Approaching Mean-Variance Efficiency for Large Portfolios *

Mengmeng Ao[†] Yingying Li[‡] Xinghua Zheng[§]

First Draft: October 6, 2014
This Draft: November 11, 2017

Abstract

This paper studies the large dimensional Markowitz optimization problem. Given any risk constraint level, we introduce a new approach for estimating the optimal portfolio, which is developed through a novel *unconstrained* regression representation of the mean-variance optimization problem, combined with high-dimensional sparse regression methods. Our estimated portfolio, under a mild sparsity assumption, asymptotically achieves mean-variance efficiency and meanwhile effectively controls the risk. To the best of our knowledge, this is the first time that these two goals can be simultaneously achieved for large portfolios. The superior properties of our approach are demonstrated via comprehensive simulation and empirical studies.

Keywords: Markowitz optimization; Large portfolio selection; Unconstrained regression, LASSO; Sharpe ratio

*Research partially supported by the RGC grants GRF16305315, GRF 16502014 and GRF 16518716 of the HKSAR, and The Fundamental Research Funds for the Central Universities 20720171073.

[†]Wang Yanan Institute for Studies in Economics & Department of Finance, School of Economics, Xiamen University, China. mengmengao@xmu.edu.cn

[‡]Hong Kong University of Science and Technology, HKSAR. yyli@ust.hk

[§]Hong Kong University of Science and Technology, HKSAR. xhzheng@ust.hk

1 INTRODUCTION

1.1 Markowitz Optimization Enigma

The groundbreaking mean-variance portfolio theory proposed by Markowitz (1952) continues to play significant roles in research and practice. The optimal mean-variance portfolio has a simple explicit expression[1] that only depends on two population characteristics, the mean and the covariance matrix of asset returns. Under the ideal situation when the underlying mean and covariance matrix are known, mean-variance investors can easily compute the optimal portfolio weights based on their preferred level of risk. In the real world, however, the true parameters are unknown. Sample mean and sample covariance matrix are used as proxies, and the resulting "plug-in" portfolio has been widely adopted. Such an approach is justified by the classical statistics theory because the plug-in portfolio is an MLE of the optimal portfolio. However, as is documented in Michaud (1989) and others, the out-of-sample performance of the plug-in portfolio is poor. Moreover, the situation worsens as the number of assets increases. (For additional details, see Best and Grauer (1991), Chopra and Ziemba (1993), Britten-Jones (1999), Kan and Zhou (2007), and Basak et al. (2009) among others.) Termed the "Markowitz Optimization Enigma" by Michaud (1989), the issues of constructing the mean-variance optimal portfolio based on sample estimates limit the use of Markowitz's mean-variance framework.

1.2 Challenges for Large Portfolios

Modern portfolios often include a large number of assets. This makes the optimization problem high-dimensional in nature, and induces numerous challenges. Take the plug-in portfolio for example, as we will see below, the risk of the plug-in portfolio can be substantially higher than the pre-specified risk level even when the portfolio weights are computed based on simulated i.i.d. returns. On the other hand, such high risk is not well compensated by high returns, resulting in significantly suboptimal Sharpe ratios. The key message is, **even in the ideal situation when all assumptions of Markowitz optimization are satisfied** (i.e., no time-varying parameters, regime switching, etc.), there are intrinsic challenges towards the estimation of the mean-variance efficient portfolio.

[1] See details in Section 2.1.

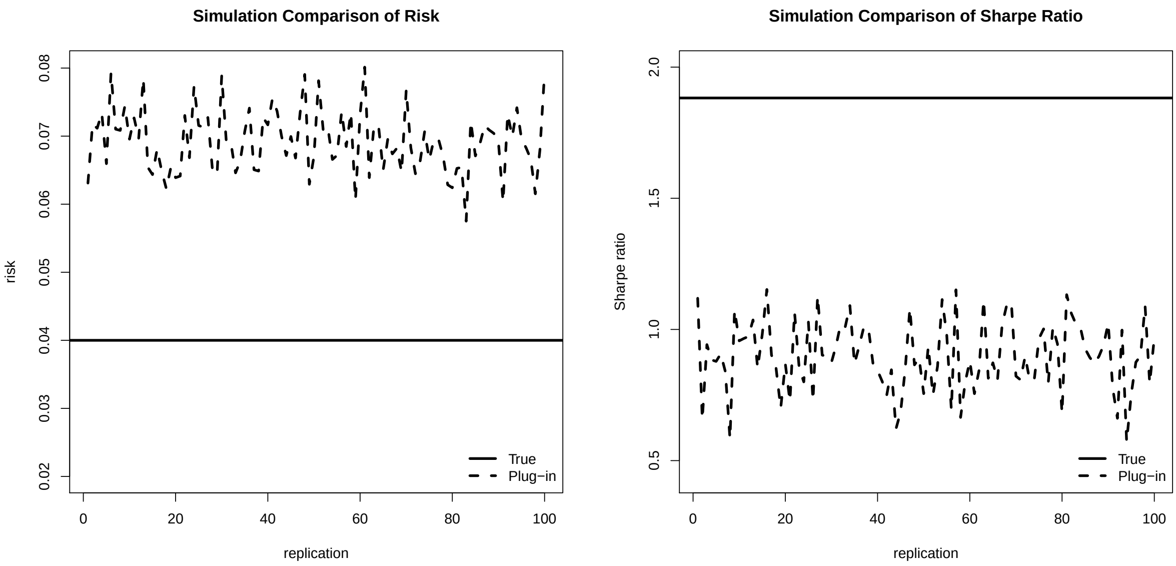

Figure 1. *Simulation comparisons of risks and Sharpe ratios of the plug-in portfolio versus the true optimal portfolio. The portfolios are constructed based on data generated from i.i.d. multivariate normal distribution with parameters specified in Section 3.2. The left panel plots the portfolio risks, and the right panel plots the Sharpe ratios. The pool of assets includes 100 stocks and 3 factors, and the number of observations is 240. The comparison is replicated 100 times.*

Figure 1 shows the comparisons between the plug-in portfolio (black dashed lines) and the theoretical optimal portfolio (black solid lines). The asset pool consists of 100 stocks and 3 (tradable) factors. The underlying mean and covariance matrix are calibrated from real data (see Section 3.2 for details). We generate 20 years of monthly returns from i.i.d. multivariate normal distribution, based on which we construct the plug-in portfolio and compare its risk and Sharpe ratio with the theoretical optimal values. Such simulated returns satisfy all the assumptions of Markowitz's mean-variance framework. However, as we observe from Figure 1, in all 100 replications, the plug-in portfolio carries a risk that is almost twice the specified level. Meanwhile, as shown in the right panel of Figure 1, the Sharpe ratio of the plug-in portfolio is only about 50% of the theoretical maximum Sharpe ratio.

