Short-term Portfolio Optimization with Loss Control

The fourth stage is to establish a loss control scheme in the SPO model with the rank-one covariance estimate $\hat{\mathbf{\Sigma}}_{RO}$. Recall the observation matrix $\mathbf{X}$ in (15), we further denote $\mathbf{X}_i$ $(1 \leqslant i \leqslant w)$ as the i-th row of $\mathbf{X}$.

For a portfolio $\mathbf{b} \in \Delta_d$, $\mathbf{X}_i\mathbf{b}$ is the increasing factor of the i-th trading period. Furthermore, $\min_{1\leqslant i\leqslant w}\mathbf{X}_i\mathbf{b}$ is the worst performance in the recent time window if we use $\mathbf{b}$ as the portfolio. Recall that there are only a small window size w of observations that are sufficiently close to the current moment t and reliable to estimate the extreme loss in the dynamic financial circumstance. To control loss and downside risk, we propose to optimize $\mathbf{b}$ such that the worst performance is maximized

$$\hat{\mathbf{b}} = \underset{\mathbf{b}\in\Delta_d}{\mathrm{argmax}}\ \min_{1\leqslant i\leqslant w}\mathbf{X}_i\mathbf{b}. \tag{50}$$

This strategy can be interpreted by Figure 3. Note that as $\mathbf{b}$ changes, the worst-performance index i may have a "hard shift" to a different index j.

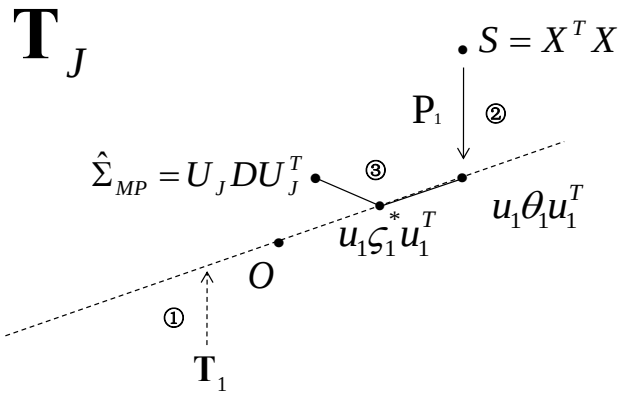


Figure 2: Interpretation of the proposed rank-one covariance estimate $\hat{\mathbf{\Sigma}}_{RO}$ via tangent spaces. The universe $\mathscr{T}_J$ is the tangent space of $\mathscr{M}(d, w)$ at the Gram matrix $\mathbf{S} = \mathbf{X}^\top \mathbf{X}$. ①Find the principal rank-one tangent space $\mathscr{T}_1$ (the dashed line). ②Project $\mathbf{S}$ orthogonally onto $\mathscr{T}_1$ as $\mathbf{u}_1 \theta_1 \mathbf{u}_1^\top$. ③Find a trade-off point $\hat{\mathbf{\Sigma}}_{RO} = \mathbf{u}_1 \zeta_1^* \mathbf{u}_1^\top$ between $\mathbf{u}_1 \theta_1 \mathbf{u}_1^\top$ and $\hat{\mathbf{\Sigma}}_{MP}$ in $\mathscr{T}_1$.

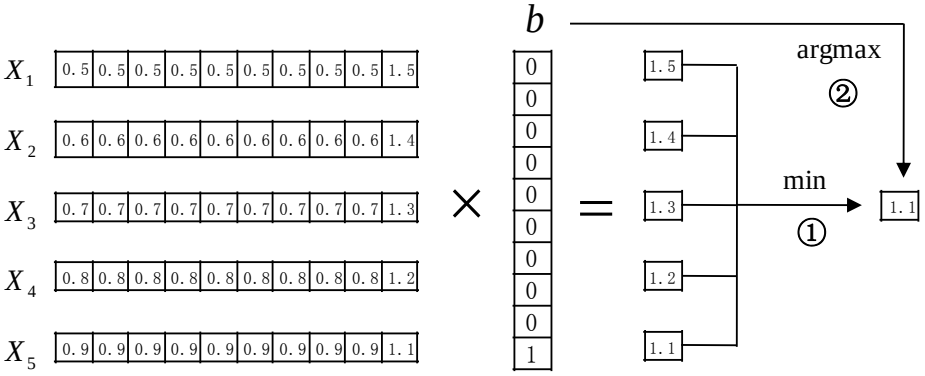


Figure 3: A toy example of the proposed loss control strategy. A portfolio $\hat{\mathbf{b}}$ is selected to maximize the minimum increasing factor $\mathbf{X}_i \mathbf{b}$ in the recent time window.

Taking the covariance estimate $\hat{\mathbf{\Sigma}}_{RO}$ into model (50), we should minimize the risk $\mathbf{b}^\top \hat{\mathbf{\Sigma}}_{RO} \mathbf{b}$ simultaneously in the optimization

$$\hat{\mathbf{b}} = \underset{\mathbf{b} \in \Delta_d}{\operatorname{argmax}} (\min_{1 \leqslant i \leqslant w} \mathbf{X}_i \mathbf{b}) - \gamma \mathbf{b}^\top \hat{\mathbf{\Sigma}}_{RO} \mathbf{b}, \tag{51}$$

where $\gamma > 0$ is a mixing parameter that trades off between the increasing factor and the risk. Since $-\gamma \mathbf{b}^\top \hat{\mathbf{\Sigma}}_{RO} \mathbf{b}$ is invariant to i, it could be put in the outer maximization, which is equivalent to minimizing $\gamma \mathbf{b}^\top \hat{\mathbf{\Sigma}}_{RO} \mathbf{b}$.

Compared with the original mean-variance model (5), our model (51) replaces the expected portfolio return $\boldsymbol{\mu}^\top \mathbf{b}$ with the worst increasing factor $\min_{1\leqslant i\leqslant w} \mathbf{X}_i\mathbf{b}$, which improves robustness to the short-term loss and the downside risk. Besides, using $\hat{\boldsymbol{\Sigma}}_{RO}$ instead of $\hat{\boldsymbol{\Sigma}}_{MP}$ could catch the instantaneous risk structure embedded in $\mathbf{X}$ with the representation via the principal rank-one tangent space $\mathscr{T}_1$. These mechanisms make our model effective in loss control and get stable growth in wealth.

To solve (51), we reformulate it as a quadratic programming. First, we introduce a new variable $q \in \mathbb{R}$ to simplify $\min_{1\leqslant i\leqslant w} \mathbf{X}_i\mathbf{b}$:

$$\min_{1\leqslant i\leqslant w} \mathbf{X}_i\mathbf{b} \quad \Longleftrightarrow \quad \max q \quad \text{s.t. } \mathbf{X}\mathbf{b} \geqslant q\mathbf{1}_w. \tag{52}$$

The inequality is componentwise, which can be inferred from the context. Now the index i no longer appears and the "hard-shift" problem of the worst-performance index is solved. Then (51) can be transformed into

$$\hat{\mathbf{b}} = \operatorname*{argmax}_{\mathbf{b}} q - \gamma \mathbf{b}^\top \hat{\boldsymbol{\Sigma}}_{RO}\mathbf{b} \quad \text{s.t. } \mathbf{b} \geqslant \mathbf{0}_d, \mathbf{1}_d^\top \mathbf{b} = 1, \mathbf{X}\mathbf{b} \geqslant q\mathbf{1}_w, \tag{53}$$

where we explicitly show the two constraints with respect to $\mathbf{b} \in \Delta_d$.

Next, we concatenate and augment some variables and matrices

$$\begin{aligned} &\mathbf{A} = [-\mathbf{X}, \mathbf{1}_w] \in \mathbb{R}^{w\times(d+1)}, \mathbf{z} = [\mathbf{b}; q] \in \mathbb{R}^{d+1}, \mathbf{f} = [\mathbf{0}_d; -1] \in \mathbb{R}^{d+1}, \\ &\mathbf{h} = [\mathbf{0}_d; -\infty] \in \mathbb{R}^{d+1}, \mathbf{y} = [\mathbf{1}_d; 0] \in \mathbb{R}^{d+1}, \mathbf{H} = \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{RO} & \mathbf{0}_d \\ \mathbf{0}_d^\top & 0 \end{bmatrix} \in \mathbb{S}_+^{d+1}, \end{aligned} \tag{54}$$

where $[\cdots,\cdots]$ and $[\cdots;\cdots]$ are concatenations in the row direction and in the column direction, respectively. Then (53) can be further transformed into a standard form of quadratic programming

$$\hat{\mathbf{z}} = \operatorname*{argmin}_{\mathbf{z}} \gamma \mathbf{z}^\top \mathbf{H}\mathbf{z} + \mathbf{f}^\top \mathbf{z} \quad \text{s.t. } \mathbf{A}\mathbf{z} \leqslant \mathbf{0}_w, \mathbf{z} \geqslant \mathbf{h}, \mathbf{y}^\top \mathbf{z} = 1. \tag{55}$$

Note that the maximization is changed to the minimization by adding a negative sign. Since $\mathbf{H}$ is positive semidefinite, the objective function is convex. Thus Slater's theorem and strong duality hold for (55), and standard primal-dual algorithms are capable to solve it (Boyd and Vandenberghe, 2004). Once $\hat{\mathbf{z}}$ is solved, the corresponding optimal portfolio is composed by the first d elements: $\hat{\mathbf{b}} = \hat{\mathbf{z}}[1:d]$. We refer to model (55) by SPOLC in the rest of this paper. The whole system can be summarized in Algorithm 1.

