Article:

Li, Gaorong, Huang, Lei, Yang, Jin et al. (1 more author) (2021) A Synthetic Regression Model for Large Portfolio Allocation. Journal of Business and Economic Statistics. pp. 1665-1677. ISSN: 0735-0015

https://doi.org/10.1080/07350015.2021.1961787

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2021, VOL. 00, NO. 0, 1–13
https://doi.org/10.1080/07350015.2021.1961787

A Synthetic Regression Model for Large Portfolio Allocation

Gaorong Li[a], Lei Huang[b], Jin Yang[c], and Wenyang Zhang[d]

[a]School of Statistics, Beijing Normal University, Beijing, China; [b]School of Mathematics, Southwest Jiaotong University; [c]School of Statistics and Data Sciences, Nankai University, Tianjin, China; [d]Department of Mathematics, University of York, York, UK

ABSTRACT
Portfolio allocation is an important topic in financial data analysis. In this article, based on the mean-variance optimization principle, we propose a synthetic regression model for construction of portfolio allocation, and an easy to implement approach to generate the synthetic sample for the model. Compared with the regression approach in existing literature for portfolio allocation, the proposed method of generating the synthetic sample provides more accurate approximation for the synthetic response variable when the number of assets under consideration is large. Due to the embedded leave-one-out idea, the synthetic sample generated by the proposed method has weaker within sample correlation, which makes the resulting portfolio allocation more close to the optimal one. This intuitive conclusion is theoretically confirmed to be true by the asymptotic properties established in this article. We have also conducted intensive simulation studies in this article to compare the proposed method with the existing ones, and found the proposed method works better. Finally, we apply the proposed method to real datasets. The yielded returns look very encouraging.

ARTICLE HISTORY
Received January 2019
Accepted July 2021

KEYWORDS
Leave-one-out; Penalized least-square estimation; Portfolio allocation

1. Introduction

Portfolio allocation plays a key role in determining returns for an investment portfolio. It attempts to balance risk versus reward by adjusting the percentage of each asset in an investment portfolio. The Markowitz mean–variance portfolio theory, Markowitz (1952), is very influential in portfolio allocation. To form a portfolio allocation by the Markowitz formula, the covariance matrix of returns of the assets under consideration usually needs to be estimated, and its sample covariance matrix is usually taken as its estimator. When the number of assets under consideration is big, the sample covariance matrix may not work very well as the estimation errors would accumulate, in the formed portfolio allocation, very quickly to reach an unacceptable level, which makes the formed portfolio allocation performs poorly; see Fan, Fan, and Lv (2008), Basak, Jagannathan, and Ma (2009), DeMiguel, Garlappi, and Uppal (2009), Ledoit and Wolf (2017), and the references therein.

One cause of the poor performance of a portfolio allocation formed by the Markowitz formula is that the inverse of sample covariance matrix can be very poor when the size of the covariance matrix concerned is big, as an estimator of the inverse of a covariance matrix, which is the case in forming a portfolio allocation by the Markowitz formula. One approach to improve the performance is to find a better estimator for the inverse of the covariance matrix in the Markowitz formula. Over the past decades, there is much literature devoted to find more accurate estimation for high-dimensional covariance matrices, see Sun, Zhang, and Tong (2007), Fan, Fan, and Lv (2008), Bickel and Levina (2008a), Bickel and Levina (2008b), El Karoui (2008), Rothman, Levina, and Zhu (2009), Yuan (2010), Fan, Liao, and Mincheva (2011), Fan, Liao, and Micheva (2013), Berthet and Rigollet (2013), Birnbaum et al. (2013), Lam (2016), Guo, Box, and Zhang (2017), Ledoit and Wolf (2017), Avella-Medina et al. (2018), and the references therein.

With the improvement in the estimation of covariance matrices alone, we still cannot improve significantly the performance of a portfolio allocation formed by the Markowitz formula when the number of assets under consideration is big. Intuitively, this is understandable, because the return of a portfolio would be very unstable if every asset is included in the portfolio when the number of assets under consideration is very big. To make the return more stable, some assets have to be excluded from the portfolio, namely the vector of portfolio weights has to be sparse. This makes the idea very promising, that if we can transform the problem of portfolio allocation to a problem of regression, we may be able to find a better portfolio allocation by the penalized least-square estimation. This is exactly what we are going to do in this article.

The idea of applying regression models for portfolio allocation has appeared in the literature for many years. See, Britten-Jones (1999), Brodie et al. (2009), Ao, Li, and Zheng (2019), and the reference therein. The scaling involved in Britten-Jones (1999) can be very challenging, and the method in Brodie et al. (2009) is a constrained regression which is not very easy to implement. Ao, Li, and Zheng (2019) proposed a very interesting unconstrained regression representation for the mean-variance portfolio problem. Because there is no constraint attached with the regression model, the method in Ao, Li, and Zheng (2019) is easier to implement, and the methodology is more promising.

CONTACT Wenyang Zhang wenyang.zhang@york.ac.uk Department of Mathematics, University of York, York YO10 5DD, UK.
Lei Huang is co-first author.

The response in Ao, Li, and Zheng (2019) is set to be a constant rather than a variable, and that constant is an estimator of $\sigma(1+\theta)\theta^{-1/2}$, obtained by using all observations of the returns of assets concerned, where θ is the squared maximum Sharpe ratio and σ is the given risk constraint. Because the tth observation of their covariate is set to be the vector of returns of all assets concerned at time point t, their response is a function of the observations of their covariate at all time points, and free of time. This is not a good idea as it creates within sample correlation. In addition to that, their method doesn't apply to real high dimensional cases where the number of assets concerned is larger than the sample size. This is because they have to have the inverse of the sample covariance matrix of the vector of returns of assets concerned, in order to get the response, and the inverse of that sample covariance matrix does not exist for real high dimensional cases.

In this article, based on the basis of unconstrained regression representation for the mean-variance portfolio problem in Ao, Li, and Zheng (2019), we propose a synthetic regression model for large portfolio allocation. We embed a leave-one-out idea in the generation of synthetic response variable, which is intuitively more reasonable. We also borrow the idea in Fan, Fan, and Lv (2008) to apply the Fama–French factor models, Fama and French (1993), to derive a structure for the covariance matrix of the vector of returns of assets concerned, and estimate the covariance matrix based on the derived structure. The proposed method applies to the cases where the number of assets concerned is larger than the sample size, and performs well. Indeed, both our simulation results and real data analysis show our proposed method outperforms the commonly used methods, which include MAXSER, proposed in Ao, Li, and Zheng (2019), see Sections 4 and 5.

The rest of this article is organized as follows. We begin in Section 2 with a detailed description of the proposed synthetic regression model for large portfolio allocation. In Section 3, the asymptotic properties of the portfolio allocation formed by the proposed synthetic regression model are presented to justify the proposed methodology theoretically. Intensive simulation studies are conducted in Section 4 to show how well the portfolio allocation formed by the proposed synthetic regression model works, compared with other existing portfolio allocation approaches. In Section 5, we apply the portfolio allocation, formed by the proposed synthetic regression model, to datasets which are freely available from the home page of Kenneth R. French,[1] and compare its returns with that of some commonly used approaches. Finally, we conclude the article by Section 6. We leave the technical conditions and theoretical proofs of all asymptotic properties in the appendix.

2. Estimation of Optimal Large Portfolio Allocation

Suppose $(\mathbf{X}_i^{\mathrm{T}}, \mathbf{Y}_i^{\mathrm{T}})$, $i = 1, \ldots, n$, is a sample from $(\mathbf{X}^{\mathrm{T}}, \mathbf{Y}^{\mathrm{T}})$, where $\mathbf{Y}$ is a p_n dimensional vector and $\mathbf{X}$ is a q dimensional factor. An underlying assumption throughout this article is that $p_n/n \longrightarrow \infty$ when $n \longrightarrow \infty$, and q is fixed.

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

As far as this article is concerned, $\mathbf{Y}$ can be more specifically defined as the vector of returns of p_n assets concerned, based on the Fama-French factor models, we can reasonably assume

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \boldsymbol{\epsilon}, \quad E(\boldsymbol{\epsilon}|\mathbf{X}) = \mathbf{0}, \quad \mathrm{cov}(\boldsymbol{\epsilon}|\mathbf{X}) = \Sigma_0, \tag{1}$$

where $\mathbf{A}$ is a $p_n \times q$ matrix of factor loadings, $\boldsymbol{\epsilon}$ is a $p_n \times 1$ vector of idiosyncratic errors, and Σ_0 is a diagonal matrix.

Model (1) is the model we assume for $\mathbf{Y}$ in this article. It is the base for us to construct the estimator of the needed covariance matrix of $\mathbf{Y}$ in portfolio allocation when the number of assets concerned, p_n, is much larger than the sample size n.

2.1. Optimal Portfolio Allocation

We first present a result from Ao, Li, and Zheng (2019), which gives the theoretical optimal portfolio allocation.

Let

$$\boldsymbol{\mu} = E(\mathbf{Y}), \quad \mathrm{cov}(\mathbf{Y}) = \Sigma, \quad \theta = \boldsymbol{\mu}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu},$$

where θ is the squared maximum Sharpe ratio. Ao, Li, and Zheng (2019) have shown the optimal portfolio allocation $\mathbf{w}$ subject to

$$\mathrm{var}(\mathbf{w}^{\mathrm{T}}\mathbf{Y}) \leq \sigma^2$$

is the minimizer of

$$E\left(\sigma(1+\theta)\theta^{-1/2} - \mathbf{w}^{\mathrm{T}}\mathbf{Y}\right)^2, \tag{2}$$

where σ is the given risk constraint. See Ao, Li, and Zheng (2019) for more details.

Equation (2) is the basis of unconstrained regression representation for mean–variance portfolio problem. Based on Equation (2), Ao, Li, and Zheng (2019) applied the idea of the penalized least-square estimation to get an estimated optimal large portfolio allocation $\hat{\mathbf{w}}$ by minimizing

$$\sum_{i=1}^{n}\left(\sigma(1+\hat{\theta})\hat{\theta}^{-1/2} - \mathbf{w}^{\mathrm{T}}\mathbf{Y}_i\right)^2 \tag{3}$$

subject to

$$\|\mathbf{w}\|_1 \leq \delta,$$

where

$$\hat{\theta} = n^{-1}\left\{(n - p_n - 2)\hat{\theta}_s - p_n\right\}$$

and $\hat{\theta}_s$ is the estimator of θ, obtained by simply replacing $\boldsymbol{\mu}$ and Σ in θ by the sample mean and sample covariance matrix of $\{\mathbf{Y}_i, i = 1, \cdots, n\}$.

Notice that $\hat{\theta}$ may take negative values, which is not reasonable as an estimator of θ. To overcome this problem, Kan and Zhou (2007) made an adjustment on $\hat{\theta}$. Ao, Li, and Zheng (2019) suggested using the adjusted estimator proposed in Kan and Zhou (2007) rather than $\hat{\theta}$ when it comes to implementation of their method.

In the regression model (3), the response variable is $\sigma(1+\hat{\theta})\hat{\theta}^{-1/2}$, which does not depend on i, namely a constant, and is obtained by using all observations of the returns of assets concerned. On the other hand, the ith observation $\mathbf{Y}_i$ of the covariate is the vector of returns of assets concerned at time point i. Theoretically speaking, the response variable here is a function of the observations of the covariate at all time points, and is free of time. Intuitively, this would create within sample correlation and affect the performance of the resulting portfolio allocation.