The above phenomenon has been noted in Kan and Zhou (2007), and further investigated in Bai et al. (2009) and El Karoui (2010). These papers document that the deviation of the plug-in portfolio from the optimal portfolio is systematic, and the bias is due to the dimension (number of assets) being not negligible compared with the sample size. For global minimum variance portfolio (GMV), Basak et al. (2009) derive a result in a similar spirit, which says that the plug-in GMV carries, on average, a risk that is a bigger than 1 multiple of the true minimum risk, and the multiplier explicitly depends on the number of assets and sample size.

1.3 Existing Alternative Methods

The plug-in portfolio is obtained by replacing the population mean and covariance matrix in the formula for the optimal portfolio with their sample estimates. Alternatively, people seek to improve portfolio performance by plugging in better estimates of the underlying mean and covariance matrix. For estimation of covariance matrix, a widely used alternative estimator is the linear shrinkage estimator proposed in Ledoit and Wolf (2003, 2004), which estimates the covariance matrix by a suitable linear combination of the sample covariance matrix and a target matrix (e.g., identity or single-index matrix). More recently, Ledoit and Wolf (2017) propose a nonlinear shrinkage estimator of the covariance matrix and its factor-model-adjusted version that are suitable for portfolio optimization. For estimation of mean, among other works, Black and Litterman (1991) propose a quasi-Bayesian approach by combining investors' views with returns implied by CAPM. This quasi-Bayesian approach is extended to a fully Bayesian approach by Lai et al. (2011), who consider the mean-variance problem from a different angle and aim to maximize a certain utility function. Garlappi et al. (2007) propose to adjust estimates of expected returns by a multi-prior approach, also with an aim of maximizing a utility function. The aforementioned paper Bai et al. (2009), which analyze the systematic bias in the plug-in portfolio, propose a "bootstrap-corrected" method to estimate the optimal portfolio. However, as pointed out in a more recent working paper (Bai et al. 2013), the bootstrap-corrected method fails to satisfy the risk constraint.

Another direction to improve portfolio performance is to modify the original framework by imposing various constraints on portfolio weights. Most research in this direction focuses on the GMV portfolio. Imposing constraints on the weights has been empirically shown to be helpful; see, for example, Jagannathan and Ma (2003), DeMiguel et al. (2009), Brodie et al. (2009) and Fastrich et al. (2012). Fan et al. (2012b) give theoretical justifications to the empirical results in Jagannathan and Ma (2003), and also investigate the GMV portfolio with gross-exposure constraints where some short positions are allowed. Fan et al. (2012a) consider GMV portfolio with high-frequency data under gross-exposure constraints.

In addition to the approaches mentioned above, combinations of different portfolios have been studied; see, for example, Kan and Zhou (2007) and Tu and Zhou (2011).

The aforementioned methods lead to improved portfolio performance. However, they are still suboptimal. Take the latest development, the nonlinear shrinkage method in Ledoit and Wolf (2017) as an example, we see in Figure 2 that although its risk is substantially lower than that of the plug-in portfolio, it still violates the risk constraint, and is also significantly suboptimal in terms of Sharpe ratio.

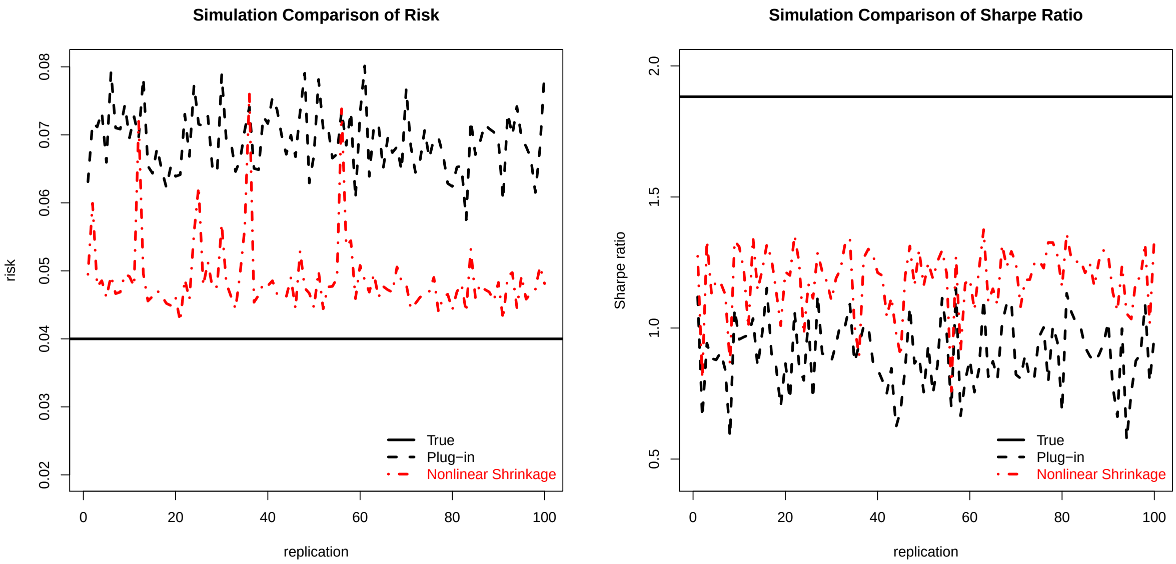

Figure 2. *Simulation comparisons of the plug-in and nonlinear shrinkage portfolios. The portfolios are constructed based on the same data used for Figure 1. The left and right panels plot the portfolio risks and Sharpe ratios, respectively. There are 100 stocks and 3 factors in the asset pool, and the sample size is 240. The comparison is replicated 100 times.*

More comprehensive comparisons including several other benchmark strategies are made in Sections 3 and 4, based on both simulated and empirical data. The comparisons reveal similar conclusions.