4. Experimental Results

The experiments include three main parts: the first is to set the parameters for SPOLC and verify that the rank-one covariance estimate cannot be replaced by simplified variants or LPO methods; The second is to assess the ability of controlling extreme losses and downside risk; The third is to evaluate the investing incomes of SPO systems versus the Market strategy and the $1/d$ strategy.

ALGORITHM 1: Short-term Portfolio Optimization with Loss Control (SPOLC)

Input: The observation matrix of asset price relatives $\mathbf{X} \in \mathbb{R}^{w \times d}$ with a window size of w and a dimensionality of d. Set the mixing parameter γ.

1. Conduct the singular value decomposition $\mathbf{X} = \mathbf{V}_J \Xi \mathbf{U}_J^\top$. Correspondingly, The principal eigenvalue and eigenvector of $\mathbf{S} = \mathbf{X}^\top \mathbf{X}$ are $\theta_1 = \Xi_{11}^2$ and $\mathbf{u}_1$, respectively.
2. Compute the transform matrix $\mathbf{D} = \Xi^2 - \frac{1}{w} \Xi \mathbf{V}_J^\top \mathbf{1}_w \mathbf{1}_w^\top \mathbf{V}_J \Xi$.
3. Let $\zeta_1^* = \left(\frac{\mathrm{tr}(\mathbf{D})}{d(w-1)} \right)^{-\frac{1}{2}} \theta_1$. Compute the rank-one covariance estimate by $\hat{\mathbf{\Sigma}}_{RO} = \mathbf{u}_1 \zeta_1^* \mathbf{u}_1^\top$.
4. Concatenate and augment some variables and matrices: $\mathbf{A} = [-\mathbf{X}, \mathbf{1}_w] \in \mathbb{R}^{w \times (d+1)}, \mathbf{z} = [\mathbf{b}; q] \in \mathbb{R}^{d+1}, \mathbf{f} = [\mathbf{0}_d; -1] \in \mathbb{R}^{d+1}$, $\mathbf{h} = [\mathbf{0}_d; -\infty] \in \mathbb{R}^{d+1}, \mathbf{y} = [\mathbf{1}_d; 0] \in \mathbb{R}^{d+1}, \mathbf{H} = \begin{bmatrix} \hat{\mathbf{\Sigma}}_{RO} & \mathbf{0}_d \\ \mathbf{0}_d^\top & 0 \end{bmatrix} \in \mathbb{S}_+^{d+1}$.
5. Solve the quadratic programming: $\hat{\mathbf{z}} = \mathrm{argmin}_{\mathbf{z}} \, \gamma \mathbf{z}^\top \mathbf{H} \mathbf{z} + \mathbf{f}^\top \mathbf{z} \quad \text{s.t. } \mathbf{A}\mathbf{z} \leqslant \mathbf{0}_w, \mathbf{z} \geqslant \mathbf{h}, \mathbf{y}^\top \mathbf{z} = 1$.

Output: The optimized portfolio $\hat{\mathbf{b}} = \hat{\mathbf{z}}[1:d]$.

4.1. Date Sets and Assessments

We conduct extensive experiments on 7 benchmark data sets from real-world financial markets. 4 standard data sets consist of daily price relatives of constituent stocks from Dow Jones Industrial Average (DJIA, Borodin et al. 2004), Standard & Pool 500 (SP500, Borodin et al. 2004), Toronto Stock Exchange (TSE, Borodin et al. 2004) and China Stock Index 300 (HS300, Lai et al. 2018a). Besides, we propose 2 new data sets: NYSE19 and FOF. We extract the constituents that have no more than 1 missing record during a recent period 2015 $\sim$ 2019 from the New York Stock Exchange (NYSE), which results in 47 stocks (their codes are given in Appendix A). Then we fill in each missing close price with the previous close price and convert all the prices to price relatives. FOF consists of daily price relatives of 24 mutual funds in China (details of these funds are given in Appendix A). A portfolio on a set of mutual funds functions with the same business logic as a fund of funds (FOF). The financial properties (e.g., volatility, correlation, liquidity, etc.) of mutual funds are quite different from those of stocks, thus the FOF data set could bring in new challenges and materials for new discoveries in SPO. Besides daily data sets, we also adopt a monthly data set (denoted by "FRENCH") from Kenneth R. French's Data Library[1]. It consists of 32 North American portfolios formed on size, book-to-market and investment. It is converted from return sequences to price relative sequences. We use this monthly data set to further investigate the performance of SPO systems in the LPO situation. Profiles of the 7 data sets are given in Table 1, which cover a wide range of financial markets, frequencies, asset types and time spans.

For comparison, we evaluate 5 SPO systems with active investments and state-of-the-art investing performance: OLMAR, RMR, PPT, AICTR, and SSPO. For the baseline of investing incomes, we use 2 common passive investments: the Market strategy (3) and

1. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Data set	Region	Asset type	Frequency	Time	Periods	Number of assets
NYSE19	US	Stock	Daily	$2/Jan/2015 \sim 4/Sep/2019$	1167	47
DJIA	US	Stock	Daily	$14/Jan/2001 \sim 14/Jan/2003$	507	30
SP500	US	Stock	Daily	$2/Jan/1998 \sim 31/Jan/2003$	1276	25
TSE	CA	Stock	Daily	$4/Jan/1994 \sim 31/Dec/1998$	1259	88
HS300	CN	Stock	Daily	$21/Jan/2016 \sim 16/Oct/2017$	421	44
FOF	CN	Fund	Daily	$4/Jan/2013 \sim 29/Dec/2017$	1215	24
FRENCH	NA	Portfolio	Monthly	$Jul/1990 \sim Feb/2020$	356	32

Table 1: Profiles of 7 benchmark data sets from real-world financial markets.

the $1/d$ strategy (4), as well as 1 LPO method combined with our loss control framework: NSCM (8). A good active SPO system should outperform these passive strategies and LPO method in the investing income while keeping the downside risk as low as possible at the same time. We use the default parameters in the original papers for the competitors (Li et al., 2015, 2016; Huang et al., 2016; Lai et al., 2018b,a,c): OLMAR: $w = 5, \epsilon = 10$; RMR: $w = 5, \epsilon = 5$; PPT: $w = 5$, $\epsilon = 100$; AICTR: $w = 5$, $\sigma_l^2 = 0.0025$, $\epsilon = 1000$; SSPO: $w = 5, \lambda = 0.5, \gamma = 0.01, \eta = 0.005, \zeta = 500$. The window size $w = 5$ is consistent among all these systems, which is much smaller than the numbers of assets (see Table 1). The parameters for the LPO method NSCM will be set in Section 4.2.4.

We use the following 8 key evaluating indices to assess the SPO systems. Indices $1 \sim 4$ offer a comprehensive assessment on the extreme losses and downside risk, while Indices $5 \sim 8$ evaluate the investing incomes. It is very challenging to achieve state-of-the-art performance in all these downside risk indices and keep competitive investing incomes simultaneously, since a higher investing income comes with a higher risk in general.

1. Value at Risk (VaR, Jorion 1997): the potential extreme loss over a trading period in a disastrous situation with a given small probability (confidence level).

2. Downside Semideviation (DSDV, Roy 1952; Markowitz 1959): the deviation that only counts downward terms.

3. Maximum Drawdown (MDD, Magdon-Ismail and Atiya 2004): the maximum percentage loss from a past peak to a past valley in the whole investment.

4. Calmar Ratio (CR, Young 1991): a downside-risk-adjusted return that smooths out overachievement or underachievement.

5. Cumulative Wealth (CW): the main index to assess investing incomes.

6. Mean Excess Return (MER, Jegadeesh 1990): the average excess income of a system versus the Market.

7. Sharpe Ratio (SR, Sharpe 1966): risk-adjusted average return.

8. Information Ratio (IR, Treynor and Black 1973): risk-adjusted MER.

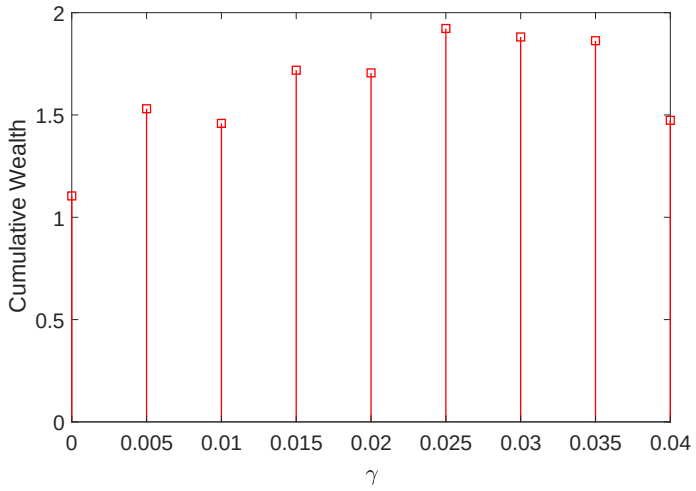


Figure 4: Cumulative wealths of SPOLC with respect to γ (fix $w = 5$).