Another problem with Equation (3) is that the response variable $\sigma(1+\hat{\theta})\hat{\theta}^{-1/2}$ involves the inverse of sample covariance matrix of $\mathbf{Y}_i$, $i = 1, \ldots, n$. When p_n is larger than n, the inverse of the sample covariance matrix would not exist, therefore, the response variable would not be available. So, the portfolio allocation proposed in Ao, Li, and Zheng (2019) would not apply to real large portfolio allocation problem.

To overcome the problems mentioned above, we propose a synthetic regression model for large portfolio allocation.

2.2. A Synthetic Regression Model For Large Portfolio Allocation

The proposed synthetic regression model is still based on Equation (2). However, $\mathbf{Y}_i$ is excluded to reduce within sample correlation when generating the ith observation of the response variable. Furthermore, we estimate the covariance matrix of $\mathbf{Y}$ based on model (1), which makes the inverse of the estimated covariance matrix available, therefore, makes the proposed synthetic regression model work for the construction of real large portfolio allocation.

2.2.1. Estimation of Σ

We first present the estimation of the covariance matrix Σ of $\mathbf{Y}$, because it is involved in the response variable of the proposed synthetic regression model.

Based on Equation (1), by simple calculation, we have

$$\Sigma = \mathbf{A}\Sigma_x\mathbf{A}^{\mathrm{T}} + \Sigma_0, \tag{4}$$

where $\Sigma_x = \mathrm{cov}(\mathbf{X})$. To get the estimator of Σ, we only need to get the estimators of $\mathbf{A}$, Σ_x and Σ_0.

Applying the standard least-square estimation, we can get the estimator $\hat{\mathbf{A}}$ of $\mathbf{A}$ by minimizing

$$\sum_{i=1}^{n} \|\mathbf{Y}_i - \mathbf{A}\mathbf{X}_i\|^2.$$

By simple calculation, we have

$$\hat{\mathbf{A}} = \boldsymbol{\mathcal{Y}}^{\mathrm{T}}\boldsymbol{\mathcal{X}}(\boldsymbol{\mathcal{X}}^{\mathrm{T}}\boldsymbol{\mathcal{X}})^{-1},$$
$$\boldsymbol{\mathcal{X}} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)^{\mathrm{T}},$$
$$\boldsymbol{\mathcal{Y}} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)^{\mathrm{T}}.$$

Furthermore, based on the residual sum squares, we use

$$\hat{\Sigma}_0 = \mathrm{diag}\left(\hat{\boldsymbol{\epsilon}}_1^2, \ldots, \hat{\boldsymbol{\epsilon}}_{p_n}^2\right)$$

to estimate Σ_0, where $\hat{\boldsymbol{\epsilon}}_i^2$ is the ith element on the diagonal of the matrix

$$\frac{1}{n-q}\sum_{k=1}^{n}(\mathbf{Y}_k - \hat{\mathbf{A}}\mathbf{X}_k)(\mathbf{Y}_k - \hat{\mathbf{A}}\mathbf{X}_k)^{\mathrm{T}}.$$

Because the dimension of $\mathbf{X}$ is usually small, for example, it is $q = 3$ for the Fama–French three-factor models, therefore, we can simply use the sample covariance matrix of $\mathbf{X}_i$, $i = 1, \ldots, n$, to estimate Σ_x, namely the estimator of Σ_x is taken to be

$$\hat{\Sigma}_x = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^{\mathrm{T}}, \quad \bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i.$$

Finally, we use

$$\hat{\Sigma} = \hat{\mathbf{A}}\hat{\Sigma}_x\hat{\mathbf{A}}^{\mathrm{T}} + \hat{\Sigma}_0$$

to estimate Σ.

2.2.2. A Synthetic Regression Model

Let $\hat{\Sigma}^{\backslash i}$ be the estimator of Σ, obtained by the method in Section 2.2.1, without using the ith observation, and $\bar{\mathbf{Y}}^{\backslash i}$ be the sample mean of $\mathbf{Y}_k$, $k = 1, \ldots, i-1, i+1, \ldots, n$. Let

$$z_i = \sigma(1+\theta_i)\theta_i^{-1/2}, \quad \theta_i = (\bar{\mathbf{Y}}^{\backslash i})^{\mathrm{T}}(\hat{\Sigma}^{\backslash i})^{-1}\bar{\mathbf{Y}}^{\backslash i}. \tag{5}$$

Treating $(z_i, \mathbf{Y}_i^{\mathrm{T}})$, $i = 1, \ldots, n$, as a synthetic sample, we propose the following synthetic regression model:

$$z_i = \mathbf{Y}_i^{\mathrm{T}}\mathbf{w} + e_i, \quad i = 1, \ldots, n, \tag{6}$$

for estimating the minimizer of Equation (2).

Due to the high dimensionality of $\mathbf{Y}$ in large portfolio allocation, we apply the penalized least-square estimation to the synthetic regression model (6) to estimate $\mathbf{w}$, that is the estimated optimal large portfolio allocation, $\hat{\mathbf{w}}$, is taken to be the minimizer of

$$\frac{1}{2n}\sum_{i=1}^{n}\left(z_i - \mathbf{Y}_i^{\mathrm{T}}\mathbf{w}\right)^2 + \lambda\|\mathbf{w}\|_1, \tag{7}$$

where λ is a tuning parameter, and

$$\mathbf{w} = (w_1, \ldots, w_{p_n})^{\mathrm{T}}, \quad \|\mathbf{w}\|_1 = \sum_{i=1}^{p_n}|w_i|.$$

Our proposed large portfolio allocation is this estimated optimal large portfolio allocation $\hat{\mathbf{w}}$, we term it SRM.

The tuning parameter λ in Equation (7) can be chosen by cross-validation (CV). Indeed, in the simulation studies and real data analysis in this article, we use the 10-fold CV to select this tuning parameter.

3. Asymptotic Properties

In this section, we are going to build asymptotic theory to justify our proposed portfolio allocation. We first introduce some notations. Let $S = \mathrm{supp}(\mathbf{w}^*)$ be the support of the true optimal large portfolio allocation $\mathbf{w}^*$, and S^c be its complement, where $\mathbf{w}^* = \frac{\sigma}{\sqrt{\theta}}\Sigma^{-1}\boldsymbol{\mu}$ is the minimizer of Equation (2). Let $s_n = |S|$ be the cardinality of the set S. In order to establish the asymptotic theory, we need the following regularity assumptions.

Assumption 1. We assume $\mathbf{Y} \sim N(\boldsymbol{\mu}, \Sigma)$, and there exists some positive constants $L < \infty$ and $M < \infty$ such that $\max\left\{\boldsymbol{\mu}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}, \max_{1\le j\le p_n}|\mu_j|\right\} \le L$ and $\max_{1\le j\le p_n}|\sigma_{jj}| \le M$, where μ_j is the jth component of $\boldsymbol{\mu}$ and $\Sigma = (\sigma_{ij})_{1\le i,j\le p_n}$.

Assumption 2. For some constants $\alpha \geq 1$ and $\phi_0 > 0$, we define the set $\mathcal{T}(S,\alpha) = \{\boldsymbol{\delta} \in \mathbb{R}^{p_n}, \|\boldsymbol{\delta}_{S^c}\|_1 \leq \alpha \|\boldsymbol{\delta}_S\|_1\}$, and assume that the $p_n \times p_n$ covariance matrix Σ satisfies

$$\phi_0^2 = \phi_0^2(S,\alpha) = \min_{\boldsymbol{\delta} \neq 0, \boldsymbol{\delta} \in \mathcal{T}(S,\alpha)} \frac{\boldsymbol{\delta}^{\mathrm{T}} \Sigma \boldsymbol{\delta}}{\|\boldsymbol{\delta}_S\|_2^2} > 0.$$

Assumption 3. The number of factors, q, is bounded, and $p_n^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{A} \to \Omega$ as $n \to \infty$, Ω is a $q \times q$ symmetric positive semidefinite matrix.

Assumption 4. Assume that $s_n^{3/2} \log p_n/n \to 0$ as $n \to \infty$.

Assumption 1 is a mild technical condition that facilitates the proofs of the main theorems, and similar assumption can be found in Ao, Li, and Zheng (2019). In practice, our proposed procedure can deal with returns with heavier-tailed distribution numerically. Assumption 2 is the restricted eigenvalue condition (REC) introduced in Bickel, Ritov, and Tsybakov (2009), and this assumption is often used to derive the oracle inequalities for the Lasso estimator and Dantzig selector (see the details in Candès and Tao (2007), Bickel, Ritov, and Tsybakov (2009), and Raskutti, Wainwright, and Yu (2010)). Assumption 3 is used in Fan, Fan, and Lv (2008) and Fan, Liao, and Mincheva (2011) to establish the asymptotic properties of the covariance estimator. Assumptions 4 is used to show the asymptotic properties of the proposed portfolio allocation, and this assumption is stronger than that in Meinshausen and Yu (2009) because we require the optimal estimation rate of $\theta = \boldsymbol{\mu}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}$. Bunea, Tsybakov, and Wegkamp (2007), van de Geer (2006), and Zou, Ke, and Zhang (2020) also used the sparsity condition to derive the consistency of the Lasso estimator in linear model and generalized linear model respectively, but they don't need to estimate $\theta = \boldsymbol{\mu}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}$. Fan, Weng, and Zhou (2021) provided the similar sparsity $\|\Sigma^{-1}\boldsymbol{\mu}\|_0 \leq s_n$ and $s_n \log p_n/n = o(1)$ to derive the minimax estimation rate of $\theta = \boldsymbol{\mu}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}$, where $\|\mathbf{a}\|_0 = \sum_{i=1}^{p_n} |a_i|^0$ with convention $0^0 = 0$ and $\mathbf{a} = (a_1, \ldots, a_{p_n})^{\mathrm{T}} \in \mathbb{R}^{p_n}$.

Theorem 1. Under Assumptions 1–4, if the tuning parameter $\lambda \asymp (s_n \log p_n/n) \vee \sqrt{\log p_n/n}$, we have

$$\left|\hat{\mathbf{w}}^{\mathrm{T}}\boldsymbol{\mu} - \sigma\theta^{1/2}\right| = O_p(\lambda s_n^{1/2}).$$

Theorem 1 shows that the mean of the return of the proposed portfolio tends, with rate $\lambda s_n^{1/2}$, to the maximum one can get under the risk constraint $\mathrm{var}(\mathbf{w}^{\mathrm{T}}\mathbf{Y}) \leq \sigma^2$.

Theorem 2. Under the conditions of Theorem 1, we have

$$\left|\hat{\mathbf{w}}^{\mathrm{T}}\Sigma\hat{\mathbf{w}} - \sigma^2\right| = O_p(\lambda s_n^{1/2}).$$

Theorem 2 shows the variance of the proposed portfolio tends, with rate $\lambda s_n^{1/2}$, to σ^2 which is the maximum risk allowed. This together with Theorem 1 show the proposed portfolio allocation is asymptotically equal to the theoretical optimal portfolio allocation.

4. Simulation Studies

The performances of the proposed SRM portfolio and various benchmark strategies will be examined and compared in this section. Since it has been demonstrated that the MAXSER method proposed by Ao, Li, and Zheng (2019) outperforms other strategies, it would be quite interesting to see whether the SRM approach is better or not than MAXSER under similar settings. More specifically, both stocks and factors are used in the simulated asset pool, the way to generate the returns are described in Section 4.2.

Table 1. Portfolios under comparison and their abbreviations.