1.4 Our Contributions

In this paper, a new methodology for estimating the mean-variance efficient portfolio is proposed, which we call the MAXimum-Sharpe-ratio Estimated & sparse Regression (MAXSER) method. MAXSER is a general approach which can be applied to various situations when the number of assets in portfolios is not negligible compared with sample size. We show that, under mild assumptions, the MAXSER portfolio can asymptotically (1) achieve mean-variance efficiency and (2) satisfy the risk constraint. To the best of our knowledge, this is the first time that both objectives can be simultaneously achieved for large portfolio optimization.

Our first main contribution is establishing an equivalent unconstrained regression representation of the Markowitz optimization problem. This special regression representation is a novel finding made in this paper. There is existing literature on using regression to estimate the optimal portfolio. Two most closely related approaches are Britten-Jones (1999), who use 1 as the response; and Brodie et al. (2009), who use the maximum expected return as the response. The issues with the two approaches are that, the regression in Britten-Jones (1999) results in a biased solution; while Brodie et al. (2009) deal with a constrained regression, which is challenging and involves error and biases induced by sample estimates. Our regression representation, in contrast, is on the one hand, *equivalent* to the original optimization problem, and on the other hand, *unconstrained*, so that it can be conveniently combined with high-dimensional regression techniques. See more details in Section 2.1.

Our method further involves the following important aspects:

(1) Consistent estimation of the response in our regression representation, which depends on consistent estimation of the maximum Sharpe ratio achieved by the tangency portfolio.

(2) Proper sparse regression and rigorous analysis under the framework of our regression which possesses some unique features.

(3) The estimation in (1) and the sparse regression in (2) consists of the core of MAXSER. Depending on whether factor structure is present in the asset pool or not, MAXSER can be used directly on all assets or on idiosyncratic components.

- When there is no factor structure in returns, the combination of estimation of the response and the sparse regression is directly applied to the assets. See Section 2.2.
- When factor structure does present, we develop a framework that decomposes the portfolio estimation into the estimation of the optimal portfolio on factors and that on idiosyncratic components. See Section 2.3.

Under both settings, we theoretically prove the convergences of the expected return and risk of our estimated portfolio towards the theoretical maximum expected return and risk constraint, respectively. These properties guarantee that our estimator can asymptotically achieve mean-variance efficiency as both the number of assets and sample size get large.

The theoretical properties of the MAXSER portfolio are supported by simulation and empirical studies. We compare our method with a number of benchmark methods including the plug-in portfolio, the equally weighted portfolio, the linear/nonlinear shrinkage portfolio of Ledoit and Wolf (2004) and Ledoit and Wolf (2017), and several other variations of MV/GMV portfolios with constraints on portfolio weights. The complete simulation results are given in Section 3. Figure 3 below shows the comparison among the plug-in, nonlinear shrinkage and MAXSER portfolios. The added blue dashed lines plot the results of our portfolio. We see that our portfolio effectively controls the risk to **satisfy the risk constraint**. More importantly, the comparisons of Sharpe ratios show that our MAXSER portfolio ***nearly achieves the mean-variance efficiency*** and significantly outperforms others.

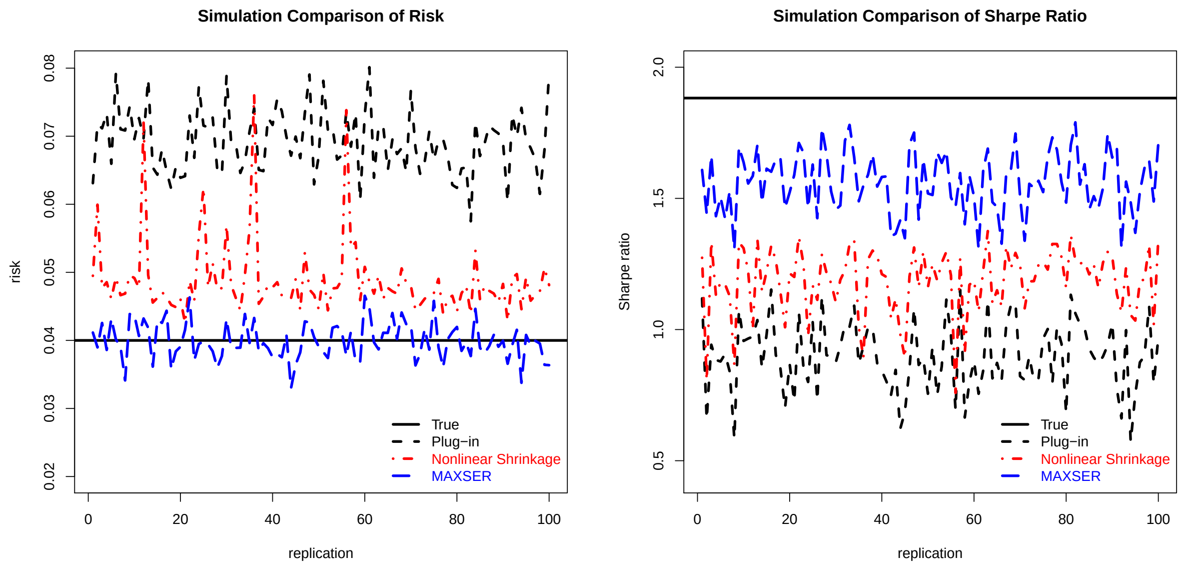

Figure 3. *Comparison of our MAXSER portfolio with the plug-in and nonlinear shrinkage portfolios, based on the same simulated data as for Figures 1 and 2. The added blue dashed lines represent the MAXSER portfolio. The comparison is replicated 100 times.*

Comprehensive empirical studies are presented in Section 4, in which we demonstrate the favorable performance of our proposed strategy.