4.2. Parameter Setting and Other Covariance Estimates

In this subsection, we set the parameters for SPOLC and further investigate how it works if we replace the rank-one covariance estimate $\hat{\mathbf{\Sigma}}_{RO}$ with other estimates, such as the unbiased covariance $\hat{\mathbf{\Sigma}}_{UB}$, the Gram matrix $\mathbf{S} = \mathbf{X}^\top \mathbf{X}$ or the NSCM $\hat{\mathbf{\Sigma}}_{NS}$.

4.2.1. Parameter Setting for SPOLC

At first, we set $w = 5$ to be consistent with other SPO systems. Then we set the mixing parameter γ in (55) according to the final CW in the FOF data set, which is in a similar way to Li et al. (2015); Huang et al. (2016); Lai et al. (2018b,a,c). Let γ change in $[0, 0.04]$, the corresponding CWs of SPOLC are plotted in Figure 4. The $\gamma = 0$ case is the non-risk model (50), which is worse than the $\gamma > 0$ case. Hence the proposed risk term $\mathbf{b}^\top \hat{\mathbf{\Sigma}}_{RO} \mathbf{b}$ is effective in SPO. Since $\gamma = 0.025$ achieves a good CW of 1.9228, we fix it in the rest of this paper.

As for the window size, $w = 5$ is a conventional and popular setting in SPO and in many indices of stock-trading softwares. Including more distant observations may affect the estimation at the current moment, since the financial characteristics are dynamic. But if one does need to tune w, we suggest that it could be increased by a little. At the same time, γ may also be adjusted by a little to fit the new window size. We give some examples for $w \in [5, 10]$ in Table 2.

4.2.2. SPOLC vs. Original Mean-variance

To verify that SPOLC is more suitable for SPO than the original mean-variance (OMV) model (5), we search in a wide range $[0.5e{-}3, 2.5e{+}11]$ for a good parameter γ for OMV with the unbiased covariance estimate $\hat{\mathbf{\Sigma}}_{UB}$ in (17). But the best case is only $CW = 1.3881$, which is much worse than $CW = 1.9228$ of SPOLC. Nor does OMV perform well on other benchmark data sets in this range of γ. For simplicity, we just show the results of

	NYSE19	DJIA	SP500	TSE	HS300	FOF	FRENCH
$w = 5, \gamma = 0.025$	2.4530	3.0097	58.1662	345.1481	1.5647	1.9228	6.5739
$w = 6, \gamma = 0.022$	0.7926	2.7925	29.6149	183.7565	1.5782	1.5090	9.0157
$w = 7, \gamma = 0.018$	1.0925	1.3318	28.8835	407.1672	1.1642	1.6725	7.8942
$w = 8, \gamma = 0.03$	1.3620	1.6184	12.1070	284.5688	1.5239	0.9372	5.6033
$w = 9, \gamma = 0.035$	2.7581	1.5830	10.7396	227.6947	1.3011	0.7258	3.8060
$w = 10, \gamma = 0.03$	0.9009	2.4711	25.1636	96.7747	1.4768	0.5971	5.5834

Table 2: Cumulative wealths of SPOLC with respect to w and γ.

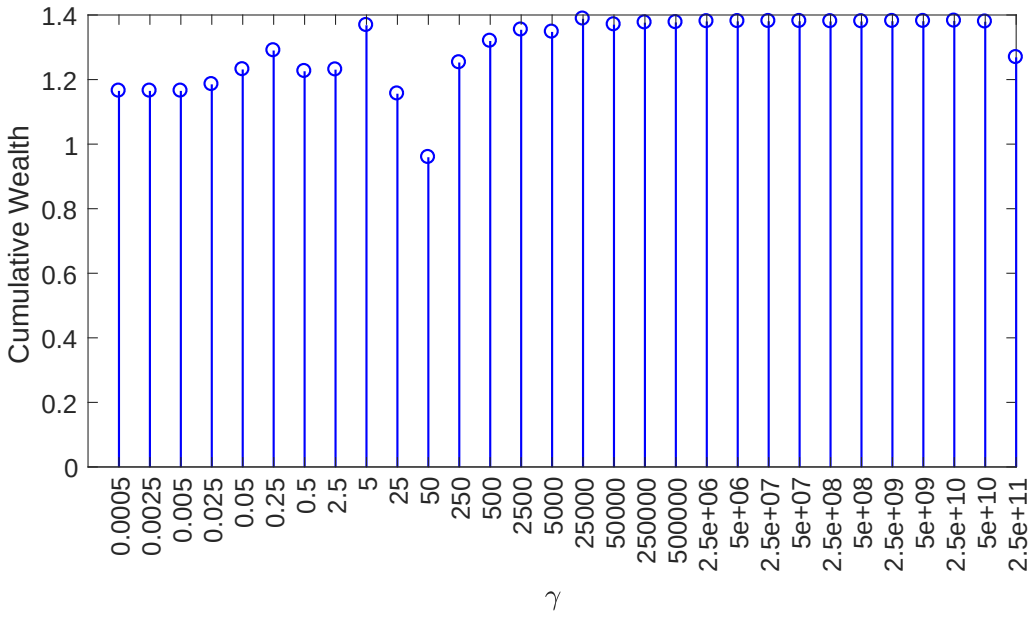


Figure 5: Cumulative wealths of the original mean-variance model with respect to γ.

FOF in Figure 5. It indicates that OMV stays at about $CW = 1.38$ when γ changes in $[2.5e + 4, 5e + 10]$. Hence this traditional model for LPO is not effective in SPO.

4.2.3. Rank-one Covariance vs. Gram Matrix

(41) shows that the Gram matrix $\mathbf{S} = \mathbf{X}^\top \mathbf{X}$ seems to be an approximate rank-one covariance in structure since $\theta_1 \gg \theta_i$ $(i \neq 1)$. However, the remaining spectrums $\{\mathbf{u}_i \theta_i \mathbf{u}_i^\top\}_{i \neq 1}$ in the minor tangent spaces still weaken the ability of catching the instantaneous principal risk structure. Besides, $\mathbf{S}$ does not extract any asset volatility information from $\hat{\mathbf{\Sigma}}_{MP}$, while $\hat{\mathbf{\Sigma}}_{MP}$ could be useful in setting the magnitude of covariance estimate. Even if $\mathbf{S}$ were useful, it cannot do without the detailed analyses of the rank-related structure of $\mathbb{S}^d$ in Section 3.

To see how $\mathbf{S}$ works, we substitute $\hat{\mathbf{\Sigma}}_{RO}$ in (51) for $\mathbf{S}$ and search γ in a wide range $[0.5e - 3, 2.5e + 11]$ for a good mixing parameter. The overall best case is $\gamma = 2.5$ and the corresponding CWs on the benchmark data sets are shown in Table 3. It indicates that $\hat{\mathbf{\Sigma}}_{RO}$ is superior to $\mathbf{S}$ on almost all the data sets except for NYSE19. The gaps are rather

Cov.	NYSE19	DJIA	SP500	TSE	HS300	FOF	FRENCH
$\hat{\mathbf{\Sigma}}_{RO}$	2.4530	**3.0097**	**58.1662**	**345.1481**	**1.5647**	**1.9228**	6.5739
$\mathbf{S}$	**3.4868**	2.9660	43.7565	161.9418	1.4978	1.9142	5.4281
$\hat{\mathbf{\Sigma}}_{NS}$	1.1233	0.7307	1.0758	1.5738	1.3099	1.4283	**20.4540**

Table 3: Cumulative wealths of SPOLC with different covariance estimates. $\hat{\mathbf{\Sigma}}_{RO}$ is the proposed rank-one covariance estimate, $\mathbf{S}$ is the Gram matrix $\mathbf{X}^\top\mathbf{X}$, and $\hat{\mathbf{\Sigma}}_{NS}$ is the NSCM method.

w_{NS}	NYSE19	DJIA	SP500	TSE	HS300	FOF	FRENCH
5	0.8150	0.4655	0.2910	0.2182	0.6224	1.1134	34.0923
100	0.7913	0.4643	0.2920	0.2184	0.6225	1.1305	35.0910
200	0.7893	0.4642	0.2916	0.2165	0.6225	1.1341	35.0785
300	0.7900	0.4642	0.2920	0.2173	0.2915	1.1349	35.0792

Table 4: Cumulative wealths of NSCM with respect to w_{NS} (fix $\gamma = 0.025$).

significant on SP500, TSE, HS300 and FRENCH. Thus it is beneficial to use $\hat{\mathbf{\Sigma}}_{RO}$ instead of $\mathbf{S}$ in SPOLC.