Portfolio	Abbreviation
Synthetic regression model	SRM
SRM without leave-one-out	$\mathrm{SRM}^{-\mathrm{LOO}}$
Maximum Sharpe ratio estimated and sparse regression	MAXSER
MAXSER with leave-one-out	$\mathrm{MAXSER}^{+\mathrm{LOO}}$
MV with nonlinear shrinkage cov	MV-NLS
MV-NLS with short-sale constraint and CV	MV-NLS-SSCV
MV-NLS with ℓ_1 constraint and CV	MV-NLS-L1CV

NOTE: "MV" represents for mean-variance portfolio, "CV" means cross-validation.

4.1. Portfolios Under Comparison

To demonstrate how well the proposed SRM portfolio works, we are going to compare the SRM portfolio with other portfolio allocation strategies including MAXSER in details, and portfolios under comparison are listed and annotated in Table 1. The portfolio "MAXSER" represents the method proposed by Ao, Li, and Zheng (2019). For other portfolios, they are formed by replacing the covariance matrices in MV with their various estimators, such as nonlinear shrinkage estimator, see Ledoit and Wolf (2004, 2017) for details.

The portfolios with either a short-sale or ℓ_1-norm constraint on the portfolio weights are also formed. For examples, "MV-NLS-SSCV" stands for the MV portfolio with nonlinear shrinkage covariance estimator and a short-sale constraint on the portfolio weights, while "MV-NLS-L1CV" means imposing an ℓ_1-norm constraint on its weights. These portfolios and MAXSER portfolio enjoy the same benefit in terms of risk control as our SRM portfolio does. Because one of the main adjustments in SRM compared to MAXSER is the leave-one-out method, it is of interest to check whether MAXSER can be improved by applying leave-one-out method, and if SRM really benefits from leave-one-out method. Thus, we also compare SRM without leave-one-out ($\mathrm{SRM}^{-\mathrm{LOO}}$) and MAXSER with leave-one-out ($\mathrm{MAXSER}^{+\mathrm{LOO}}$). By making such comparison, we can reveal that the advantages of SRM essentially come from its methodology and ideas.

4.2. Parameter Setting

The proposed SRM method applies directly to high dimensional cases where $p_n > n$. Although the MAXSER assumes that $p_n < n$, but it can also apply to $p_n > n$ after subpool selection. Thus, in the simulation studies, to make the comparison complete and fair, we consider two scenarios including both $p_n < n$ and $p_n > n$. We will see that the proposed SRM method outperforms MAXSER under each scenario.

To make our simulations more realistic, all parameters are set based on real data. Specifically, in our data generation, the parameters such as the mean $\boldsymbol{\mu}_x = E(\mathbf{X})$ and covariance matrix $\Sigma_x = \mathrm{cov}(\mathbf{X})$ are set to be the sample mean and sample covariance matrix of the monthly returns of the Fama-French Three Factors (FF3) from 2007 to 2019, respectively. To set the loading matrix $\mathbf{A}$, $p_n = 100$ stocks are randomly selected from those in the *S&P* 500 index for the entire period 2007 to 2019. By regression of the monthly excess returns of each selected stock on the returns of FF3, each row of the loading matrix $\mathbf{A}$ is set to be the coefficients of each regression. We generate the returns, $\mathbf{Y}_i$s, through (1) with $\boldsymbol{\epsilon}_i$ being generated from $N(\mathbf{0}_{p_n},\ 0.155\mathbf{I}_{p_n})$ and $\mathbf{X}_i$s from $N(\boldsymbol{\mu}_x,\ \Sigma_x)$, $\mathbf{0}_{p_n}$ is a p_n-dimensional vector with each component being 0, $\mathbf{I}_{p_n}$ is an identity matrix of size p_n. We set the level of risk constraint to be $\sigma = 0.04$ across all simulations.

4.3. Comparisons

In the simulations, the Fama–French three factors are used as $\mathbf{X}_i$ in Model (1) of Section 2, meaning that the factors are only applied to estimate Σ_x in Equation (4), not being considered as portfolios in the full asset pool.

We set the sample size to be $n = 120$ ($p_n < n$) and $n = 72$ ($p_n > n$), and for each scenario, we do $L = 1000$ simulations to evaluate the portfolio performance in terms of risk and Sharpe ratio. The results for both $n = 120$ and $n = 72$ are presented in Table 2. Even the $\{n = 120, p_n = 100\}$ scenario means quite large dimensionality for MAXSER, to make MAXSER work better, the subpool selection proposed by Ao, Li, and Zheng (2019) is implemented for MAXSER, and the subpool size is 50 by default according to Ao, Li, and Zheng (2019). Because SRM applies well to high dimensional cases, thus the subpool selection is not implemented for SRM hereafter.

The risks and Sharpe ratios in Table 2 are obtained as follows: for each simulation, say the ℓth simulation, based on the generated data, a portfolio allocation $\hat{\mathbf{w}}_{<\ell>}$ is formed by each of the methods under comparison. The conditional mean and variance of the portfolio $\hat{\mathbf{w}}_{<\ell>}$, given the data, are $\hat{\mathbf{w}}^{\mathrm{T}}_{<\ell>}\boldsymbol{\mu}$ and $\hat{\mathbf{w}}^{\mathrm{T}}_{<\ell>}\Sigma\hat{\mathbf{w}}_{<\ell>}$, where $\boldsymbol{\mu}$ and Σ are the true mean and covariance matrix of the vector of the asset returns. The risk of this portfolio is defined as the average of its conditional standard deviations over the L simulations, namely $\frac{1}{L}\sum_{\ell=1}^{L}\sqrt{\hat{\mathbf{w}}^{\mathrm{T}}_{<\ell>}\Sigma\hat{\mathbf{w}}_{<\ell>}}$, where $L = 1000$, and its Sharpe ratio is the averge of its conditional Sharpe ratios over the L simulations. Values in the brackets are standard deviation over L simulations.

Table 2 shows that the risk of the SRM portfolio is more close to the given constraint than any strategy of portfolio allocation under comparison. Besides, it can be seen that, the leave-one-out method improves both SRM and MAXSER to some extent. When the sample size $n = 120$, the Sharpe ratio of SRM reaches approximately 63.3% of the theoretical maximum of the Sharpe ratio on average, while the MAXSER portfolio only reaches 57.4%. When the sample size n equals to 72, which is the scenario of $p_n > n$, the Sharpe ratio of the SRM portfolio still outperforms the others.

Moreover, we also examine the performances of candidate portfolios without assuming the exact factor structure. Here, we generate the returns, $\mathbf{Y}_i$'s, from multivariate normal distribution with parameters $\boldsymbol{\mu}_y$ and Σ_y, which are set to be the sample mean and sample covariance matrix of the 100 stocks. The results are presented in Table 3, which shows that the SRM still outperforms MAXSER in this situation.

Because both SRM and MAXSER are developed for high-dimensional situation with assumptions on sparsity of optimal allocation $\mathbf{w}^*$, we conduct another simulation study by letting

$$(w_1, \ldots, w_d, 0, \ldots, 0)_{p\times 1} = C_0\Sigma_y^{-1}\boldsymbol{\mu}_{y0},$$

from which we can obtain $\boldsymbol{\mu}_{y0}$. Then, we generate the returns, $\mathbf{Y}_i$s, from multivariate normal distribution with parameters $\boldsymbol{\mu}_{y0}$ and Σ_y, which ensures that the theoretical allocation $\mathbf{w}^*$ is sparse. Here we choose $d = 30$, the $\{w_j, 1 \leq j \leq d\}$ come from uniform distribution $U(0, 1)$, C_0 is a constant to make $\boldsymbol{\mu}_{y0}$ be relatively close to the sample mean $\boldsymbol{\mu}_y$. In our simulation, we choose $C_0 = 1/500$. The results in Table 4 are consistent to Table 3, which shows that the SRM methods is better than MAXSER under sparsity condition of allocations $\mathbf{w}^*$.

Moreover, to test the robustness of the proposed SRM method, we have also conducted a simulation where Σ_x in Section 4.2 is misspecified. More specifically, in Case I, we generate the returns $\mathbf{Y}_i$'s based on Fama and French 3 factors, but using Carhart-4 factors (Fama and French 3 factors plus a Momentum factor) to construct the portfolio; in Case II, we generate the returns $\mathbf{Y}_i$'s based on Carhart-4 factors, but using Fama and French 3 factors to construct the portfolio; in Case III, we generate the returns $\mathbf{Y}_i$'s based on Fama and French 3 factors, but using Fama and French 5 factors to construct the portfolio; in Case IV, we generate the returns $\mathbf{Y}_i$'s based on Fama and French 5 factors, but using Fama and French 3 factors to

Table 2. Risks and Sharpe Ratios of candidate portfolios.

Normal distribution	$\sigma = 0.04$			
	$n = 120$		$n = 72$	
Portfolio	Risk	S-Ratio	Risk	S-Ratio
SRM	0.042(0.004)	**1.193**(0.248)	0.042(0.005)	**1.109**(0.320)
SRM^{-LOO}	0.043(0.004)	1.179(0.251)	0.043(0.005)	1.105(0.322)
MAXSER	0.044(0.005)	1.080(0.303)	0.045(0.006)	0.953(0.391)
MAXSER^{+LOO}	0.043(0.005)	1.161(0.297)	0.044(0.006)	0.994(0.384)
MV-NLS	0.055(0.016)	0.947(0.185)	0.058(0.021)	0.833(0.239)
MV-NLS-SSCV	0.036(0.024)	0.848(0.257)	0.034(0.031)	0.750(0.331)
MV-NLS-L1CV	0.028(0.012)	0.975(0.168)	0.020(0.015)	0.867(0.195)

NOTES: "S-Ratio" represents for Sharpe Ratio hereafter. The risk constraint is set to 0.04. The theoretical maximum Sharpe ratio is 1.881. The average value and standard deviation (in parentheses) of each performance measure are reported.

Table 3. Risks and Sharpe Ratios of candidate portfolios without factor structure.

Normal Distribution	$\sigma = 0.04$			
	$n = 120$		$n = 72$	
Portfolio	Risk	S-Ratio	Risk	S-Ratio
SRM	0.040(0.004)	**1.061**(0.276)	0.041(0.005)	**0.990**(0.356)
SRM^{-LOO}	0.040(0.004)	1.050(0.239)	0.041(0.005)	0.983(0.306)
MAXSER	0.041(0.005)	0.878(0.246)	0.042(0.006)	0.784(0.318)
MAXSER^{+LOO}	0.041(0.005)	0.932(0.272)	0.042(0.006)	0.796(0.351)
MV-NLS	0.053(0.017)	0.770(0.149)	0.056(0.023)	0.688(0.192)
MV-NLS-SSCV	0.035(0.025)	0.689(0.207)	0.033(0.033)	0.615(0.268)
MV-NLS-L1CV	0.029(0.012)	0.793(0.133)	0.022(0.016)	0.711(0.152)

NOTE: "S-Ratio" represents for Sharpe Ratio hereafter. The risk constraint is set to 0.04. The theoretical maximum Sharpe ratio is 1.670. The average value and standard deviation (in parentheses) of each performance measure are reported.

Table 4. Risks and Sharpe Ratios of candidate portfolios without factor structure and with sparsity.

Normal Distribution		$\sigma = 0.04$		
		$n = 120$		$n = 72$
Portfolio	Risk	S-Ratio	Risk	S-Ratio
SRM	0.039(0.004)	**0.620**(0.263)	0.040(0.005)	**0.584**(0.331)
SRM^{-LOO}	0.040(0.004)	0.608(0.227)	0.041(0.006)	0.579(0.280)
MAXSER	0.040(0.004)	0.581(0.231)	0.042(0.006)	0.488(0.288)
MAXSER^{+LOO}	0.040(0.005)	0.595(0.259)	0.041(0.005)	0.502(0.323)
MV-NLS	0.054(0.017)	0.451(0.134)	0.057(0.022)	0.406(0.165)
MV-NLS-SSCV	0.034(0.024)	0.403(0.192)	0.032(0.032)	0.363(0.237)
MV-NLS-L1CV	0.030(0.012)	0.464(0.120)	0.021(0.016)	0.420(0.135)

NOTES: "S-Ratio" represents for Sharpe Ratio hereafter. The risk constraint is set to 0.04. The theoretical maximum Sharpe ratio is 0.973. The average value and standard deviation (in parentheses) of each performance measure are reported.

construct the portfolio. These misspecified cases include both missing factors and useless factors.