2 The MAXSER Methodology

2.1 An Unconstrained Regression Representation

Suppose that we have a pool of N risky assets. Denote their (random excess) returns by $\boldsymbol{r} = (r_1, r_2, \cdots, r_N)'$, where for any vector $\boldsymbol{v}$, $\boldsymbol{v}'$ stands for the transpose of $\boldsymbol{v}$. Let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be the mean vector and covariance matrix of $\boldsymbol{r}$, respectively, and let $\boldsymbol{w}$ be a vector of portfolio weights on the risky assets. For a given level of risk constraint σ, the Markowitz optimization problem is

$$\arg\max_{\boldsymbol{w}} E(\boldsymbol{w}'\boldsymbol{r}) = \boldsymbol{w}'\boldsymbol{\mu} \quad \text{subject to} \quad \mathrm{Var}(\boldsymbol{w}'\boldsymbol{r}) = \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \leq \sigma^2. \tag{2.1}$$

If we denote by $\theta = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ the square of the maximum Sharpe ratio of the optimal portfolio, then the optimization problem (2.1) can be represented in its dual form with a return constraint $r^* = \sigma\sqrt{\theta}$:

$$\arg\min_{\boldsymbol{w}} \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{w}'\boldsymbol{\mu} = r^*. \tag{2.2}$$

The optimal portfolio, $\boldsymbol{w}^*$, admits the following explicit expression:

$$\boldsymbol{w}^* = \frac{\sigma}{\sqrt{\theta}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}. \tag{2.3}$$

The optimal portfolio $\boldsymbol{w}^*$ satisfies the following linear relationship:

$$r^* = \boldsymbol{w}'\boldsymbol{r} + \varepsilon, \quad \varepsilon \sim (0, \sigma^2), \tag{2.4}$$

where $\varepsilon = r^* - \boldsymbol{w}'\boldsymbol{r}$ is a random variable with mean 0 and variance σ^2. Based on such an equation, a natural way to solve for $\boldsymbol{w}^*$ is the ordinary least squares (OLS):

$$\boldsymbol{w}_0 := \arg\min_{\boldsymbol{w}} E(r^* - \boldsymbol{w}'\boldsymbol{r})^2. \tag{2.5}$$

However, $\boldsymbol{w}_0$ is not the optimal portfolio. In fact, we have

$$\boldsymbol{w}_0 = \frac{\theta}{1+\theta}\boldsymbol{w}^*. \tag{2.6}$$

In particular, we see that the OLS solution yields a smaller expected return and carries a lower risk.

The result above shows that the OLS (2.5) leads to biased solution. The bias is due to a unique feature of the OLS (2.5): Unlike in the conventional regression where the noise is independent of (or at least uncorrelated with) the predictor, in (2.5) the noise ε and the predictor are *correlated.* To correct for the bias, if still r^* is used as the response, then a constraint must be added to least squares:

$$\arg\min_{\boldsymbol{w}} E(r^* - \boldsymbol{w}'\boldsymbol{r})^2 \quad \text{subject to} \quad E(\boldsymbol{w}'\boldsymbol{r}) = r^* \text{ or } \text{Var}(\boldsymbol{w}'\boldsymbol{r}) = \sigma^2.$$

The difficulty with such an approach is that the constraint has to be replaced with a sample version as what is done in Brodie et al. (2009). In such a way, estimation error and even bias are introduced. Our proposal, in contrast, is the following novel unconstrained regression.

Proposition 1. *The unconstrained regression*

$$\arg\min_{\boldsymbol{w}} E(r_c - \boldsymbol{w}'\boldsymbol{r})^2, \quad \textit{where} \quad r_c := \frac{1+\theta}{\theta} r^* \equiv \frac{1+\theta}{\sqrt{\theta}}\sigma. \tag{2.7}$$

is equivalent to the Markowitz optimization (2.1) *or* (2.2).

The *corrected* response in our regression representation (2.7), r_c, adjusts the bias of the OLS solution $\boldsymbol{w}_0$ by rescaling the maximum expected return r^*. In such a way, the Markowtiz optimization problem is translated into an *equivalent unconstrained regression.*

Remark 1. *In Britten-Jones (1999), the author connects the estimation of the tangency portfolio with the regression coefficients in an OLS regression. There, the response is simply 1, which results in a multiple of the plug-in tangency portfolio, hence is biased. To obtain the tangency portfolio, one scales the weights so that the weights add up to one. After such an adjustment, one gets the* plug-in *tangency portfolio, which, unfortunately, will not be mean-variance efficient.*

We emphasize that our regression representation (2.7) is intrinsically different from existing regression representations in the literature for mean-variance portfolio estimation. Our representation is *unconstrained* and *equivalent* to the Markowitz problem. The elimination of constraint is particularly helpful for the estimation of the optimal portfolio of a large pool of assets, for which important techniques like sparse regression becomes directly applicable.

2.2 When There Is No Factor Structure

Let us first consider the situation when asset returns do not possess a factor structure, for example, when the asset pool consists of factor portfolios.

2.2.1 Estimating the Maximum Sharpe Ratio and the Regression Response

In our regression representation, the response r_c is unknown[2] and needs to be estimated. The estimation of the response r_c is closely related to the estimation of the maximum Sharpe ratio, which has been considered in Kan and Zhou (2007). It is shown there that the square of the plug-in Sharpe ratio follows a non-centralized F-distribution; see equation (49) in their paper. Utilizing such a fact, we establish the following

Proposition 2. *Define the following estimator of θ:*

$$\widehat{\theta} := \frac{(T-N-2)\widehat{\theta}_s - N}{T}, \tag{2.8}$$

where $\widehat{\theta}_s := \widehat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}$ is the sample estimate of θ. Under normality assumption on returns, and assuming that $N/T \to \rho \in (0,1)$, we have

$$|\widehat{\theta} - \theta| \xrightarrow{P} 0. \tag{2.9}$$

Consequently,

$$\widehat{r_c} := \frac{1+\widehat{\theta}}{\sqrt{\widehat{\theta}}} \tag{2.10}$$

satisfies that

$$|\widehat{r_c} - r_c| \xrightarrow{P} 0. \tag{2.11}$$

We emphasize that our estimation of $\theta = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ is *not* via consistently estimating $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, which would be impossible without imposing strong structural assumptions on them. Instead, we estimate θ directly, and in such a way we overcome the challenge due to high-dimensionality.

2.2.2 A LASSO-type Estimator

On the basis of our unconstrained regression representation (2.7), we aim to estimate the optimal portfolio $\boldsymbol{w}^*$ for a given risk constraint σ.