4.2.4. Rank-one Covariance vs. NSCM

The NSCM $\hat{\mathbf{\Sigma}}_{NS}$ in (8) is an improved covariance estimate for LPO. To see whether $\hat{\mathbf{\Sigma}}_{NS}$ or $\hat{\mathbf{\Sigma}}_{RO}$ is more effective, we substitute $\hat{\mathbf{\Sigma}}_{RO}$ in (51) for $\hat{\mathbf{\Sigma}}_{NS}$. Moreover, we try both small and large window sizes $w_{NS} = 5, 100, 200, 300$ in the computation of $\hat{\mathbf{\Sigma}}_{NS}$, because more observations might better show its strength according to its theory of consistency. Note that the window size of the first term $(\min_{1\leqslant i\leqslant w} \mathbf{X}_i\mathbf{b})$ in (51) remains $w = 5$, in order to conduct a fair comparison between $\hat{\mathbf{\Sigma}}_{NS}$ and $\hat{\mathbf{\Sigma}}_{RO}$.

First, we fix $\gamma = 0.025$ and conduct CW experiments with these window sizes. The results are shown in Table 4. It indicates that increasing the window size would not bring about better performance for $\hat{\mathbf{\Sigma}}_{NS}$. Second, we just keep $w_{NS} = 100$ and search γ in a wide range $[0.5e-3, 2.5e+11]$ for a good mixing parameter. The overall best case is $\gamma = 2.5e+7$ and the corresponding CWs are shown in Table 3. The results match the expectation that $\hat{\mathbf{\Sigma}}_{NS}$ works well on the monthly data set FRENCH but quite badly on all the daily data sets. Hence the proposed $\hat{\mathbf{\Sigma}}_{RO}$ is a better choice for SPO than $\hat{\mathbf{\Sigma}}_{NS}$ as an improved covariance estimate.

4.3. Assessment on Extreme Losses and Downside Risk

In this subsection, we mainly assess the extreme losses and downside risk of different SPO systems with several common risk metrics.

4.3.1. Value at Risk

The most important and popular index of loss assessment is the value at risk (VaR, Jorion 1997), because it not only measures the percentage loss but also considers its prob-

ability. It is formally the quantile of the strategy return at a given confidence level c

$$\Pr\{r_s < VaR\} = 1 - c, \tag{56}$$

where $\Pr\{r_s < VaR\}$ is the probability of the event $\{r_s < VaR\}$, r_s is the return of an investing strategy on one trading period. For example, a $VaR = -0.1$ at $c = 0.95$ means that the investing strategy may suffer a percentage loss $> 10\%$ with a probability of 5% on one trading period.

With the problem setting of this paper, we can use the empirical return distributions of different SPO systems to assess their VaRs. Once an SPO system runs through a data set with n trading periods, we convert the increasing factors to returns

$$r_{s,t} = \hat{\mathbf{b}}_{s,t}^{\top}\mathbf{x}_t - 1, \quad t = 1, 2, \cdots, n, \tag{57}$$

where $r_{s,t}$ and $\hat{\mathbf{b}}_{s,t}$ are the return and the optimized portfolio of the SPO system at the t-th period, respectively. We could sort $\{r_{s,t}\}_{t=1}^{n}$ as $\{r_{s,(k)}\}_{k=1}^{n}$ in the ascending order, and pick up the $(1-c)$ quantile as $\widehat{VaR}$. To be specific, the empirical distribution of $\Pr\{r_s < VaR\}$ is

$$\widehat{\Pr}\{r_s < \widehat{VaR}\} = \frac{\#\{r_{s,t} : r_{s,t} < \widehat{VaR}\}}{n} \approx 1 - c, \tag{58}$$

where $\#\{r_{s,t} : r_{s,t} < \widehat{VaR}\}$ denotes the count for $\{r_{s,t}\}_{t=1}^{n}$ such that $r_{s,t} < \widehat{VaR}$. Given a confidence level c, we use the k-th smallest $r_{s,t}$ as $\widehat{VaR}$ such that

$$\begin{aligned} k &\triangleq \lfloor n(1-c) \rfloor, \quad r_{s,(1)} \leqslant r_{s,(2)} \leqslant \cdots \leqslant r_{s,(k)} \leqslant \cdots \leqslant r_{s,(n)}, \\ \widehat{VaR} &\triangleq r_{s,(k)}, \end{aligned} \tag{59}$$

where $\lfloor \cdot \rfloor$ denotes the floor operator.

An SPO system with a higher VaR endures a smaller loss at the given confidence level. We show the VaRs at two common confidence levels $c = 90\%$ and $c = 95\%$ for different SPO systems in Table 5. SPOLC achieves the highest VaRs at both confidence levels on all the daily or monthly data sets. It outperforms other SPO systems by percentages of about $9\% \sim 15\%$ when $c = 90\%$ and $2\% \sim 15\%$ when $c = 95\%$, which shows state-of-the-art performance in loss control. For example, the VaR of SPOLC at $c = 90\%$ on FOF is -0.0166, which means that SPOLC may suffer a percentage loss greater than 1.66% on one trading day with 10% probability. Such a relatively low level of downside risk indicates that SPOLC is effective in catching the instantaneous risk structure of the current financial circumstance.

4.3.2. Downside Semideviation

Most ordinary investors gain profits mainly by the long position in the portfolio. Although some may be able to make profits by pure short position, these profits could not reflect the real growth of the whole financial system. Hence the impact of downside risk is generally much greater than that of upside risk, but the traditional standard deviation (STD) treats both risks equally as a common risk metric. To stress the importance of

		NYSE19	DJIA	SP500	TSE	HS300	FOF	FRENCH
$c = 90\%$	OLMAR	−0.0390	−0.0368	−0.0354	−0.0427	−0.0167	−0.0205	−0.0740
	RMR	−0.0381	−0.0367	−0.0360	−0.0408	−0.0161	−0.0214	−0.0751
	PPT	−0.0415	−0.0359	−0.0366	−0.0435	−0.0179	−0.0244	−0.0833
	AICTR	−0.0400	−0.0368	−0.0360	−0.0435	−0.0179	−0.0217	−0.0810
	SSPO	−0.0405	−0.0358	−0.0366	−0.0435	−0.0179	−0.0243	−0.0810
	SPOLC	**−0.0325**	**−0.0313**	**−0.0301**	**−0.0370**	**−0.0143**	**−0.0166**	**−0.0672**
$c = 95\%$	OLMAR	−0.0554	−0.0498	−0.0492	−0.0710	−0.0260	−0.0352	−0.1149
	RMR	−0.0554	−0.0482	−0.0488	−0.0625	−0.0257	−0.0354	−0.1148
	PPT	−0.0591	−0.0500	−0.0512	−0.0692	−0.0267	−0.0421	−0.1148
	AICTR	−0.0563	−0.0498	−0.0508	−0.0698	−0.0283	−0.0361	−0.1149
	SSPO	−0.0589	−0.0498	−0.0512	−0.0678	−0.0267	−0.0421	−0.1149
	SPOLC	**−0.0511**	**−0.0473**	**−0.0440**	**−0.0526**	**−0.0219**	**−0.0310**	**−0.0973**

Table 5: VaRs at two confidence levels $c = 90\%$ and $c = 95\%$ for SPO systems on 7 benchmark data sets.

System	NYSE19	DJIA	SP500	TSE	HS300	FOF	FRENCH
OLMAR	0.0253	0.0219	0.0217	0.0343	0.0137	0.0172	0.0483
RMR	0.0249	0.0219	0.0215	0.0327	0.0135	0.0169	0.0483
PPT	0.0260	0.0221	0.0230	0.0341	0.0135	0.0189	0.0494
AICTR	0.0256	0.0218	0.0227	0.0339	0.0140	0.0179	0.0476
SSPO	0.0254	0.0218	0.0231	0.0340	0.0135	0.0189	0.0495
SPOLC	**0.0224**	**0.0203**	**0.0192**	**0.0275**	**0.0113**	**0.0155**	**0.0416**

Table 6: Downside semideviations for SPO systems on 7 benchmark data sets.

downside risk in risk portfolio management, Roy (1952) and Markowitz (1959) propose the downside semideviation (DSDV) that only counts the downward terms

$$\sigma_s^- = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\{(r_{s,t} - \bar{r}_s)^-\}^2}, \quad (r_{s,t} - \bar{r}_s)^- \triangleq \begin{cases} 0 & \text{if } r_{s,t} \geqslant \bar{r}_s \\ \bar{r}_s - r_{s,t} & \text{if } r_{s,t} < \bar{r}_s \end{cases}, \tag{60}$$

where $\bar{r}_s = \frac{1}{n}\sum_{t=1}^{n} r_{s,t}$ is the mean return for the SPO system, and $(\cdot)^-$ denotes the negative part of a real number.