In the following simulation, dataset of sample size $n + 1$ is generated, the first n observations are used as training dataset to form a portfolio allocation $\hat{\mathbf{w}}_n$, the $(n + 1)$th observation serves for the computation of the return of the formed portfolio, that is, the return of the formed portfolio is $\hat{\mathbf{w}}_n^{\mathrm{T}}\mathbf{Y}_{n+1}$. We still do $L = 1000$ simulations and risk constraint is still set to be 0.04. We use $r_{n+1,\ell}$ to denote the return of a portfolio in the ℓth simulation, and call $\{r_{n+1,\ell}, \ell = 1, \ldots, L\}$ the out-of-sample returns of this portfolio. The mean return and Sharpe ratio of this portfolio are calculated through

$$\bar{r} = \frac{1}{L}\sum_{\ell=1}^{L} r_{n+1,\ell}, \quad SR = \frac{(L-1)^{1/2}\bar{r}}{\left\{\sum_{\ell=1}^{L}(r_{n+1,\ell} - \bar{r})^2\right\}^{1/2}}. \tag{8}$$

To compare the proposed SRM method and MAXSER, we conduct the paired Sharpe ratio tests, see Ledoit and Wolf (2008), the null hypothesis is

$$H_0 : Sr_s < Sr_m. \tag{9}$$

Based on the out-of-sample returns of SRM portfolio and MAXSER portfolio, (9) can be tested, where Sr_s is the Sharpe ratio of SRM portfolio, Sr_m is the Sharpe ratio of the MAXSER portfolio. The p-values under all four cases are presented in Table 5, where the p-value, under every case, is very close to 0. This means the proposed SRM method is significantly better than MAXSER even when the structure of Σ_x is misspecified to some extent.

Table 5. The Sharpe Ratio tests between SRM and MAXSER.

Normal Distribution: $\sigma = 0.04$	p-value
Case I: FF3-Carhart4	1.3×10^{-4}
Case II: Carhart4-FF3	2.9×10^{-4}
Case III: FF3-FF5	3.0×10^{-4}
Case IV: FF5-FF3	2.8×10^{-4}

NOTE: "FF3-Carhart4" means that we generate the returns $\mathbf{Y}_i$s based on Fama and French 3 factors, but using Carhart-4 factors to construct the portfolio. Others can be similarly explained.

5. Real Data Analysis

In this section, we are going to use five real datasets to illustrate how to use the proposed SRM method and how well it works in practice. Because our simulation studies in Section 4 have shown the performances of all the seven portfolios in the comparison, in the sake of consistency, we also primarily focus on applying the seven portfolio allocation strategies to the real datasets and compare the obtained results. The datasets for us to study are downloaded from the home page of Kenneth R. French.[2] Specifically, four pools of portfolios are downloaded from this website, and each pool consists of monthly returns of p_n (100 or 49) portfolios from June 1990 to May 2020. The time span is in total 360 months, that is, 30 years. Each of the 100 portfolios in the first pool is formed by the two factors: Size and Book-to-Market ratio. We denote this pool of portfolios by Pool A hereafter. Each of the 100 portfolios in the second pool is formed by Size and Investment. We denote this pool of portfolios by Pool B. The third pool consists of 100 portfolios formed by Size and Operating Profit, denoted by Pool C. The forth pool includes the 49 industry portfolios, denoted by Pool D. The last one, Pool E, represents the first 100 available stocks of Standard Poor's list by alphabetical order of their abbreviations. The Fama-French three factors of the same period are also downloaded as the factors $\mathbf{X}_i$ in Model (1) of Section 2, meaning that the three factors are only used to estimate Σ_x in (4), not being considered as portfolios in any pool.

In the downloaded datasets, there are very few observations unavailable (less than 0.15%), they are assigned as −99.99 in the original dataset, we recode them as 0 in our analysis. The moving average approach could also be used for the imputation of the unavailable observations, however we find it makes little difference to setting them to be 0.

In real stock market, the gold standard for evaluating different strategies of portfolio allocation is based on their out-of-sample returns. Therefore, we start with splitting the whole dataset to two parts, the first part is from June 1990 to May 2000, called training set, it has 120 months. The second part is from June 2000 to May 2020, called test set, it has 240 months. For each portfolio allocation under comparison, we compute its return at each month in the test set, and its risk and Sharpe ratio are computed based its returns at the 240 months in the test set. The return of each portfolio allocation under comparison at each month in the test set is computed based on the rolling window approach, namely, we form the portfolio allocation based on the data in the first 120 months, which is the training set, and compute its return at month $t = 121$, which is the first month in the test set. We then roll the training data by one month, that is to form the portfolio allocation based on the data from month $t = 2$ to month $t = 121$, and compute its return at month $t = 122$. We continuously do this until the return of the portfolio allocation at the last month is obtained. This way, the return of the portfolio allocation at each month in the test set is obtained.

As did in simulations studies, we also compare different portfolios when $n = 72$, which is a real high dimensional case for $p_n = 100$. Similarly, we split the whole dataset into

[2] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

training set and test set, where the rolling window approach is also applied. The initial training set consists of the first 6 years' data($n = 72$), the test set has 24×12 months. Moreover, following Engle, Ferstenberg, and Russell (2012), the portfolio return net of transaction costs in each period is computed as follows:

$$r_{\text{net}}(t) = \Big(1 - \sum_j c_{t,j} |w_j(t+1) - w_j(t+)|\Big)\Big(1 + r(t)\Big) - 1, \quad (10)$$

where $w_j(t+1)$ is the weight on asset j at the beginning of period $t+1$, $w_j(t+)$ is the weight of the same asset at the end of period t, $c_{t,j}$ is a cost level and $r(t)$ is the portfolio return without transaction cost at period t. For the cost level $c_{t,j}$, Ao, Li, and Zheng (2019) set it to be constant 0.1% from 1991 to 2016. Since most assets are portfolios in our empirical analysis, we set it to be 0.4% throughout the empirical analysis.

The risk and Sharp ratio of each portfolio allocation under each situation is presented in Tables 6 to 10.

Some conclusions can be drawn from Tables 6 to 8. First, since the portfolios in these Pools are formed by pairs of Fama and French factors, the covariance decomposition of Equation (4) is easy to be satisfied, thus the performances of SRM is always

Table 6. Risks and Sharpe Ratios of candidate portfolios for Pool A.

$\sigma = 0.04$	Without transaction cost				With transaction cost			
	$n = 120$		$n = 72$		$n = 120$		$n = 72$	
Portfolio	Risk	S-Ratio	Risk	S-Ratio	Risk	S-Ratio	Risk	S-Ratio
SRM	0.043	**0.308**	0.048	**0.260**	0.043	**0.306**	0.048	**0.258**
SRM^{-LOO}	0.044	0.304	0.049	0.258	0.044	0.301	0.049	0.257
MAXSER	0.046	0.224	0.055	0.145	0.046	0.223	0.055	0.144
MAXSER^{+LOO}	0.045	0.240	0.054	0.151	0.045	0.239	0.054	0.149
MV-NLS	0.058	0.196	0.069	0.127	0.058	0.195	0.070	0.126
MV-NLS-SSCV	0.038	0.176	0.036	0.114	0.038	0.176	0.036	0.113
MV-NLS-L1CV	0.029	0.202	0.023	0.131	0.029	0.201	0.024	0.130

Table 7. Risks and Sharpe Ratios of candidate portfolios for Pool B.

$\sigma = 0.04$	Without transaction cost				With transaction cost			
	$n = 120$		$n = 72$		$n = 120$		$n = 72$	
Portfolio	Risk	S-Ratio	Risk	S-Ratio	Risk	S-Ratio	Risk	S-Ratio
SRM	0.044	**0.249**	0.047	**0.245**	0.044	**0.248**	0.047	**0.244**
SRM^{-LOO}	0.045	0.247	0.048	0.244	0.045	0.246	0.048	0.242
MAXSER	0.045	0.191	0.052	0.158	0.045	0.190	0.052	0.157
MAXSER^{+LOO}	0.044	0.206	0.051	0.165	0.044	0.205	0.051	0.164
MV-NLS	0.056	0.167	0.065	0.139	0.056	0.167	0.065	0.139
MV-NLS-SSCV	0.037	0.149	0.034	0.125	0.037	0.148	0.034	0.124
MV-NLS-L1CV	0.029	0.172	0.022	0.143	0.029	0.171	0.022	0.142

Table 8. Risks and Sharpe Ratios of candidate portfolios for Pool C.

$\sigma = 0.04$	Without transaction cost				With transaction cost			
	$n = 120$		$n = 72$		$n = 120$		$n = 72$	
Portfolio	Risk	S-Ratio	Risk	S-Ratio	Risk	S-Ratio	Risk	S-Ratio
SRM	0.040	**0.190**	0.051	**0.127**	0.040	**0.189**	0.051	**0.127**
SRM^{-LOO}	0.041	0.187	0.052	0.126	0.041	0.185	0.052	0.125
MAXSER	0.044	0.113	0.056	0.099	0.044	0.112	0.056	0.099
MAXSER^{+LOO}	0.043	0.121	0.055	0.103	0.043	0.120	0.055	0.103
MV-NLS	0.055	0.099	0.070	0.087	0.055	0.098	0.070	0.087
MV-NLS-SSCV	0.036	0.088	0.033	0.078	0.036	0.087	0.033	0.077
MV-NLS-L1CV	0.028	0.102	0.023	0.090	0.028	0.102	0.023	0.090

Table 9. Risks and Sharpe Ratios of candidate portfolios for Pool D.

$\sigma = 0.04$	Without transaction cost				With transaction cost			
	$n = 120$		$n = 72$		$n = 120$		$n = 72$	
Portfolio	Risk	S-Ratio	Risk	S-Ratio	Risk	S-Ratio	Risk	S-Ratio
SRM	0.040	0.085	0.047	**0.142**	0.040	0.084	0.047	**0.140**
SRM^{-LOO}	0.041	0.084	0.048	0.141	0.041	0.083	0.048	0.140
MAXSER	0.040	0.087	0.048	0.126	0.040	0.086	0.048	0.124
MAXSER^{+LOO}	0.040	**0.094**	0.047	0.132	0.040	**0.094**	0.047	0.130
MV-NLS	0.050	0.077	0.060	0.110	0.050	0.077	0.060	0.109
MV-NLS-SSCV	0.033	0.069	0.031	0.099	0.033	0.069	0.031	0.098
MV-NLS-L1CV	0.025	0.079	0.020	0.113	0.025	0.078	0.020	0.112

Table 10. Risks and Sharpe Ratios of candidate portfolios for Pool E.