Suppose that $\boldsymbol{R_t} = (R_{t1}, \cdots, R_{tN})'$, $t = 1, \cdots, T$, are T i.i.d. copies of the (excess) return. Let $\boldsymbol{R} = (\boldsymbol{R_1}, ..., \boldsymbol{R_n})'$ be the $T \times N$ observation matrix. We focus on the case where N and T are both large and the optimal portfolio has a bounded ℓ_1-norm. In terms of

[2] In the case when return constraint r^* is given, the response is still unknown because the bias-correction scalar depends on θ, the square of the underlying maximum Sharpe ratio.

the regression (2.7), this amounts to a high-dimensional regression problem with bounded ℓ_1-norm regression coefficients. Such a connection leads us to the widely used approach in high-dimensional regression, LASSO (Tibshirani (1996)). More specifically, take a $\lambda > 0$ and define a LASSO-type estimator $\widehat{\boldsymbol{w}^*} = \left(\widehat{w_1^*}, ..., \widehat{w_N^*}\right)'$ as follows:

$$\widehat{\boldsymbol{w}^*} = \arg\min_{\boldsymbol{w}} ||\widehat{\boldsymbol{r}_c} - \boldsymbol{R}\boldsymbol{w}||_2^2 \quad \text{subject to } ||\boldsymbol{w}||_1 \leq \lambda, \tag{2.12}$$

where $\widehat{\boldsymbol{r}_c} = (\widehat{r_c}, ..., \widehat{r_c})' \in \mathbb{R}^T$ is the estimated response. Before we give the theoretical property of $\widehat{\boldsymbol{w}^*}$, we first list the assumptions that will be needed.

Assumption:

A1 The (excess) return $\boldsymbol{r} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$;

A2 There exists an $L < \infty$ such that for all dimension N, $\boldsymbol{\Sigma} = (\sigma_{jk})_{N\times N}$ satisfies that $\max_{1\leq j,k\leq N} |\sigma_{jk}| \leq L$;

A3 The eigenvalues of $\boldsymbol{\Sigma}$ are bounded from both above and below;

A4 There exists $M < \infty$ such that for all dimension N, $||\boldsymbol{\mu}||_2 \leq M$;

A5 $||\boldsymbol{w}^*||_1 \leq \lambda$ for some constant λ;

A6 The number of assets N and the sample size T satisfy that $\rho_T := N/T \to \rho \in (0, 1)$.

Remark 2. *About Assumption A1, in this paper we focus on the most fundamental form of the mean-variance problem, so we assume normal distribution of returns. Numerically, we find that our proposed method works well even when heavy-tailedness is present. Extensions incorporating heteroscedasticity and heavy-tailedness will be studied in subsequent papers.*

Remark 3. *Under Assumption A3, Assumption A4 is equivalent to the common belief that the theoretically optimal Sharpe ratio is bounded. Note that Assumption A4 also implies that* $\max_{1\leq j\leq p} |\mu_j| \leq M$.

Remark 4. *Assumption A5 is our sparsity requirement on the optimal portfolio when there is no factor structure. Assumptions A3 and A4 imply that* $||\boldsymbol{w}^*||_2$ *is bounded, so the only difference is between the* ℓ_2 *norm and the* ℓ_1 *norm. Note that Assumption A5 does not require most weights to be zero. For example, it does not rule out value-weighted portfolios. We will see in Section 2.3 that when there is factor structure, the requirement will be only imposed upon the weights on idiosyncratic components.*

More generally, if indeed $||\boldsymbol{w}^*||_1$ *is unbounded, then one can truncate* $\boldsymbol{w}^*$ *at the level of* $\varepsilon/\sqrt{N}$ *for some* $\varepsilon > 0$ *to reduce nonzero but small elements and possibly largely reduce the*

ℓ_1 norm. Note that the difference between $\boldsymbol{w}^*$ and its truncated version has a difference of ℓ_2 norm at most ε. Using Assumptions A3 and A4, one can show that the differences between the variances and expected returns of the two portfolios are both $O(\varepsilon)$. In other words, one sacrifices little in terms of both return and risk and yet can possibly largely reduce the number of nonzero but small positions.

Remark 5. *Assumption A6 says that we are in a high-dimensional setting where the number of assets N and the sample size T proportionally grow up to infinity. We require the sample size T to be larger than the dimension N due to that we need to take inverse of the sample covariance matrix in estimating the maximum Sharpe ratio; see Proposition 2.*

2.2.3 Main Result I: MAXSER Without Factor Structure

We now state our first main result, which establishes the near-optimality of our MAXSER portfolio when returns do not admit factor structures.

Theorem 1. *Under Assumptions A1∼A6, the MAXSER portfolio $\widehat{\boldsymbol{w}^*}$ defined in* (2.12) *with $\widehat{r_c}$ given by* (2.10) *satisfies that, as $N \to \infty$,*

$$||\boldsymbol{w}^* - \widehat{\boldsymbol{w}^*}||_2 \xrightarrow{P} 0, \tag{2.13}$$

$$|r^* - \boldsymbol{\mu}'\widehat{\boldsymbol{w}^*}| \xrightarrow{P} 0, \tag{2.14}$$

and

$$\left|\sqrt{\widehat{\boldsymbol{w}^*}'\boldsymbol{\Sigma}\widehat{\boldsymbol{w}^*}} - \sigma\right| \xrightarrow{P} 0. \tag{2.15}$$

Theorem 1 guarantees that our MAXSER portfolio $\widehat{\boldsymbol{w}^*}$ asymptotically (1) achieves the maximum expected return and (2) satisfies the risk constraint. An immediate implication is that $\widehat{\boldsymbol{w}^*}$ also approaches the mean-variance efficiency.

2.3 When Factor Structure Presents

2.3.1 The Optimal Portfolio: A Factor-Idiosyncratic Components Separation

Motivated by the large literature on factor models for stock returns, we next propose an alternative version of MAXSER under the following approximate factor model:

$$r_i = \alpha_i + \sum_{j=1}^{K} \beta_{ij} f_j + e_i := \sum_{j=1}^{K} \beta_{ij} f_j + u_i, \qquad i = 1, \cdots, N, \tag{2.16}$$

where the β_{ij}'s represent individual stock sensitivities to the factors, and e_i's are idiosyncratic disturbances independent of the factor returns (f_j). The factors can be any well-recognized factors like Fama-French three factors or other factors identified in the large literature of asset pricing (see, e.g., Jegadeesh and Titman (2001) and Korajczyk and Sadka (2008)). They can also be statistical factors identified via principal component analysis (see, e.g., Connor et al. (2010)). As to the idiosyncratic disturbances (e_i)'s, we emphasize that they can still be cross-sectionally dependent. The terms $u_i = \alpha_i + e_i$ will be referred to as the idiosyncratic returns. Model (2.16) can be written in a compact form as