The DSDVs for different SPO systems are shown in Table 6. SPOLC outperforms all the other SPO systems on all the daily or monthly data sets. For example, SPOLC achieves $\sigma^-_{SPOLC} = 0.0275$ and 0.0113 on TSE and HS300, respectively, which are much smaller than $\sigma^-_{RMR} = 0.0327$ and 0.0135 for the second best competitor RMR. It indicates that SPOLC is effective in controlling downside risk, which is more attractive to the risk-averting investors with heavy long positions.

4.3.3. Maximum Drawdown

In real-world investments, risk-averting investors are highly concerned about the worst case in the losses they might suffer. Such an extreme loss can be measured by the maximum drawdown (MDD, Magdon-Ismail and Atiya 2004), which is the maximum percentage loss of wealth from a past peak to a past valley in the whole investment with a particular investing

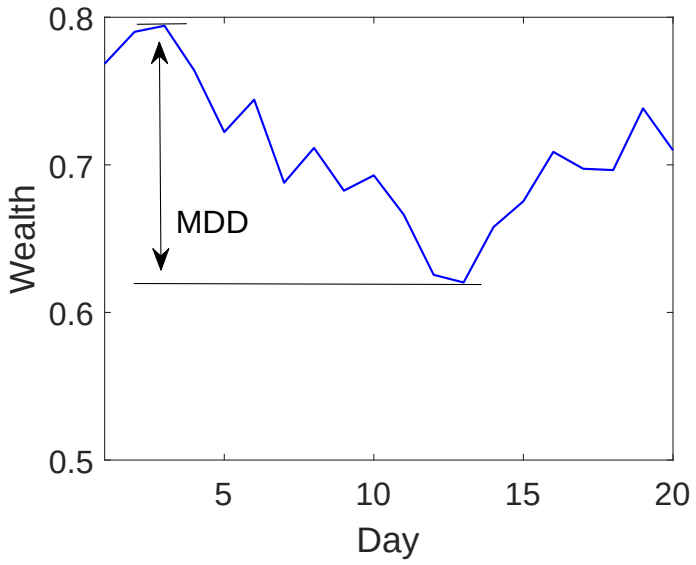


Figure 6: Diagram of maximum drawdown.

strategy

$$MDD \triangleq \max_{\tau\in[1,n]} \frac{\max\limits_{t\in[1,\tau]} \hat{S}_t - \hat{S}_\tau}{\max\limits_{t\in[1,\tau]} \hat{S}_t} = 1 - \min_{\tau\in[1,n]} \left(\frac{\hat{S}_\tau}{\max\limits_{t\in[1,\tau]} \hat{S}_t} \right), \tag{61}$$

where $\hat{S}_t$ denotes the CW of the investing strategy at time t. As the current time τ passes, it searches the past time $t \in [1, \tau]$ for the peak and the valley CWs and computes the maximum percentage loss. Note that to be consistent with the custom in the finance industry, the MDD here is positive (i.e., the absolute value of the actual percentage loss).

A diagram of MDD is shown in Figure 6. It is a **hard index** that will not decrease as the investment proceeds. Even if an investing strategy works well for the most time but fails seriously for only once in a disadvantageous financial circumstance, the MDD will be very high and it may take much time to recover. Besides, the investors may lose confidence in this strategy and withdraw their investments, which may lead to the liquidation of the corresponding fund. Hence MDD is one of the main downside risk metrics adopted by most institutional investors to supervise their portfolios and funds.

The MDDs for different SPO systems are shown in Table 7. SPOLC outperforms other SPO systems to a large extent on 6 data sets. Although it is slightly behind AICTR (0.3324) or RMR (0.3469) on DJIA, the gap is small: $0.3552 - 0.3324 = 0.0228$. Besides, all the other SPO systems suffer high MDDs on at least one of the data sets. The highest MDD for SPOLC is 0.7186, while the highest MDDs for other systems are: 0.8357 (OLMAR), 0.8236 (RMR), 0.8160 (PPT), 0.8100 (AICTR), and 0.8230 (SSPO), all on NYSE19. The advantage of SPOLC over other SPO systems in the highest MDD is nearly 0.1. Other SPO systems also suffer high MDDs on TSE, FOF and FRENCH. Hence SPOLC achieves much more robust performance than other SPO systems in controlling extreme losses.

		NYSE19	DJIA	SP500	TSE	HS300	FOF	FRENCH
MDD	OLMAR	0.8357	0.3685	0.4117	0.8201	0.2277	0.6043	0.7952
	RMR	0.8236	0.3469	0.4631	0.7488	0.2333	0.5932	0.8051
	PPT	0.8160	0.4056	0.5249	0.7692	0.2440	0.7120	0.7375
	AICTR	0.8100	**0.3324**	0.4161	0.6637	0.2265	0.6596	0.7748
	SSPO	0.8230	0.3853	0.5190	0.7490	0.2391	0.7006	0.7459
	SPOLC	**0.7186**	0.3552 (3)	**0.3379**	**0.5095**	**0.1902**	**0.5097**	**0.6426**
CR	OLMAR	−0.0242	1.5970	1.7679	1.5340	0.5267	0.0997	0.0253
	RMR	0.0707	1.8124	1.1188	2.4463	0.8353	0.1659	0.0177
	PPT	0.0711	1.8470	1.1415	2.7089	0.2226	0.0746	0.0322
	AICTR	0.1015	**3.1115**	1.6564	**4.3761**	1.0199	0.0188	0.0240
	SSPO	0.0790	2.3643	1.4436	3.0137	0.1969	0.0845	0.0359
	SPOLC	**0.2976**	2.0529 (3)	**3.6432**	4.3594 (2)	**1.6158**	**0.2849**	**0.1020**

Table 7: Maximum drawdowns (MDD) and Calmar ratios (CR) for SPO systems on 7 benchmark data sets. The numbers in the parentheses are the rankings of SPOLC on the corresponding data sets when it is not the best SPO system.

4.3.4. Calmar Ratio

In contrast to risk-averting investors, risk-seeking investors are more concerned about how much return they could make if they are willing to take a certain amount of downside risk. Calmar ratio (CR, Young 1991) is such a downside-risk-adjusted return that smooths out overachievement or underachievement. It is not so hard as MDD, since a sufficient growth could cover a high drawdown. According to its definition, we first convert the CW to the annualized return (AR)

$$AR = \hat{S}_n^{\frac{\nu}{n}} - 1, \tag{62}$$

where ν denotes the total number of trading days (or months) for one year while the investment lasts for n trading days (or months). We set $\nu = 252$ (daily data) or $\nu = 12$ (monthly data) according to the custom in the finance industry. Then CR is defined as follows

$$CR = \frac{AR}{MDD}. \tag{63}$$

The CRs for different SPO systems are shown in Table 7. SPOLC achieves the highest CRs on 5 data sets, and is very close to the first-place AICTR on TSE. In particular, SPOLC achieves much better CRs than other SPO systems on NYSE19, SP500, FOF and FRENCH, which involve different data types. Therefore, SPOLC is competitive in trading off between downside risk and return in various scenarios.

4.4. Assessment on Investing Incomes

In this subsection, we turn to evaluate the investing incomes for the active SPO systems. An effective active SPO system should not sacrifice the excess return just to avoid risk. At least, it should outperform some passive strategies like the Market strategy (3) and the $1/d$ strategy (4), or the LPO method NSCM.

		NYSE19	DJIA	SP500	TSE	HS300	FOF	FRENCH
CW	Market	1.3300	0.7644	1.3416	1.6129	1.3419	1.8202	19.3634
	$1/d$	1.3829	0.8127	1.6487	1.5952	1.3280	1.8951	18.3226
	NSCM	1.1233	0.7307	1.0758	1.5738	1.3099	1.4283	20.4540
	OLMAR	0.9096	2.5372	15.9435	58.5127	1.2083	1.3258	1.8040
	RMR	1.2999	2.6682	8.2800	181.3437	1.3464	1.5723	1.5226
	PPT	1.2982	3.0800	10.7760	277.5791	1.0924	1.2834	2.0062
	AICTR	1.4417	**4.1734**	14.2208	**902.5091**	1.4151	1.0612	1.7284
	SSPO	1.3389	3.6800	16.9677	364.9443	1.0799	1.3197	2.1920
	SPOLC	**2.4530**	3.0097 (4)	**58.1662**	345.1481 (3)	**1.5647**	**1.9228**	**6.5739**
MER	$1/d$	$3.6469e-05$	0.0001	0.0001	0.0000	0.0000	0.0000	$-2.5081e-04$
	NSCM	$-1.7585e-04$	$-1.4008e-04$	$-2.2472e-04$	$-4.1182e-05$	$-7.0407e-05$	$-3.0926e-04$	$-2.8438e-04$
	OLMAR	$3.0440e-04$	0.0028	0.0024	0.0045	-0.0001	-0.0001	-0.0056
	RMR	$5.9776e-04$	0.0029	0.0019	0.0053	0.0001	0.0000	-0.0061
	PPT	$6.5369e-04$	0.0032	0.0022	0.0058	-0.0004	0.0000	-0.0052
	AICTR	$7.1689e-04$	**0.0038**	0.0024	**0.0067**	0.0003	-0.0003	-0.0058
	SSPO	$6.5703e-04$	0.0036	0.0025	0.0060	-0.0004	0.0000	-0.0049
	SPOLC	**0.0011**	0.0030 (4)	**0.0033**	0.0055 (4)	**0.0005**	**0.0002**	**−0.0024**

Table 8: Cumulative wealths (CW) and mean excess returns (MER) for SPO systems on 7 benchmark data sets. The numbers in the parentheses are the rankings of SPOLC on the corresponding data sets when it is not the best SPO system.