$\sigma = 0.04$	Without transaction cost				With transaction cost			
	$n = 120$		$n = 72$		$n = 120$		$n = 72$	
Portfolio	Risk	S-Ratio	Risk	S-Ratio	Risk	S-Ratio	Risk	S-Ratio
SRM	0.041	0.066	0.054	**0.071**	0.041	0.065	0.054	**0.071**
SRM^{-LOO}	0.042	0.065	0.055	0.070	0.042	0.064	0.055	0.070
MAXSER	0.040	0.068	0.053	0.063	0.040	0.067	0.053	0.063
MAXSER^{+LOO}	0.040	**0.070**	0.052	0.066	0.040	**0.069**	0.052	0.065
MV-NLS	0.050	0.060	0.066	0.055	0.050	0.059	0.066	0.054
MV-NLS-SSCV	0.032	0.054	0.030	0.049	0.032	0.054	0.030	0.048
MV-NLS-L1CV	0.026	0.062	0.022	0.057	0.026	0.061	0.022	0.056

better than MAXSER and other strategies. Second, the leave-one-out method embedded in SRM is useful, it can also improve MAXSER to some extent. Third, whether $n > p_n$ or $n < p_n$, SRM still outperforms MAXSER and other strategies.

From Tables 9 and 10, one can see that the leave-one-out method is quite useful. In addition to that, although SRM is not always better than MAXSER, when considering $n = 72$, SRM has ensured its competitiveness. It is well known that the relative performances of portfolio allocation strategies depend on underlying datasets (we have shown only five datasets here), rolling windows, performance measures and estimation methods, therefore, we are not intended to claim that our SRM is overwhelmingly superior to its alternatives. However, the empirical findings above do show the powerfulness and competitiveness of the proposed SRM in constraining the risk and maximizing the Sharpe ratios, especially for high-dimensional cases. We would also like to point out that SRM method only uses factors to achieve the covariance decomposition, and factor investing is not considered here. Since Ao, Li, and Zheng (2019) suggests that MAXSER with factor investing is more preferable to MAXSER without factor investing, we only claim that SRM performs better than MAXSER when factor investing is not allowed.

6. Conclusion

In this article, we propose a synthetic regression model for large portfolio allocation. Appealing the leave-one-out idea, we have successfully reduced the within sample correlation, which makes the estimated optimal portfolio allocation much more close to the theoretical optimal portfolio allocation. Due to the use of the structure of the factor model, an estimation method of high dimensional covariance matrices, and the penalized least-square estimation, the proposed method applies to the real large portfolio allocation where the number of assets under concern is much larger than the sample size. We have conducted intensive simulation studies and shown the proposed method outperforms its alternatives under some circumstances. We have also applied the proposed method to some publicly available real datasets and demonstrated the portfolio formed by the proposed method yields much higher return than its alternatives in most scenarios. In addition to the numerical demonstration of the superiority of the proposed method over its alternatives, in this article, we have also established the asymptotic theory of the proposed method, which has theoretically justified the proposed method.

The appendix contains the proofs of Theorems 1 and 2, and Lemmas 1–4 and their additional technical details.

Acknowledgments

The authors are grateful for the Editor, Associate Editor and two referees for their helpful comments that substantially improve this work.

Funding

This research is supported by National Natural Science Foundation of China (Grant Numbers 11931014, 11871001, 11901315, 72033002), the Beijing Natural Science Foundation (Grant Number 1182003) and the Fundamental Research Funds for the Central Universities (Grant Numbers 2019NTSS18, 2682020ZT113).

Appendix A: Proofs of the Theorems

For simplicity, we first introduce some notations. Let $\|\mathbf{X}\|_\Sigma^2 = \mathbf{X}^\mathrm{T}\Sigma\mathbf{X}$ denote the norm induced by matrix Σ for any vector $\mathbf{X} \in \mathbb{R}^{p_n}$.

Proof of Theorem 1. Let $\theta = \boldsymbol{\mu}^\mathrm{T}\Sigma^{-1}\boldsymbol{\mu}$ denotes the square of the maximum Sharpe ratio of the optimal portfolio, then it is easy to show that $\sigma\theta^{1/2} = \frac{\sigma}{\sqrt{\theta}}\boldsymbol{\mu}^\mathrm{T}\Sigma^{-1}\boldsymbol{\mu}$. As shown in Ao, Li, and Zheng (2019), the optimal portfolio $\mathbf{w}^*$ has the explicit expression: $\mathbf{w}^* = \frac{\sigma}{\sqrt{\theta}}\Sigma^{-1}\boldsymbol{\mu}$. Using Cauchy–Schwarz inequality, we have

$$\begin{aligned}\left|\hat{\mathbf{w}}^\mathrm{T}\boldsymbol{\mu} - \sigma\theta^{1/2}\right| &= \left|\hat{\mathbf{w}}^\mathrm{T}\boldsymbol{\mu} - \mathbf{w}^{*\mathrm{T}}\boldsymbol{\mu}\right| = \left|(\hat{\mathbf{w}} - \mathbf{w}^*)^\mathrm{T}\Sigma^{1/2}\Sigma^{-1/2}\boldsymbol{\mu}\right| \\ &\le \sqrt{(\hat{\mathbf{w}} - \mathbf{w}^*)^\mathrm{T}\Sigma(\hat{\mathbf{w}} - \mathbf{w}^*) \times \boldsymbol{\mu}^\mathrm{T}\Sigma^{-1}\boldsymbol{\mu}} \\ &= \sqrt{\theta\|\hat{\mathbf{w}} - \mathbf{w}^*\|_\Sigma^2},\end{aligned} \tag{A1}$$

where $\hat{\mathbf{w}}$ is the estimated optimal large portfolio allocation, which is the minimizer of Equation (7).

We first consider the convergence rate of $\|\hat{\mathbf{w}} - \mathbf{w}^*\|_\Sigma$. By the definition of $\hat{\mathbf{w}}$ in Equation (7) and the minimization property, we have

$$\frac{1}{2n}\sum_{i=1}^{n}\left(z_i - \mathbf{Y}_i^\mathrm{T}\hat{\mathbf{w}}\right)^2 + \lambda\|\hat{\mathbf{w}}\|_1 \le \frac{1}{2n}\sum_{i=1}^{n}\left(z_i - \mathbf{Y}_i^\mathrm{T}\mathbf{w}^*\right)^2 + \lambda\|\mathbf{w}^*\|_1. \tag{A2}$$

By Equation (6), (A2) and some simple calculations, we have

$$\begin{aligned}
&\frac{1}{2n}\sum_{i=1}^{n}\left(z_i-\mathbf{Y}_i^{\mathrm{T}}\hat{\mathbf{w}}\right)^2+\lambda\|\hat{\mathbf{w}}\|_1\\
&=\frac{1}{2n}\sum_{i=1}^{n}\left(z_i-\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^*+\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^*-\mathbf{Y}_i^{\mathrm{T}}\hat{\mathbf{w}}\right)^2+\lambda\|\hat{\mathbf{w}}\|_1\\
&=\frac{1}{2n}\sum_{i=1}^{n}\left(\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^*-\mathbf{Y}_i^{\mathrm{T}}\hat{\mathbf{w}}\right)^2+\frac{1}{2n}\sum_{i=1}^{n}(z_i-\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^*)^2\\
&\quad-\frac{1}{n}\sum_{i=1}^{n}(z_i-\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^*)\mathbf{Y}_i^{\mathrm{T}}\left(\hat{\mathbf{w}}-\mathbf{w}^*\right)+\lambda\|\hat{\mathbf{w}}\|_1\\
&\le\frac{1}{2n}\sum_{i=1}^{n}\left(z_i-\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^*\right)^2+\lambda\|\mathbf{w}^*\|_1.
\end{aligned}$$

Thus, we have the following inequality:

$$\begin{aligned}
&\frac{1}{2n}\sum_{i=1}^{n}\left(\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^*-\mathbf{Y}_i^{\mathrm{T}}\hat{\mathbf{w}}\right)^2+\lambda\|\hat{\mathbf{w}}\|_1\\
&\le\frac{1}{n}\sum_{i=1}^{n}\left(z_i-\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^*\right)\mathbf{Y}_i^{\mathrm{T}}\left(\hat{\mathbf{w}}-\mathbf{w}^*\right)+\lambda\|\mathbf{w}^*\|_1.
\end{aligned}\tag{A3}$$

For simplicity, let $f(\theta)=\sigma(1+\theta)\theta^{-1/2}$ and $f(\theta_i)=z_i=\sigma(1+\theta_i)\theta_i^{-1/2}$. By $\mathbf{w}^*=\frac{\sigma}{\sqrt{\theta}}\Sigma^{-1}\boldsymbol{\mu}$, it is easy to show that $f(\theta)=\frac{1+\theta}{\theta}\boldsymbol{\mu}^{\mathrm{T}}\mathbf{w}^*$. Thus, we have the following decomposition for the first term in Equation (A3), that is,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(z_i-\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^*\right)\mathbf{Y}_i^{\mathrm{T}}\left(\hat{\mathbf{w}}-\mathbf{w}^*\right)\\
&=\frac{1}{n}\sum_{i=1}^{n}\left[f(\theta_i)-f(\theta)+f(\theta)-\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^*\right]\mathbf{Y}_i^{\mathrm{T}}\left(\hat{\mathbf{w}}-\mathbf{w}^*\right)\\
&=: I_1+I_2,
\end{aligned}$$

where

$$I_1=\frac{1}{n}\sum_{i=1}^{n}\left[f(\theta_i)-f(\theta)\right]\mathbf{Y}_i^{\mathrm{T}}\left(\hat{\mathbf{w}}-\mathbf{w}^*\right)$$

and

$$I_2=\frac{1}{n}\sum_{i=1}^{n}\left[f(\theta)-\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^*\right]\mathbf{Y}_i^{\mathrm{T}}\left(\hat{\mathbf{w}}-\mathbf{w}^*\right).$$

We first consider I_1, and we can show that

$$\begin{aligned}
|I_1|&=\left|\frac{1}{n}\sum_{i=1}^{n}[f(\theta_i)-f(\theta)]\sum_{i=1}^{p_n}Y_{ij}(\hat{w}_j-w_j^*)\right|\\
&\le\|\hat{\mathbf{w}}-\mathbf{w}^*\|_1\cdot\max_{1\le i\le n}|f(\theta_i)-f(\theta)|\cdot\max_{1\le j\le p_n}\left|\frac{1}{n}\sum_{i=1}^{n}Y_{ij}\right|.
\end{aligned}\tag{A4}$$

Let $\bar{Y}_j=\frac{1}{n}\sum_{i=1}^{n}Y_{ij}=\mu_j+\frac{\sigma_j}{\sqrt{n}}\xi_j$, where $\xi_j, j=1,\dots,p_n$, are correlated standard normal random variables. By Lemma 1, we have $E\Big(\max_{1\le j\le p_n}|\xi_j|\Big)\le\sqrt{2\log(2p_n)}$. By Assumption 1 and $\log(p_n)/n\to 0$ as $n\to\infty$, then we have

$$\begin{aligned}
\max_{1\le j\le p_n}|\bar{Y}_j|&=\max_{1\le j\le p_n}\left|\mu_j+\frac{\sigma_j}{\sqrt{n}}\xi_j\right|\\
&\le\max_{1\le j\le p_n}|\mu_j|+\max_{1\le j\le p_n}|\sigma_j|\frac{\max_{1\le j\le p_n}|\xi_j|}{\sqrt{n}}\\
&\le L+O_p\left(\sqrt{\frac{2M\log(2p_n)}{n}}\right)=L+o_p(1).
\end{aligned}\tag{A5}$$