$$\boldsymbol{r} = \boldsymbol{\beta}\boldsymbol{f} + \boldsymbol{u}, \tag{2.17}$$

where $\boldsymbol{\beta} = (\beta_{ij})_{N\times K}$, $\boldsymbol{f} = (f_1, \cdots, f_K)'$, and $\boldsymbol{u} = (u_1, \cdots, u_N)'$. Let $\boldsymbol{\mu}_f$ be the mean of the factor returns, and $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_N)'$ which is the mean of idiosyncratic return $\boldsymbol{u}$. Let $\boldsymbol{\Sigma}_f$ and $\boldsymbol{\Sigma}_u$ be the covariance matrix of factor and idiosyncratic returns, respectively. Then the return vector $\boldsymbol{r}$ has the following mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{\mu} = \boldsymbol{\beta}\boldsymbol{\mu}_f + \boldsymbol{\alpha}, \qquad \boldsymbol{\Sigma} = \boldsymbol{\beta}\boldsymbol{\Sigma}_f\boldsymbol{\beta}' + \boldsymbol{\Sigma}_u. \tag{2.18}$$

Denote the mean and covariance matrix of the returns on the full pool of factors and assets by $\boldsymbol{\mu}_{all}$ and $\boldsymbol{\Sigma}_{all}$, respectively.

Such a model has been widely adopted in the literature. In particular, in light of Section 5 of Chamberlain and Rothschild (1983), we make the following

Assumption:

B1 The eigenvalues of $\boldsymbol{\Sigma}_u := \mathrm{Cov}(\boldsymbol{u})$ is bounded from both above and below, and $\sigma_{ij} := \boldsymbol{\Sigma}_u(i,j)$ satisfies that $\max_{1\le i,j\le N} |\sigma_{i,j}| \le L < \infty$.

B2 $||\boldsymbol{\alpha}||_2 := \sqrt{\sum_{j=1}^N \alpha_j^2}$ is bounded. In particular, there exists $M < \infty$ such that $\max_{1\le j\le N} |\alpha_j| \le M$.

B3 The number of factors K is bounded.

Given the factor structure, when we build portfolios, we shall invest in not only the N assets but also the K factors. Such a strategy is straightforward when the factors are taken to be investable factors like the Fama-French factors and many others (see, e.g., the supplementary file of Feng et al. (2017)), and is also feasible when the factors are statistical factors thanks to that both the number of factors K and the factors can be consistently estimated (see, for example, Bai and Ng (2002)).

We aim to find an optimal weight $(w_1^f, \dots, w_K^f; w_1, \dots, w_N) := (\boldsymbol{w}_f, \boldsymbol{w})$, where $\boldsymbol{w}_f$ and $\boldsymbol{w}$ represent the weight vectors put on the K factors and N assets, respectively. The following result shows that finding such an optimal weight can be decomposed into three steps:

(i) Find the optimal portfolio on the factors with 1 unit of risk (denoted by $\boldsymbol{w}_f^*$),

(ii) Find the optimal portfolio on the idiosyncratic components with 1 unit of risk (denoted by $\boldsymbol{w}_u^*$), and

(iii) Suitably combine these two portfolios.

Proposition 3. *For any given risk constraint level σ, the optimal portfolio $\boldsymbol{w}_{all} := (\boldsymbol{w}_f, \boldsymbol{w})$ is given by*

$$\left(\sigma\sqrt{\frac{\theta_f}{\theta_{all}}}\boldsymbol{w}_f^* - \sigma\sqrt{\frac{\theta_u}{\theta_{all}}}\boldsymbol{\beta}'\boldsymbol{w}_u^*, \quad \sigma\sqrt{\frac{\theta_u}{\theta_{all}}}\boldsymbol{w}_u^*\right),$$

where $\theta_f = \boldsymbol{\mu}_f'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\mu}_f$, $\theta_u = \boldsymbol{\alpha}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\alpha}$, and $\theta_{all} = \boldsymbol{\mu}_{all}'\boldsymbol{\Sigma}_{all}^{-1}\boldsymbol{\mu}_{all}$ are the squared maximum Sharpe ratios of portfolios on the factors, the idiosyncratic components, and the full set of factors and individual assets, respectively. Moreover, $\boldsymbol{w}_f^*$ and $\boldsymbol{w}_u^*$ admit the following explicit expressions:

$$\boldsymbol{w}_f^* = \frac{1}{\sqrt{\theta_f}}\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\mu}_f, \quad \boldsymbol{w}_u^* = \frac{1}{\sqrt{\theta_u}}\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\alpha}. \tag{2.19}$$

Remark 6. *In the case when $\alpha_i \equiv 0$, we have $\theta_u = 0$ and hence the optimal weight is given by $\big(\sigma\boldsymbol{w}_f^*, \mathbf{0}\big)$. In other words, the optimal portfolio will be fully invested in factors. However, this is only the case when all the underlying factors are identified and included, which is unlikely in practice especially when one wants to only include a small number of strong factors.*

According to Proposition 3, in order to estimate the optimal portfolio $(\boldsymbol{w}_f, \boldsymbol{w})$, we need to estimate θ_f, θ_u, $\boldsymbol{w}_f^*$ and $\boldsymbol{w}_u^*$. We will deal with them one by one, starting with the estimation of the maximum Sharpe ratios.

2.3.2 Estimating the Maximum Sharpe ratios

Estimation of the maximum Sharpe ratio is still essential under the current setting with factor structure. There are three Sharpe ratios that are of interest in this setting: θ_f, θ_u and θ_{all}, the squared maximum Sharpe ratios on factors, idiosyncratic components and all assets, respectively. For estimating θ_f and θ_{all}, parallel to Proposition 2 we have the following