4.4.1. Cumulative Wealth

The most important assessment is the final CWs for different SPO systems, which are shown in Table 8. SPOLC outperforms Market, $1/d$ and NSCM on all the 6 daily data sets and achieves the highest CWs on 4 daily data sets. Besides, SPOLC outperforms each competitor on at least 5 data sets, which shows a competitive performance in gaining investing incomes. As for the monthly data set FRENCH, the passive strategies and NSCM perform quite well as they are designed for this situation, but SPOLC still outperforms other SPO systems to a large extent. Although some competitors get better results than SPOLC on DJIA and TSE, they deteriorate badly on NYSE19, HS300 and FOF. Besides, only SPOLC outperforms Market or $1/d$ on the fund data set FOF. Thus SPOLC gains stable and favourable incomes due to its mechanism of downside risk control. We also draw the CW plots for different SPO systems on HS300 in Figure 7. It shows that SPOLC gains more wealths and suffers less drawdowns than other SPO systems on most days.

4.4.2. Mean Excess Return

The final CW is only a final score at the end of an investment. If we want to assess the average performance during the investment, we can use the mean excess return (MER, Jegadeesh 1990). It compares the mean return of an active SPO system with the Market baseline (see Lai et al. 2018c for its computation). The MERs for different SPO systems are shown in Table 8. SPOLC is the only system that achieves positive MERs on all the daily data sets, and it achieves the highest MERs on 4 daily data sets. Besides, SPOLC outperforms each competitor on at least 5 data sets. Hence SPOLC is competitive in gaining excess returns besides controlling downside risk. Although some competitors achieve high MERs on DJIA and TSE, they get worse than Market on HS300 or FOF. Thus they do not fit the financial properties in the two new data sets. By contrast, SPOLC is effective to fill this gap.

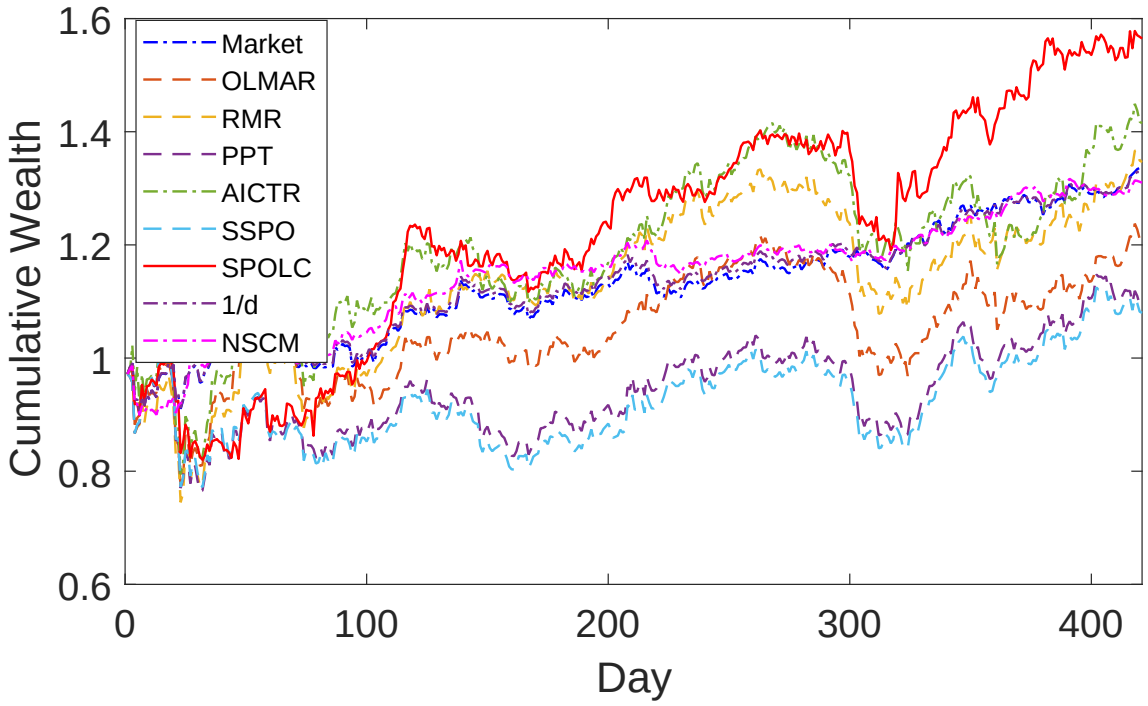


Figure 7: Cumulative wealth plots for SPO systems on HS300.

4.4.3. Sharpe Ratio

The Capital Asset Pricing Model (CAPM, Sharpe 1964) is a fundamental theory that prices a return at a reasonable level of risk. It leads to a popular risk-adjusted return in the finance industry, the Sharpe ratio (SR, Sharpe 1966). Its computation can be found in Lai et al. (2018c). The SRs for different SPO systems are shown in Table 9. SPOLC outperforms other SPO systems to a large extent on 5 data sets. It is notable that SPOLC gets better rankings in SR than in CW or MER on DJIA and TSE (see also Table 8). It indicates that SPOLC gets more competitive and advantageous when risk is taken into consideration. Furthermore, SPOLC is very close to the first-place AICTR on TSE. To summarize, SPOLC is effective in gaining returns on the premise of rigorous downside risk control.

4.4.4. Information Ratio

We further use the information ratio (IR, Treynor and Black 1973) as a combination of MER and SR. Its computation can be found in Lai et al. (2018c). The IRs for different SPO systems are shown in Table 9. Similar to the results on SR, SPOLC outperforms other SPO systems to a large extent on 5 data sets and gets better rankings in IR than in CW or MER on DJIA and TSE. Moreover, SPOLC outperforms each competitor on at least 5 data sets. As an improved covariance estimate, SPOLC outperforms NSCM on all the daily data sets. Note that while SPOLC keeps good IRs on NYSE19, HS300 and FOF, other SPO systems deteriorate badly on these three data sets. It indicates that SPOLC can

		NYSE19	DJIA	SP500	TSE	HS300	FOF	FRENCH
SR	$1/d$	0.0361	−0.0179	0.0349	0.0480	0.0813	0.0426	0.2036
	NSCM	0.0282	−0.0488	0.0103	0.0598	0.0996	0.0816	0.2457
	OLMAR	0.0166	0.0731	0.0806	0.0820	0.0372	0.0217	0.0603
	RMR	0.0248	0.0763	0.0659	0.0981	0.0512	0.0277	0.0531
	PPT	0.0253	0.0820	0.0698	0.1018	0.0246	0.0211	0.0649
	AICTR	0.0275	**0.0994**	0.0762	**0.1179**	0.0555	0.0144	0.0583
	SSPO	0.0258	0.0919	0.0791	0.1054	0.0231	0.0219	0.0686
	SPOLC	**0.0398**	0.0877 (3)	**0.1165**	0.1134 (2)	**0.0768**	**0.0356**	**0.1216**
IR	$1/d$	0.0207	0.1156	0.0371	−0.0073	−0.0276	0.0400	−0.0456
	NSCM	−0.0268	−0.0154	−0.0233	−0.0097	−0.0115	−0.0226	−0.0140
	OLMAR	0.0089	0.1060	0.0848	0.0768	−0.0079	−0.0059	−0.1504
	RMR	0.0177	0.1094	0.0678	0.0931	0.0099	0.0026	−0.1688
	PPT	0.0187	0.1177	0.0728	0.0970	−0.0255	−0.0029	−0.1344
	AICTR	0.0207	**0.1366**	0.0810	**0.1133**	0.0181	−0.0155	−0.1567
	SSPO	0.0190	0.1307	0.0840	0.1007	−0.0275	−0.0014	−0.1297
	SPOLC	**0.0334**	0.1272 (3)	**0.1303**	0.1084 (2)	**0.0343**	**0.0115**	**−0.0744**

Table 9: Sharpe ratios (SR) and information ratios (IR) for SPO systems on 7 benchmark data sets. The numbers in the parentheses are the rankings of SPOLC on the corresponding data sets when it is not the best SPO system.

obtain excess risk-adjusted returns in different financial scenarios while other SPO systems cannot.