Obviously, $f(\theta)$ is a continuous function of θ, and its derivative is $f'(\theta)=\frac{1}{2}\sigma\theta^{-1/2}(1-\theta^{-1})$. For a small constant $0<l<L$ and the closed interval $[l,L]$, there exists a sufficiently large constant $C>0$ such that $|f(\theta_i)-f(\theta)|\le\sup_{\varsigma\in[l,L]}|f'(\varsigma)|\cdot|\theta_i-\theta|\le C|\theta_i-\theta|$ for each $i=1,\dots,p_n$. In order to obtain the convergence rate of $|f(\theta_i)-f(\theta)|$ for each $i=1,\dots,p_n$, we only need to bound $|\theta_i-\theta|$ for each $i=1,\dots,p_n$. Combining the results in Fan, Weng, and Zhou (2021) and Fan, Liao, and Mincheva (2011), and invoking Assumptions 1 and 3, we can obtain that $|\theta_i-\theta|=O_p\left(\frac{s_n\log p_n}{n}\vee\frac{1}{\sqrt{n}}\right)$ holds uniformly for each $i=1,\dots,n$ as $n\to\infty$. Thus, we have $|f(\theta_i)-f(\theta)|\le C|\theta_i-\theta|=O_p\left(\frac{s_n\log p_n}{n}\vee\frac{1}{\sqrt{n}}\right)$ for each $i=1,\dots,p_n$. Combining this result with (A4) and (A5), we have

$$|I_1|\le\|\hat{\mathbf{w}}-\mathbf{w}^*\|_1\cdot O_p\left(\frac{s_n\log p_n}{n}\vee\frac{1}{\sqrt{n}}\right).\tag{A6}$$

Now we consider I_2. By some simple calculations, we have

$$\begin{aligned}
I_2&=\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{w}^*-\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^*\right)\mathbf{Y}_i^{\mathrm{T}}\left(\hat{\mathbf{w}}-\mathbf{w}^*\right)\\
&\quad+\frac{\boldsymbol{\mu}^{\mathrm{T}}\mathbf{w}^*}{\theta}\frac{1}{n}\sum_{i=1}^{n}\mathbf{Y}_i^{\mathrm{T}}\left(\hat{\mathbf{w}}-\mathbf{w}^*\right)\\
&=\sum_{j=1}^{p_n}(w_j^*-\hat{w}_j)\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{p_n}w_k^*[(Y_{ij}-\mu_j)(Y_{ik}-\mu_k)-\sigma_{jk}]\\
&\quad+\sum_{j=1}^{p_n}(w_j^*-\hat{w}_j)\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{p_n}w_k^*(Y_{ik}-\mu_k)\mu_j\\
&\quad+\sum_{j=1}^{p_n}(w_j^*-\hat{w}_j)\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{p_n}w_k^*\sigma_{jk}\\
&\quad-\frac{\boldsymbol{\mu}^{\mathrm{T}}\mathbf{w}^*}{\theta}\sum_{j=1}^{p_n}(w_j^*-\hat{w}_j)\frac{1}{n}\sum_{i=1}^{n}Y_{ij}\\
&=:\frac{1}{n}\left[\sum_{j=1}^{p_n}(w_j^*-\hat{w}_j)I_{21,j}+\sum_{j=1}^{p_n}(w_j^*-\hat{w}_j)I_{22,j}\right.\\
&\quad\left.-\sum_{j=1}^{p_n}(w_j^*-\hat{w}_j)I_{23,j}\right],
\end{aligned}\tag{A7}$$

where

$$I_{21,j} = \sum_{i=1}^{n}\sum_{k=1}^{p_n} w_k^*[(Y_{ij}-\mu_j)(Y_{ik}-\mu_k)-\sigma_{jk}],$$
$$I_{22,j} = \sum_{i=1}^{n}\sum_{k=1}^{p_n} w_k^*(Y_{ik}-\mu_k)\mu_j,$$
$$I_{23,j} = \frac{\boldsymbol{\mu}^{\mathrm{T}}\mathbf{w}^*}{\theta}\sum_{i=1}^{n} Y_{ij} - \sum_{i=1}^{n}\sum_{k=1}^{p_n} w_k^*\sigma_{jk}.$$

For $I_{21,j}$, we first denote

$$\rho_j = \mathrm{corr}\left(Y_{ij}-\mu_j, \sum_{k=1}^{p_n} w_k^*(Y_{ik}-\mu_k)\right) = \frac{\sum_{k=1}^{p_n} w_k^*\sigma_{jk}}{\sigma_j\sigma},$$

where $\sigma_j = \mathrm{sd}(Y_{ij})$ for $j = 1, \dots, p_n$. Let $\{\xi_i, i = 1, \dots, n\}$ and $\{\eta_{ij}, i = 1, \dots, n\}$ be iid standard normal random variables, where $j = 1, \dots, p_n$. Thus, it is easy to show that

$$\begin{aligned} I_{21,j} &\overset{d}{=} \sum_{i=1}^{n}\sigma\sigma_j\left[\xi_i\left(\rho_j\xi_i + \sqrt{1-\rho_j^2}\eta_{ij}\right) - \rho_j\right] \\ &= \sigma\sigma_j\rho_j\sum_{i=1}^{n}(\xi_i^2-1) + \sigma\sigma_j\sqrt{1-\rho_j^2}\sum_{i=1}^{n}\xi_i\eta_{ij}, \end{aligned} \tag{A8}$$

where "$\overset{d}{=}$" denotes equal in distribution. By Assumption 1, we have

$$\begin{aligned} E\left(\max_{1\le j\le p_n}\left|\sigma\sigma_j\rho_j\sum_{i=1}^{n}(\xi_i^2-1)\right|\right) &\le \sigma\sqrt{M}\sqrt{E\left(\sum_{i=1}^{n}(\xi_i^2-1)\right)^2} \\ &= \sigma\sqrt{2nM}, \end{aligned}$$

By Lemma 3 and Assumption 1, we have

$$E\left(\max_{1\le j\le p_n}\left|\sigma\sigma_j\sqrt{1-\rho_j^2}\sum_{i=1}^{n}\xi_i\eta_{ij}\right|\right) \le 2\sigma\sqrt{nM\log(2p_n)}. \tag{A9}$$

For $I_{22,j}$, since w_k^* is the optimal portfolio, then we have $I_{22,j} \sim N(0, n\mu_j^2\sigma^2)$. Invoking Lemma 1 and Assumption 1, we have

$$E\left(\max_{1\le j\le p_n}|I_{22,j}|\right) \le \sigma L\sqrt{2n\log(2p_n)}. \tag{A10}$$

For $I_{23,j}$, by $\mathbf{w}^* = \frac{\sigma}{\sqrt{\theta}}\Sigma^{-1}\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\mathrm{T}}\mathbf{w}^* = \sigma\sqrt{\theta}$, we have

$$I_{23,j} = \frac{\sigma}{\sqrt{\theta}}\sum_{i=1}^{n}\left(Y_{ij} - \boldsymbol{\mu}^{\mathrm{T}}\Sigma^{-1}\Sigma(,j)\right) = \frac{\sigma}{\sqrt{\theta}}\sum_{i=1}^{n}(Y_{ij}-\mu_j),$$

where $\Sigma(,j)$ is the j-th column of Σ. Again using Lemma 1, we have

$$\begin{aligned} E\left(\max_{1\le j\le p_n}|I_{23,j}|\right) &\le \frac{\sigma}{\sqrt{\theta}}E\left(\max_{1\le j\le p_n}\left|\sum_{i=1}^{n}(Y_{ij}-\mu_j)\right|\right) \\ &\le \frac{\sigma\sqrt{M}}{\sqrt{\theta}}\sqrt{2n\log(2p_n)}. \end{aligned} \tag{A11}$$

Summarizing the above results from Equations (A7) to (A11), we have

$$|I_2| \le \|\hat{\mathbf{w}} - \mathbf{w}^*\|_1 \cdot O_p\left(\sqrt{\frac{\log p_n}{n}}\right). \tag{A12}$$

By Equations (A3), (A6), and (A12), we have the following inequality:

$$\begin{aligned} &\frac{1}{2n}\sum_{i=1}^{n}\left(\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^* - \mathbf{Y}_i^{\mathrm{T}}\hat{\mathbf{w}}\right)^2 + \lambda\|\hat{\mathbf{w}}\|_1 \\ &\le \frac{1}{n}\sum_{i=1}^{n}\left(z_i - \mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^*\right)\mathbf{Y}_i^{\mathrm{T}}\left(\hat{\mathbf{w}} - \mathbf{w}^*\right) + \lambda\|\mathbf{w}^*\|_1 \\ &\le \|\hat{\mathbf{w}} - \mathbf{w}^*\|_1 \cdot O_p\left(\frac{s_n\log p_n}{n} \vee \sqrt{\frac{\log p_n}{n}}\right) + \lambda\|\mathbf{w}^*\|_1. \end{aligned} \tag{A13}$$

By Assumption 4 and $\log(p_n)/n \to 0$ as $n \to \infty$, it is easy to show that $O_p\left(\sqrt{\log p_n/n}\right)\|\mathbf{w}^* - \hat{\mathbf{w}}\|_1$ has the faster convergence rate than $O_p\left(\sqrt{\log p_n/n}\right)$. Thus, we can show that

$$\begin{aligned} &O_p\left(\frac{s_n\log p_n}{n} \vee \sqrt{\frac{\log p_n}{n}} \vee \sqrt{\frac{\log p_n}{n}}\|\mathbf{w}^* - \hat{\mathbf{w}}\|_1\right) \\ &= O_p\left(\frac{s_n\log p_n}{n} \vee \sqrt{\frac{\log p_n}{n}}\right). \end{aligned}$$

Letting $\lambda_0 = C_0((s_n\log p_n/n) \vee \sqrt{\log p_n/n})$ with the large enough constant $C_0 > 0$, by Equation (A13) and Lemma 4, in probability, we have

$$\frac{1}{2}\|\hat{\mathbf{w}} - \mathbf{w}^*\|_\Sigma^2 + \lambda\|\hat{\mathbf{w}}\|_1 \le \lambda_0\|\hat{\mathbf{w}} - \mathbf{w}^*\|_1 + \lambda\|\mathbf{w}^*\|_1. \tag{A14}$$

Let $S = \{1 \le j \le p_n : w_j^* \ne 0\}$ denote the nonzero position set for optimal portfolio allocation $\mathbf{w}^*$, and S^c be complement of S. Note that $\|\hat{\mathbf{w}}\|_1 = \|\hat{\mathbf{w}}_S\|_1 + \|\hat{\mathbf{w}}_{S^c}\|_1$ and $\|\mathbf{w}^*\|_1 = \|\mathbf{w}_S^*\|_1 + \|\mathbf{w}_{S^c}^*\|_1 = \|\mathbf{w}_S^*\|_1$, where $\mathbf{w}_{S^c}^* = \mathbf{0}$. By (A14) and the inequality $\|\hat{\mathbf{w}}_S - \mathbf{w}_S^*\|_1 \ge \|\mathbf{w}_S^*\|_1 - \|\hat{\mathbf{w}}_S\|_1$, we have

$$\frac{1}{2}\|\hat{\mathbf{w}} - \mathbf{w}^*\|_\Sigma^2 + \lambda\|\hat{\mathbf{w}}_{S^c}\|_1 \le \lambda_0\|\hat{\mathbf{w}} - \mathbf{w}^*\|_1 + \lambda\|\hat{\mathbf{w}}_S - \mathbf{w}_S^*\|_1. \tag{A15}$$