Proposition 4. *Define the following estimators of* θ_f *and* θ_{all}*:*

$$\widehat{\theta}_f := \frac{(T-K-2)\widehat{\theta}_{sf} - K}{T}, \tag{2.20}$$

$$\widehat{\theta}_{all} := \frac{(T-N-K-2)\widehat{\theta}_{sa} - N - K}{T}, \tag{2.21}$$

where $\widehat{\theta}_{sf} := \widehat{\boldsymbol{\mu}}'_f \widehat{\boldsymbol{\Sigma}}_f^{-1} \widehat{\boldsymbol{\mu}}_f$ *and* $\widehat{\theta}_{sa} := \widehat{\boldsymbol{\mu}}'_{all} \widehat{\boldsymbol{\Sigma}}_{all}^{-1} \widehat{\boldsymbol{\mu}}_{all}$ *are the sample estimates of* θ_f *and* θ_{all}*, respectively. Under normality assumption on factor and idiosyncratic returns, assuming that* $N/T \to \rho \in (0,1)$*, we have*

$$|\widehat{\theta}_f - \theta_f| \xrightarrow{P} 0, \quad \text{and} \quad |\widehat{\theta}_{all} - \theta_{all}| \xrightarrow{P} 0.$$

There is one more Sharpe ratio to be estimated, $\sqrt{\theta_u}$, the maximum Sharpe ratio on the idiosyncratic components. This quantity is a bit trickier to deal with, because the idiosyncratic return $\boldsymbol{U}$ is not observable. A natural idea is to work with $\widehat{\boldsymbol{U}}$, the estimated idiosyncratic return. However, it can be shown that an estimator similar to (2.20) and (2.21) applied to $\widehat{\boldsymbol{U}}$ will be biased.

The solution to the aforementioned difficulty lies in the relationship among θ_f, θ_u and θ_{all}. Based on the model (2.17), one can show that

$$\theta_{all} = \theta_f + \theta_u.$$

By Proposition 4, both θ_f and θ_u can be consistently estimated, so we get the following

Proposition 5. *Define*

$$\widehat{\theta}_u := \widehat{\theta}_{all} - \widehat{\theta}_f.$$

Under the assumptions of Proposition 4, we have

$$|\widehat{\theta}_u - \theta_u| \xrightarrow{P} 0.$$

Therefore for $r_c := (1+\theta_u)/\sqrt{\theta_u}$*, if we define*

$$\widehat{r_c} = \frac{1+\widehat{\theta}_u}{\sqrt{\widehat{\theta}_u}}, \tag{2.22}$$

then

$$|\widehat{r_c} - r_c| \xrightarrow{P} 0.$$

2.3.3 Estimating the Optimal Portfolio on Idiosyncratic Components

The optimal portfolio on the idiosyncratic components, $\boldsymbol{w}_u^*$, solves the following Markowitz optimization problem:

$$\arg\max_{\boldsymbol{w}} \boldsymbol{\alpha}'\boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{w}'\boldsymbol{\Sigma}_u\boldsymbol{w} \leq 1. \tag{2.23}$$

By (2.19), the optimal portfolio yields an expected return of

$$r_u^* = \sqrt{\theta_u}, \tag{2.24}$$

which is the maximum Sharpe ratio of portfolios on (u_i). Following our regression representation in Section 2.1, we will estimate $\boldsymbol{w}_u^*$ based on the following:

$$\arg\min_{\boldsymbol{w}} E\left(r_c - \boldsymbol{u}'\boldsymbol{w}\right)^2, \quad \text{where } r_c := \frac{1+\theta_u}{\theta_u} r_u^* = \frac{1+\theta_u}{\sqrt{\theta_u}}. \tag{2.25}$$

One major difference here is that the idiosyncratic returns are not observable.

Suppose that $\boldsymbol{R_t} = (R_{t1}, \cdots, R_{tN})'$ and $\boldsymbol{F_t} = (F_{t1}, \cdots, F_{tK})'$, $t = 1, \cdots, T$, are T i.i.d. copies of the (excess) return $\boldsymbol{r}$ and the factor (excess) return $\boldsymbol{f}$, respectively. Let $\boldsymbol{R} = (\boldsymbol{R_1}, ..., \boldsymbol{R_T})'$ and $\boldsymbol{F} = (\boldsymbol{F_1}, \cdots, \boldsymbol{F_T})'$. We estimate the coefficient $\boldsymbol{\beta}$ in (2.17) by regressing $\boldsymbol{R}$ on $\boldsymbol{F}$. Denote by $\widehat{\boldsymbol{\beta}}$ the estimated beta matrix. Correspondingly, let $\widehat{\boldsymbol{U}} = (\widehat{\boldsymbol{u}}_1, \cdots, \widehat{\boldsymbol{u}}_N)' = \boldsymbol{R} - \boldsymbol{F}\widehat{\boldsymbol{\beta}}$ be the estimator of $\boldsymbol{U} = \boldsymbol{R} - \boldsymbol{F}\boldsymbol{\beta}$. Our estimation of $\boldsymbol{w}_u^*$ will be based on the $T \times N$ matrix $\widehat{\boldsymbol{U}}$.

Similarly to Section 2.2.2, our estimator of $\boldsymbol{w}_u^*$ is the following LASSO-type estimator:

$$\widehat{\boldsymbol{w}_u^*} = \arg\min_{\boldsymbol{w}} ||\widehat{\boldsymbol{r}_c} - \widehat{\boldsymbol{U}}\boldsymbol{w}||_2^2 \quad \text{subject to} \quad \sum_{i=1}^N |w_i| \leq \lambda, \tag{2.26}$$

where $\widehat{\boldsymbol{r}_c} = (\widehat{r_c}, \cdots, \widehat{r_c})'$ is the estimator of the response defined in (2.22), and λ is a positive constant. To give the theoretical properties of $\widehat{\boldsymbol{w}_u^*}$, the following assumptions will be needed.

Assumption:

C1 $\boldsymbol{f} \sim N(\boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f)$, $\boldsymbol{u} \sim N(\boldsymbol{\alpha}, \boldsymbol{\Sigma}_u)$;

C2 $||\boldsymbol{w}_u^*||_1 \leq \lambda$ for some constant λ;

C3 The number of assets N and the sample size T satisfy that $\rho_T := N/T \to \rho \in (0, 1)$.