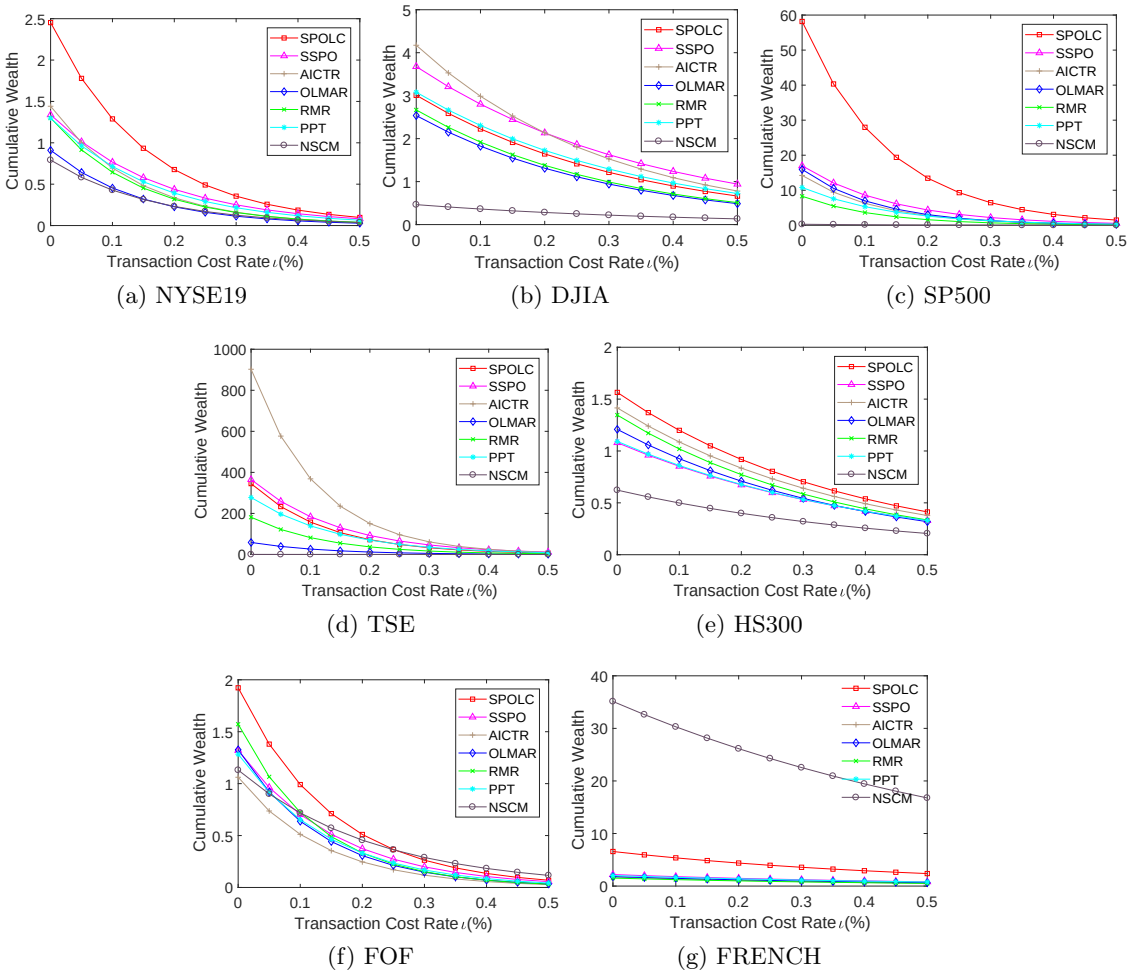
4.5. Transaction Costs and Running Times

To see how the SPO systems work with transaction costs, we use the proportional transaction cost model (Blum and Kalai, 1999; Li et al., 2015; Huang et al., 2016; Lai et al., 2018c) to calculate the CWs with respect to the transaction cost rate $\iota \in [0, 0.5\%]$. Figure 8 indicates that SPOLC outperforms other SPO systems on 5 data sets and keeps its advantage when ι increases from 0 to 0.5%. It also outperforms NSCM on all the 6 daily data sets with a moderate transaction cost rate $\iota \in [0, 0.25\%]$, and outperforms NSCM on 5 daily data sets with any $\iota \in [0, 0.5\%]$. Thus SPOLC can withstand a considerable transaction cost rate in SPO.

SPOLC enjoys high computational efficiency besides a good ability of loss control and competitive investing performance. We use an ordinary-setting computer with an Intel Core i5-4440 CPU and a 16GB DDR3 memory card to run SPOLC in the experiments. The average running times (in seconds) of SPOLC for one trade on different data sets are: NYSE19 (0.0166s), DJIA (0.0129s), SP500 (0.0112s), TSE (0.0170s), HS300 (0.0157s), FOF (0.0112s), FRENCH (0.0126s). Such a running speed is capable of large-scale and time-limited trading such as High-Frequency Trading (HFT, Aldridge 2013).

5. Conclusions and Discussions

In this paper, we propose a novel loss control scheme with a rank-one covariance estimate for SPO. In the rapidly-changing financial environment of SPO, some financial characteristics like the expected return and the true covariance might be dynamic $\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t$ rather than



(a) NYSE19 (b) DJIA (c) SP500

(d) TSE (e) HS300

(f) FOF (g) FRENCH

Figure 8: Cumulative wealths of SPO systems with respect to the transaction cost rate ι on 7 benchmark data sets.

static $\boldsymbol{\mu}, \boldsymbol{\Sigma}$. We usually have only a small window size w of observations that are sufficiently close to the current moment t and reliable to make estimations. w is usually much smaller than the number of assets d, then the statistical assumptions regarding asset price relatives are easily violated. Thus the traditional covariance estimate becomes invalid in statistical efficiency and consistency, which results in bad investing performance.

In order to solve this undersampled problem, we propose to reconsider the covariance estimation problem in the perspective of operators. It is essentially a symmetric quadratic form in $\mathbb{S}^d$. Hence we carefully investigate the rank-related structure of $\mathbb{S}^d$, including the determinantal varieties, their tangent spaces at the spectral components of the observation matrix, and the corresponding orthogonal projections. This orthogonality helps to separate the structural information of $\mathbb{S}^d$ into uncorrelated components and extract the most significant component. Based on this property, we propose a rank-one covariance estimate $\hat{\boldsymbol{\Sigma}}_{RO}$

lying in the principal tangent space together with a loss control scheme for SPO, which effectively catches the instantaneous risk structure of the current financial circumstance.

Extensive experiments are conducted on 7 real-world benchmark data sets, in which 5 contain daily stock data, 1 contains daily fund data, and 1 contains monthly portfolio data. The proposed SPOLC system outperforms other state-of-the-art active SPO systems in the comprehensive assessment of extreme losses and downside risk in most cases. SPOLC achieves the highest VaR and the smallest downside semideviation. It also keeps relatively low maximum drawdowns versus other SPO systems in all the cases. Hence it is very strong in controlling extreme losses and downside risk. Moreover, SPOLC is competitive in gaining investing incomes on the premise of rigorous downside risk control, according to the risk-adjusted return metrics (Calmar ratio, Sharpe ratio, and information ratio). It outperforms the Market benchmark, the $1/d$ strategy, and the LPO method NSCM in CW on all the daily data sets. It also outperforms other SPO systems in CW on at least 5 data sets. As an improved covariance estimate, $\hat{\mathbf{\Sigma}}_{RO}$ outperforms the original mean-variance model, the Gram matrix and the NSCM method in most cases. Thus $\hat{\mathbf{\Sigma}}_{RO}$ is an effective covariance estimate especially in SPO. Last, SPOLC can withstand a considerable transaction cost rate and enjoys high computational efficiency, which are useful in practical situations.

In conclusion, the perspective of operators and operator spaces might be a promising approach for undersampled estimations in SPO. Nevertheless, there are some limitations to be noted and they could be addressed in future works. First, the actual transaction costs include not only the service charge with a fixed rate but also the bid-ask spread, and the latter has not been considered in the problem setting of SPO. Second, the data sets may contain some illiquid assets that may produce good results. Third, whether SPO and LPO could be effectively handled in a unified framework is an interesting problem for future investigations.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grants 61703182, 61972179, 61602211, 61603152, 61772232, 61773179, 61877029, 61966009, 61602210, in part by the Fundamental Research Funds for the Central Universities under Grants 21617347, 21617404, 21617408, 21619404, in part by the Talent Introduction Foundation of Jinan University under Grants 88016653, 88016534, in part by the Science and Technology Planning Project of Guangdong under Grants 2019KTSCX010, 2018KTSCX016, 2019A050510024, in part by the Science and Technology Program of Guangzhou, China under Grants 201707010259, 201707010320, 201902010041, and in part by the Project of Guangxi Key Laboratory of Trusted Software under Grant kx202007. Liangda Fang is also affiliated to Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, China.