Noting that $\|\hat{\mathbf{w}}_{S^c}\|_1 = \|\hat{\mathbf{w}}_{S^c} - \mathbf{w}_{S^c}^*\|_1$, by Equation (A15), we further have

$$\begin{aligned} &\frac{1}{2}\|\hat{\mathbf{w}} - \mathbf{w}^*\|_\Sigma^2 + \lambda\|\hat{\mathbf{w}}_{S^c} - \mathbf{w}_{S^c}^*\|_1 \le \lambda_0\|\hat{\mathbf{w}} - \mathbf{w}^*\|_1 + \lambda\|\hat{\mathbf{w}}_S - \mathbf{w}_S^*\|_1 \\ &= \lambda_0\|\hat{\mathbf{w}}_S - \mathbf{w}_S^*\|_1 + \lambda_0\|\hat{\mathbf{w}}_{S^c} - \mathbf{w}_{S^c}^*\|_1 + \lambda\|\hat{\mathbf{w}}_S - \mathbf{w}_S^*\|_1, \end{aligned}$$

that is

$$\frac{1}{2}\|\hat{\mathbf{w}} - \mathbf{w}^*\|_\Sigma^2 + (\lambda - \lambda_0)\|\hat{\mathbf{w}}_{S^c} - \mathbf{w}_{S^c}^*\|_1 \le (\lambda + \lambda_0)\|\hat{\mathbf{w}}_S - \mathbf{w}_S^*\|_1. \tag{A16}$$

If $\lambda \ge 2\lambda_0$, we have

$$\frac{1}{2}\|\hat{\mathbf{w}} - \mathbf{w}^*\|_\Sigma^2 + \frac{\lambda}{2}\|\hat{\mathbf{w}}_{S^c} - \mathbf{w}_{S^c}^*\|_1 \le \frac{3\lambda}{2}\|\hat{\mathbf{w}}_S - \mathbf{w}_S^*\|_1. \tag{A17}$$

As $\|\hat{\mathbf{w}} - \mathbf{w}^*\|_\Sigma^2 \geq 0$, we have the basic constraint $\|\hat{\mathbf{w}}_{S^c} - \mathbf{w}^*_{S^c}\|_1 \leq 3\|\hat{\mathbf{w}}_S - \mathbf{w}^*_S\|_1$ on the set $\mathcal{T}(S,3)$ defined in Assumption 2. By Equation (A17), we further have

$$\begin{aligned}
&\frac{1}{2}\|\hat{\mathbf{w}} - \mathbf{w}^*\|_\Sigma^2 + \frac{\lambda}{2}\|\hat{\mathbf{w}} - \mathbf{w}^*\|_1 \\
&= \frac{1}{2}\|\hat{\mathbf{w}} - \mathbf{w}^*\|_\Sigma^2 + \frac{\lambda}{2}\|\hat{\mathbf{w}}_{S^c} - \mathbf{w}^*_{S^c}\|_1 + \frac{\lambda}{2}\|\hat{\mathbf{w}}_S - \mathbf{w}^*_S\|_1 \\
&\leq 2\lambda\|\hat{\mathbf{w}}_S - \mathbf{w}^*_S\|_1.
\end{aligned} \tag{A18}$$

By Assumption 2, and invoking the Cauchy–Schwarz inequality and $2ab \leq a^2/4 + 4b^2$, for $\hat{\mathbf{w}}_S - \mathbf{w}^*_S \in \mathcal{T}(S,3)$, in probability, we have

$$\begin{aligned}
2\lambda\|\hat{\mathbf{w}}_S - \mathbf{w}^*_S\|_1 &\leq 2\lambda\sqrt{s_n}\|\hat{\mathbf{w}}_S - \mathbf{w}^*_S\|_2 \leq 2\lambda\sqrt{s_n}\|\hat{\mathbf{w}} - \mathbf{w}^*\|_2 \\
&\leq 2\lambda\sqrt{s_n}\|\hat{\mathbf{w}} - \mathbf{w}^*\|_\Sigma \Big/ \phi_0 \\
&\leq \frac{\|\hat{\mathbf{w}} - \mathbf{w}^*\|_\Sigma^2}{4} + \frac{4\lambda^2 s_n}{\phi_0^2}.
\end{aligned} \tag{A19}$$

By Equations (A18) and (A19), we can show that

$$\frac{1}{2}\|\hat{\mathbf{w}} - \mathbf{w}^*\|_\Sigma^2 + \lambda\|\hat{\mathbf{w}} - \mathbf{w}^*\|_1 \leq \frac{8\lambda^2 s_n}{\phi_0^2} \tag{A20}$$

holds in probability. From the above inequality, we can obtain that

$$\|\hat{\mathbf{w}} - \mathbf{w}^*\|_1 \leq \frac{8\lambda s_n}{\phi_0^2} \quad \text{and} \quad \|\hat{\mathbf{w}} - \mathbf{w}^*\|_\Sigma \leq \frac{4\lambda s_n^{1/2}}{\phi_0} \tag{A21}$$

holds in probability.

Note that $\lambda \asymp (s_n \log p_n/n) \vee \sqrt{\log p_n/n}$, and $s_n^{3/2}\log p_n/n \to 0$ as $n \to \infty$ in Assumption 4, it is easy to show that $\|\hat{\mathbf{w}} - \mathbf{w}^*\|_\Sigma = O_p(\lambda s_n^{1/2}) = o_p(1)$. Thus, by this result, Equation (A1) and Assumption 1, we finish the proof of Theorem 1. □

Proof of Theorem 2. Noting that $\mathbf{w}^* = \dfrac{\sigma}{\sqrt{\theta}}\Sigma^{-1}\boldsymbol{\mu}$, and using the triangular inequality for the norm $\|\cdot\|_\Sigma$, we have

$$\begin{aligned}
\left|\hat{\mathbf{w}}^{\mathrm{T}}\Sigma\hat{\mathbf{w}} - \sigma^2\right| &= \left|\hat{\mathbf{w}}^{\mathrm{T}}\Sigma\hat{\mathbf{w}} - \mathbf{w}^{*\mathrm{T}}\Sigma\mathbf{w}^*\right| \\
&= \Big|\|\hat{\mathbf{w}}\|_\Sigma - \|\mathbf{w}^*\|_\Sigma\Big| \leq \|\hat{\mathbf{w}} - \mathbf{w}^*\|_\Sigma.
\end{aligned}$$

By Theorem 1, it is easy to show that $\left|\hat{\mathbf{w}}^{\mathrm{T}}\Sigma\hat{\mathbf{w}} - \sigma^2\right| = O_p(\lambda s_n^{1/2}) = o_p(1)$ under Assumption 4. Thus, we finish the proof of Theorem 2. □

Appendix B: Some Lemmas and Proofs

Lemma 1. Suppose that $\xi_i \sim N(0, \sigma_i^2)$ for $i = 1, \ldots, m$, which need not be independent, then

$$E\left(\max_{1\leq i\leq m}|\xi_i|\right) \leq \max_{1\leq i\leq m}\sigma_i\sqrt{2\log(2m)}.$$

The proof of Lemma 1 can be found in Chatterjee (2013), hence we omit the details here.

Lemma 2. Suppose that $\zeta_i \sim \chi^2(n)$ for $i = 1, \ldots, m$, which need not be independent. If $\sqrt{\log(2m)/2n} \leq 1/4$, then

$$E\left(\max_{1\leq i\leq m}|\zeta_i - n|\right) \leq 2\sqrt{2n\log(2m)}.$$

Lemma 3. Suppose that $\xi_j \sim N(0,1)$ for $j = 1, \ldots, p_n$, and $\eta_k \sim N(0,1)$ for $k = 1, \ldots, q_n$. The two sequences $\{\xi_j, j = 1, \ldots, p_n\}$ and $\{\eta_k, j = 1, \ldots, q_n\}$ are independent, but ξ_j's do not need to be independent, neither do η_k's. Let ξ_{ij} and η_{ik} be iid copies of $\{\xi_j, j = 1, \ldots, p_n\}$ and $\{\eta_k, j = 1, \ldots, q_n\}$ respectively, where $i = 1, \ldots, n$. If $\log(2p_nq_n)/n \leq 1/2$, then

$$E\left(\max_{j,k}\left|\sum_{i=1}^n \xi_{ij}\eta_{ik}\right|\right) \leq 2\sqrt{n\log(2p_nq_n)}.$$

The proofs of Lemmas 2 and 3 can be found in Ao, Li, and Zheng (2019), hence we omit the details here.

Lemma 4. Suppose that $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip_n})^{\mathrm{T}}, i = 1, \ldots, n$, are iid random vectors from $N(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{p_n})^{\mathrm{T}}$ and $\Sigma = (\sigma_{jk})_{1\leq j,k\leq p_n}$. For $j, k = 1, \ldots, p_n$, let $\xi_{jk} = E(Y_jY_k) - \dfrac{1}{n}\sum_{i=1}^n Y_{ij}Y_{ik}$. If $\max_{1\leq j\leq p_n}|\mu_j| \leq L$ and $\max_{1\leq j\leq p_n}|\sigma_{jj}| \leq M$, then

$$\|\mathbf{w}^* - \hat{\mathbf{w}}\|_\Sigma^2 \leq \frac{1}{n}\sum_{i=1}^n(\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^* - \mathbf{Y}_i^{\mathrm{T}}\hat{\mathbf{w}})^2 + O_p\left(\sqrt{\frac{\log p_n}{n}}\right)\|\mathbf{w}^* - \hat{\mathbf{w}}\|_1^2.$$

Proof. Let $\mathcal{F}$ be the σ-algebra generated by $\{Y_{ij}, i = 1, \ldots, n; j = 1, \ldots, p_n\}$, and $\mathbf{Y} = (Y_1, \ldots, Y_{p_n})^{\mathrm{T}}$ be a future return. Note that $\hat{\mathbf{w}}$ is estimated by the observed data, then $\hat{\mathbf{w}}$ is independent of $\mathbf{Y} = (Y_1, \ldots, Y_{p_n})^{\mathrm{T}}$. By some simple calculations, we have

$$\begin{aligned}
&E\left[\left(\sum_{j=1}^{p_n} w_j^* Y_j - \sum_{j=1}^{p_n}\hat{w}_j Y_j\right)^2 \Bigg| \mathcal{F}\right] \\
&= \sum_{j,k=1}^{p_n}(w_j^* - \hat{w}_j)(w_k^* - \hat{w}_k)E(Y_jY_k) \\
&= \sum_{j,k=1}^{p_n}(w_j^* - \hat{w}_j)(w_k^* - \hat{w}_k)\Big[E(Y_j - \mu_j)(Y_k - \mu_k) + \mu_j\mu_k\Big] \\
&= (\mathbf{w}^* - \hat{\mathbf{w}})^{\mathrm{T}}\Sigma(\mathbf{w}^* - \hat{\mathbf{w}}) + \left(\sum_{j=1}^{p_n}(w_j^* - \hat{w}_j)\mu_j\right)^2 \\
&\geq (\mathbf{w}^* - \hat{\mathbf{w}})^{\mathrm{T}}\Sigma(\mathbf{w}^* - \hat{\mathbf{w}}) = \|\mathbf{w}^* - \hat{\mathbf{w}}\|_\Sigma^2
\end{aligned} \tag{B1}$$

and

$$\frac{1}{n}\sum_{i=1}^n(\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^* - \mathbf{Y}_i^{\mathrm{T}}\hat{\mathbf{w}})^2 = \frac{1}{n}\sum_{i=1}^n\sum_{j,k=1}^{p_n}(w_j^* - \hat{w}_j)(w_k^* - \hat{w}_k)Y_{ij}Y_{ik}.$$