Appendix A. Information about the FOF and the NYSE19 Data Sets

Table 10 gives the codes, names, times since establishments (time) and net asset values (NAV) of the 24 mutual funds in the FOF data set. The censoring time is $29/Dec/2017$.

Code	Name	Time (year)	NAV (CNY)
150019	Yinhua SZSE 100 Index Structured Fund	7.66	3,251,532,745.20
150023	SWS MU Shenzhen Composition Index Structured Fund	7.20	2,339,036,988.22
159902	ChinaAMC China SME ETF	11.57	2,628,453,724.73
159903	China Southern Shenzhen Stock Exchange Component Stock Index ETF	8.08	454,122,828.12
159907	GF Mid/Small-Plate 300 ETF	6.58	230,156,107.84
159915	E Fund GEM ETF	6.28	5,165,768,469.64
159919	Harvest SZSE SME-CHINEXT 300 ETF	5.65	18,339,948,550.23
159920	ChinaAMC Hang Seng ETF	5.40	2,179,342,466.72
160311	ChinaAMC Blue Chip Core Hybrid Fund (LOF)	10.70	4,480,724,091.32
160505	Bosera Thematic Sectors Hybrid Fund (LOF)	12.99	10,652,841,166.39
160607	Penghua Value Advantage Hybrid Fund (LOF)	11.46	2,378,770,662.37
160916	Dacheng Selected Hybrid Fund (LOF)	5.43	841,008,868.47
161010	Fullgoal Tianfeng Consolidate Income Bond Fund	9.19	431,546,991.03
162607	Invesco Great Wall Resources And Monopoly Hybrid Fund (LOF)	11.94	1,830,082,897.16
162703	GF Small-cap Growth Hybrid Fund(LOF)	12.92	2,305,886,940.89
163402	Aegon-Industrial Trend Investment Hybrid Securities Investment Fund (LOF)	12.17	13,260,428,885.65
163503	China Nature Kernel Pullulation Hybrid Fund (LOF)	11.95	594,212,639.86
163801	BOC China Selected Hybrid Fund (LOF)	13.00	1,342,616,257.95
510050	ChinaAMC China 50 ETF	13.01	38,085,289,036.13
510060	ICBCCS SSE China Central Enterprises 50 ETF	8.35	222,993,731.63
510160	China Security Southern Well-off Industry Index ETF	7.35	837,739,227.93
510230	Guotai SSE 180 Finance ETF	6.76	3,907,590,530.78
510300	Huatai-PB CSI 300 ETF	5.66	20,321,022,282.29
510880	Huatai-PB SSE Dividend Index ETF	11.13	1,423,667,125.14

Table 10: Details of the FOF data set.

UNT	UPS	URI	USB	USM	UTI	UTL	VAC	VCRA	VG	VKQ	VLO	VLT	VLY	VMC	VNQ
VOC	VSH	VTR	VVI	VVR	WAB	WAL	WBK	WBS	WCC	WCG	WCN	WD	WEC	WES	WMK
WMS	WNC	WPX	WRI	WSM	WST	WTR	WU	WWW	WY	X	XIN	YELP	ZTR	ZTS	

Table 11: Codes of the stocks in the NYSE19 data set.

Table 11 gives the codes of the 47 stocks in the NYSE19 data set.

References

I. Aldridge. *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems*. Wiley, Hoboken, NJ, 2 edition, Apr. 2013.

A. Blum and A. Kalai. Universal portfolios with and without transaction costs. *Machine Learning*, 35(3):193–205, 1999.

W. F. M. D. Bondt and R. Thaler. Does the stock market overreact? *Journal of Finance*, 40(3):793–805, Jul. 1985.

A. Borodin, R. El-Yaniv, and V. Gogan. Can we learn to beat the best stock. *Journal of Artificial Intelligence Research*, 21(1):579–594, Jan. 2004.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2010.

J. Brodie, I. Daubechies, C. D. Giannone, and I. Loris. Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences of the United States of America*, 106(30):12267–12272, Jul. 2009.

W. Bruns and U. Vetter. *Determinantal Rings*. Springer-Verlag New York, Inc., 1988.

V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky. Rank-sparsity incoherence for matrix decomposition. *Siam Journal on Optimization*, 21(2):572–596, 2009.

T. M. Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, Jan. 1991.

P. Das, N. Johnson, and A. Banerjee. Online lazy updates for portfolio selection with transaction costs. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 202–208, 2013.

V. DeMiguel, L. Garlappi, and R. Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953, May 2009.

J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra. Efficient projections onto the ℓ^1-ball for learning in high dimensions. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2008.

M. Ho, Z. Sun, and J. Xin. Weighted elastic net penalized mean-variance portfolio design and computation. *SIAM Journal on Financial Mathematics*, 6(1), 2015.

D. Huang, J. Zhou, B. Li, S. C. H. Hoi, and S. Zhou. Robust median reversion strategy for online portfolio selection. *IEEE Transactions on Knowledge and Data Engineering*, 28 (9):2480–2493, Sep. 2016.

N. Jegadeesh. Evidence of predictable behavior of security returns. *Journal of Finance*, 45 (3):881–898, Jul. 1990.

N. Jegadeesh. Seasonality in stock price mean reversion: Evidence from the U.S. and the U.K. *Journal of Finance*, 46(4):1427–1444, Sep. 1991.

N. Jegadeesh and S. Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91, Mar. 1993.

P. Jorion. *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, New York, 1997.

D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292, Mar. 1979.

F. J. Király, L. Theran, and R. Tomioka. The algebraic combinatorial approach for low-rank matrix completion. *Journal of Machine Learning Research*, 16:1391–1436, 2015.

Z. R. Lai, D. Q. Dai, C. X. Ren, and K. K. Huang. Radial basis functions with adaptive input and composite trend representation for portfolio selection. *IEEE Transactions on Neural Networks and Learning Systems*, 29(12):6214–6226, Dec. 2018a.

Z. R. Lai, D. Q. Dai, C. X. Ren, and K. K. Huang. A peak price tracking based learning system for portfolio selection. *IEEE Transactions on Neural Networks and Learning Systems*, 29(7):2823–2832, Jul. 2018b.

Z. R. Lai, P. Y. Yang, L. Fang, and X. Wu. Short-term sparse portfolio optimization based on alternating direction method of multipliers. *Journal of Machine Learning Research*, 19(63):1–28, 2018c. URL `http://jmlr.org/papers/v19/17-558.html`.

O. Ledoit and M. Wolf. Spectrum estimation: A unified framework for covariance matrix estimation and pca in large dimensions. *Journal of Multivariate Analysis*, 135:360–384, Apr. 2015.

O. Ledoit and M. Wolf. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *The Review of Financial Studies*, 30(12):4349–4388, Dec. 2017.

B. Li and S. C. H. Hoi. On-line portfolio selection with moving average reversion. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2012.

B. Li and S. C. H. Hoi. Online portfolio selection: A survey. *ACM Computing Surveys (CSUR)*, 46(3):35:1–35:36, 2014.

B. Li, S. C. H. Hoi, D. Sahoo, and Z. Y. Liu. Moving average reversion strategy for on-line portfolio selection. *Artificial Intelligence*, 222:104–123, 2015.

B. Li, D. Sahoo, and S. C. H. Hoi. OLPS: a toolbox for on-line portfolio selection. *Journal of Machine Learning Research*, 17(1):1242–1246, 2016.

M. Magdon-Ismail and A. F. Atiya. Maximum drawdown. *Risk Magazine*, 10:99–102, 2004.

H. M. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, Mar. 1952.

H. M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press, 1959.

A. D. Roy. Safety first and the holding of assets. *Econometrica*, 20(3):431–449, Jul. 1952.

W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, Sep. 1964.

W. F. Sharpe. Mutual fund performance. *Journal of Business*, 39(1):119–138, Jan. 1966.

W. Shen, J. Wang, and S. Ma. Doubly regularized portfolio with risk minimization. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2014.

R. J. Shiller. *Irrational Exuberance*. Princeton University Press, Princeton, NJ, 2000.

R. J. Shiller. From efficient markets theory to behavioral finance. *Journal of Economic Perspectives*, 17(1):83–104, 2003.

J. L. Treynor and F. Black. How to use security analysis to improve portfolio selection. *Journal of Business*, 46(1):66–86, Jan. 1973.

Y. Vardi and C. H. Zhang. The multivariate ℓ^1-median and associated data depth. *Proceedings of the National Academy of Sciences of the United States of America*, 97(4): 1423–1426, Feb. 2000.

T. W. Young. Calmar ratio: A smoother tool. *Futures*, 20(1):40, 1991.

H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.