By the definition of $\xi_{jk} = E(Y_jY_k) - \dfrac{1}{n}\sum_{i=1}^n Y_{ij}Y_{ik}$, we have

$$\begin{aligned}
&E\left[\left(\sum_{i=1}^{p_n} w_j^* Y_j - \sum_{i=1}^{p_n}\hat{w}_j Y_j\right)^2 \Bigg| \mathcal{F}\right] - \frac{1}{n}\sum_{i=1}^n(\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^* - \mathbf{Y}_i^{\mathrm{T}}\hat{\mathbf{w}})^2 \\
&= \sum_{j,k=1}^{p_n}(w_j^* - \hat{w}_j)(w_k^* - \hat{w}_k)E(Y_jY_k) \\
&\quad - \frac{1}{n}\sum_{i=1}^n\sum_{j,k=1}^{p_n}(w_j^* - \hat{w}_j)(w_k^* - \hat{w}_k)Y_{ij}Y_{ik} \\
&= \sum_{j,k=1}^{p_n}(w_j^* - \hat{w}_j)(w_k^* - \hat{w}_k)\xi_{jk} \leq \|\mathbf{w}^* - \hat{\mathbf{w}}\|_1^2\max_{1\leq j,k\leq p_n}|\xi_{jk}|.
\end{aligned} \tag{B2}$$

Now we will bound the term $\max_{1\le j,k\le p_n} |\xi_{jk}|$ above. Some simple calculations yield that

$$\sum_{i=1}^{n} Y_{ij}Y_{ik} = \sum_{i=1}^{n}(Y_{ij}-\mu_j)(Y_{ik}-\mu_k) + \sum_{i=1}^{n}(Y_{ij}-\mu_j)\mu_k + \sum_{i=1}^{n}(Y_{ik}-\mu_k)\mu_j + n\mu_j\mu_k$$

and

$$\xi_{jk} = E(Y_jY_k) - \frac{1}{n}\sum_{i=1}^{n} Y_{ij}Y_{ik} = \sigma_{jk} - \frac{1}{n}\sum_{i=1}^{n}(Y_{ij}-\mu_j)(Y_{ik}-\mu_k) - \frac{1}{n}\sum_{i=1}^{n}(Y_{ij}-\mu_j)\mu_k - \frac{1}{n}\sum_{i=1}^{n}(Y_{ik}-\mu_k)\mu_j. \tag{B3}$$

Let $A_{jk} = \sum_{i=1}^{n}(Y_{ij}-\mu_j)(Y_{ik}-\mu_k)$, $B_{jk} = \sum_{i=1}^{n}(Y_{ij}-\mu_j)\mu_k$, $C_{jk} = \sum_{i=1}^{n}(Y_{ik}-\mu_k)\mu_j$ and $\rho_{jk} = \mathrm{corr}(Y_j, Y_k)$. Further, let $\{\xi_{ij}, i = 1,\dots,n\}$ and $\{\eta_{ik}, i = 1,\dots,n\}$ are iid standard normal random variables. For A_{jk}, we have

$$A_{jk} \overset{d}{=} \sqrt{\sigma_{jj}\sigma_{kk}}\sum_{i=1}^{n}\xi_{ij}\left(\rho_{jk}\xi_{ij} + \sqrt{1-\rho_{jk}^2}\eta_{ik}\right) = \sigma_{jk}\sum_{i=1}^{n}(\xi_{ij}^2 - 1) + n\sigma_{jk} + \sqrt{\sigma_{jj}\sigma_{kk}(1-\rho_{jk}^2)}\sum_{i=1}^{n}\xi_{ij}\eta_{ik} \tag{B4}$$

and

$$\sigma_{jk} - \frac{1}{n}A_{jk} = -\left[\frac{\sigma_{jk}}{n}\sum_{i=1}^{n}(\xi_{ij}^2 - 1) + \frac{\sqrt{\sigma_{jj}\sigma_{kk}(1-\rho_{jk}^2)}}{n}\sum_{i=1}^{n}\xi_{ij}\eta_{ik}\right]. \tag{B5}$$

By Lemmas 2 and 3, we have

$$E\left(\max_{1\le j\le p_n}\left|\sum_{i=1}^{n}(\xi_{ij}^2 - 1)\right|\right) \le 2\sqrt{2n\log(2p_n)}, \tag{B6}$$

$$E\left(\max_{1\le j\le p_n}\left|\sum_{i=1}^{n}\xi_{ij}\eta_{ik}\right|\right) \le 2\sqrt{n\log(2p_n^2)}. \tag{B7}$$

It is easy to show that $B_{jk} \sim N(0, n\mu_k^2\sigma_{jj})$ and $C_{jk} \sim N(0, n\mu_j^2\sigma_{kk})$. Thus, we can show that by Lemma 1

$$\max\left(E\left(\max_{1\le j,k\le p_n}|B_{jk}|\right), E\left(\max_{1\le j,k\le p_n}|C_{jk}|\right)\right) \le L\sqrt{2nM\log(2p_n^2)}. \tag{B8}$$

Summarizing the above results from (B3)–(B8), we have

$$E\left(\max_{1\le j,k\le p_n}|\xi_{jk}|\right) \le 2M\sqrt{\frac{2\log(2p_n)}{n}} + (2M + 2L\sqrt{2M})\sqrt{\frac{\log(2p_n^2)}{n}}, \tag{B9}$$

which implies that $\max_{1\le j,k\le p_n}|\xi_{jk}| = O_p(\sqrt{p_n/n})$. Combining this result with Equations (B1) and (B2), we have

$$\|\mathbf{w}^* - \hat{\mathbf{w}}\|_{\Sigma}^2 \le \frac{1}{n}\sum_{i=1}^{n}(\mathbf{Y}_i^{\mathrm{T}}\mathbf{w}^* - \mathbf{Y}_i^{\mathrm{T}}\hat{\mathbf{w}})^2 + O_p\left(\sqrt{\frac{\log p_n}{n}}\right)\|\mathbf{w}^* - \hat{\mathbf{w}}\|_1^2.$$

Thus, we finish the proof of Lemma 4. □

References

Ao, M., Li, Y., and Zheng, X. (2019), "Approaching Mean-Variance Efficiency for Large Portfolios," *Review of Financial Studies*, 32, 2890–2919. [1,2,3,4,5,7,8,11]

Avella-Medina, M., Battey, H., Fan, J., and Li, Q. (2018), "Robust Estimation of High Dimensional Covariance and Precision Matrices," *Biometrika*, 105, 271–284. [1]

Basak, G. K., Jagannathan, R., and Ma, T. (2009), "Jackknife Estimator for Tracking Error Variance of Optimal Portfolios," *Management Science*, 55, 990–1002. [1]

Berthet, Q., and Rigollet, P. (2013), "Optimal Detection of Sparse Principal Components in High Dimension," *The Annals of Statistics*, 41, 1780–1815. [1]

Bickel, P., and Levina, E. (2008a), "Covariance Regularization by Thresholding," *The Annals of Statistics*, 36, 2577–2604. [1]

——— (2008b), "Regularized Estimation of Large Covariance Matrices," *The Annals of Statistics*, 36, 199–227. [1]

Bickel, P., Ritov, Y., and Tsybakov, A. B. (2009), "Simultaneous Analysis of Lasso and Dantzig Selector," *The Annals of Statistics*, 37, 1705–1732. [4]

Birnbaum, A., Johnstone, I. M., Nadler, B., and Paul, D. (2013), "Minimax Bounds for Sparse PCA With Noisy High-Dimensional Data," *The Annals of Statistics*, 41, 1055–1084. [1]

Britten-Jones, M. (1999), "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights," *The Journal of Finance*, 54, 655–671. [1]

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., and Loris, I. (2009), "Sparse and Stable Markowitz Portfolios," *Proceedings of the National Academy of Sciences*, 106, 12267–12272. [1]

Bunea, F., Tsybakov, A., and Wegkamp, M. (2007), "Sparsity Oracle Inequalities for the Lasso," *Electronic Journal of Statistics*, 1, 169–194. [4]

Candès, E., and Tao, T. (2007), "The Dantzig Selector: Statistical Estimation When p is Much Larger Than n" (with discussion), *The Annals of Statistics*, 35, 2313–2351. [4]

Chatterjee, S. (2013), "Assumptionless Consistency of the Lasso," arXiv:1303.5817. [11]

DeMiguel, V., Garlappi, L., and Uppal, R. (2009), "Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?" *Review of Financial Studies*, 22, 1915–1953. [1]

El Karoui, N. (2008), "Operator Norm Consistent Estimation of a Large Dimensional Sparse Covariance Matrices," *The Annals of Statistics*, 36, 2717–2756. [1]

Engle, R., Ferstenberg, R., and Russell, J. (2012), "Measuring and Modeling Execution Cost and Risk," *Journal of Portfolio Management*, 38, 14–28. [7]

Fama, E., and French, K. (1993), "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56. [2]

Fan, J., Fan, Y., and Lv, J. (2008), "High Dimensional Covariance Matrix Estimation Using a Factor Model," *Journal of Econometrics*, 147, 186–197. [1,2,4]

Fan, J., Liao, Y., and Mincheva, M. (2011), "High Dimensional Covariance Matrix Estimation in Approximate Factor Models," *The Annals of Statistics*, 39, 3320–3356. [1,4,9]

——— (2013), "Large Covariance Estimation by Thresholding Principal Orthogonal Complements" (with discussion), *Journal of Royal Statistical Society*, Series B, 75, 603–680. [1]

Fan, J., Weng, H., and Zhou, Y. (2021), "Optimal Estimation of Functionals of High-Dimensional Mean and Covariance Matrix," arXiv:1908.07460v2. [4,9]

Guo, S., Box, J., and Zhang, W. (2017), "A Dynamic Structure for High Dimensional Covariance Matrices and Its Application in Portfolio Allocation," *Journal of the American Statistical Association*, 112, 235–253. [1]

Kan, R., and Zhou, G. (2007), "Optimal Portfolio Choice With Parameter Uncertainty," *Journal of Financial and Quantitative Analysis*, 42, 621–656. [2]

Lam, C. (2016), "Nonparametric Eigenvalue-Regularized Precision or Covariance Matrix Estimator," *The Annals of Statistics*, 44, 928–953. [1]

Ledoit, O., and Wolf, M. (2004), "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices," *Journal of Multivariate Analysis*, 88, 365–411. [4]

——— (2008), "Robust Performance Hypothesis Testing With the Sharpe Ratio," *Journal of Empirical Finance*, 15, 850–859. [6]

Ledoit, O., and Wolf, M. (2017), "Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks," *The Review of Financial Studies*, 30, 4349–4388. [1,4]

Markowitz, H. M. (1952), "Portfolio Selection," *The Journal of Finance*, 7, 77–91. [1]

Meinshausen, N., and Yu, B. (2009), "Lasso-Type Recovery of Sparse Representations for High-Dimensional Data," *The Annals of Statistics*, 37, 246–270. [4]

Raskutti, G., Wainwright, M. J., and Yu, B. (2010), "Restricted Eigenvalue Properties for Correlated Gaussian Designs," *Journal of Machine Learning Research*, 11, 2241–2259. [4]

Rothman, A. J., Levina, E., and Zhu, J. (2009), "Generalized Thresholding of Large Covariance Matrices," *Journal of the American Statistical Association*, 104, 177–186. [1]

Sun, Y., Zhang, W., and Tong, H. (2007), "Estimation of the Covariance Matrix of Random Effects in Longitudinal Studies," *The Annals of Statistics*, 35, 2795–2814. [1]

van de Geer, S. (2006), "High-Dimensional Generalized Linear Models and the Lasso," *The Annals of Statistics*, 36, 614–645. [4]

Yuan, M. (2010), "High Dimensional Inverse Covariance Matrix Estimation Via Linear Programming," *Journal of Machine Learning Research*, 11, 2261–2286. [1]

Zou, C., Ke, Y., and Zhang, W. (2020), "Estimation of Low Rank High-Dimensional Multivariate Linear Models for Multi-Response Data," *Journal of the American Statistical Association*, 1–11. [4]