Remark 7. *Assumption C2 is our sparsity requirement under factor models. We emphasize that the sparsity assumption is put on the weights on the* ***idiosyncratic components****. As* *discussed in Remark 6, $\boldsymbol{w}_u^* \equiv 0$ if $\alpha_i \equiv 0$, and the assumption is certainly satisfied. More generally, $||\alpha||_2$ is bounded by Assumption B2. If there are only finitely many none zero of α_i, then Assumption C2 would be satisfied when $\max_i \sum_{j=1,\ldots,N} |\boldsymbol{\Omega}_u(i,j)|$ is bounded, where $\boldsymbol{\Omega}_u := \boldsymbol{\Sigma}_u^{-1}$. Recall that under the Gaussian assumption C1, $\boldsymbol{\Omega}_u(i,j) = 0$ if and only if u_i and u_j are conditionally independent given $\{u_k : k \neq i, j\}$, and more generally, $\boldsymbol{\Omega}_u(i,j)$ is related to the partial correlation between u_i and u_j. Because (u_i) are idiosyncratic returns, it is possibly reasonable to expect that for each i, u_i is only correlated with finitely many other (u_j)'s, in which case the assumption naturally holds.*

We are now ready to give the asymptotic properties of the portfolio $\widehat{\boldsymbol{w}_u^*}$. Note that because $\widehat{\boldsymbol{U}}$ contains estimation errors, Theorem 1 does not readily apply to this case.

Proposition 6. *Under Assumptions B1 $\sim$ B3 and C1 $\sim$ C3, we have as $N \to \infty$,*

$$||\boldsymbol{w}_u^* - \widehat{\boldsymbol{w}_u^*}||_2 \xrightarrow{P} 0, \tag{2.27}$$

$$|\boldsymbol{\alpha}'\boldsymbol{w}_u^* - \boldsymbol{\alpha}'\widehat{\boldsymbol{w}_u^*}| \xrightarrow{P} 0, \tag{2.28}$$

and

$$\left|\sqrt{\widehat{\boldsymbol{w}_u^*}'\boldsymbol{\Sigma}_u\widehat{\boldsymbol{w}_u^*}} - 1\right| \xrightarrow{P} 0. \tag{2.29}$$

Proposition 6 states that the portfolio $\widehat{\boldsymbol{w}_u^*}$ asymptotically attains the maximum expected return and carries a risk that is close to the given risk constraint, in this case, 1.

2.3.4 Main Result II: MAXSER Under Factor Models

By far we have achieved consistency in estimating $\boldsymbol{w}_u^*$, θ_u and θ_f. There is one more item to be estimated, $\boldsymbol{w}_f^*$, the optimal weight on factors (with risk equal to 1). This is easy because the number of factors is fixed, and the simple "plug-in" estimator works. Combining these results with Proposition 3, we obtain the following main result for our estimator of the optimal full portfolio $\widehat{\boldsymbol{w}_{all}}$.

Theorem 2. *Let $\widehat{\boldsymbol{w}_f^*} := \frac{1}{\sqrt{\widehat{\theta}_f}}\widehat{\boldsymbol{\Sigma}}_f^{-1}\widehat{\boldsymbol{\mu}}_f$ be the estimator of $\boldsymbol{w}_f^*$. Our estimator of the optimal full portfolio $\boldsymbol{w}_{all}$ is*

$$\widehat{\boldsymbol{w}_{all}} := (\widehat{\boldsymbol{w}}_f, \widehat{\boldsymbol{w}}) = \left(\sigma\sqrt{\frac{\widehat{\theta}_f}{\widehat{\theta}_{all}}}\widehat{\boldsymbol{w}_f^*} - \sigma\sqrt{\frac{\widehat{\theta}_u}{\widehat{\theta}_{all}}}\widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{w}_u^*},\ \sigma\sqrt{\frac{\widehat{\theta}_u}{\widehat{\theta}_{all}}}\widehat{\boldsymbol{w}_u^*}\right). \tag{2.30}$$

Under Assumptions B1 $\sim$ B3 and C1 $\sim$ C3, as $N \to \infty$, we have

$$|\widehat{\boldsymbol{w}_{all}}'\boldsymbol{\mu}_{all} - r^*| \xrightarrow{P} 0, \quad \textit{and} \quad |\widehat{\boldsymbol{w}_{all}}'\boldsymbol{\Sigma}_{all}\widehat{\boldsymbol{w}_{all}} - \sigma^2| \xrightarrow{P} 0. \tag{2.31}$$

Theorem 2 guarantees that our MAXSER portfolio under the factor model setting can again asymptotically achieve the maximum expected return and meanwhile satisfy the risk constraint, and consequently, achieve mean-variance efficiency.

2.4 Practical Implementation of MAXSER

2.4.1 Choosing λ in (2.12) and (2.26)

In practice, it is important to choose λ in (2.12) and (2.26). Because one of our goals is to meet the risk constraint, we naturally want to choose a λ such that the estimated portfolio possesses a risk that is close to the given risk constraint. In practice, we do not know the underlying covariance matrix $\boldsymbol{\Sigma}/\boldsymbol{\Sigma}_{all}$. To circumvent this difficulty, we use a cross-validation method to choose the λ. Specifically, for a 10-fold cross-validation procedure, we randomly split the sample into 10 groups to form 10 validation sets. For each validation set, the training set is taken to be the rest of the observations. Next, for each such training set i, we let λ vary from 0 to the maximum value (the ℓ_1-norm of the least squares solution) to obtain the whole solution path $\left(\widehat{\boldsymbol{w}^*}(\lambda)\right)_\lambda$ $((\widehat{\boldsymbol{w}_{all}}(\lambda))_\lambda$ under factor model setting), and find the value of λ such that the estimated portfolio minimizes the difference between the "out-of-sample" risk computed using the validation set and the given risk constraint. Denote such a value by $\lambda(i)$. The ultimate $\widehat{\lambda}$ is then taken to be the average of $(\lambda(i),\ i = 1, \cdots, 10)$.

To our knowledge, the above cross-validation method for determining the constraint parameter in norm-constrained mean-variance portfolio optimization is new and constitutes another contribution of ours to the literature. By choosing the parameter selection criterion to be the risk, our cross-validation method effectively helps control out-of-sample risk. The procedure can be easily applied to other portfolio optimizations where norm-constraints are imposed. In our numerical studies, we apply such a cross-validation method to other norm-constrained portfolios and compare their performance with ours.

2.4.2 Adjustment of $\widehat{\theta}$, $\widehat{\theta}_f$, $\widehat{\theta}_u$ and $\widehat{\theta}_{all}$

Kan and Zhou (2007) notice that the unbiased estimator of the square of maximum Sharpe ratio $\widehat{\theta}$ often takes negative values, and they propose the adjusted estimator that improves over the unbiased one:

$$\widehat{\theta}_a = \frac{(T-N-2)\widehat{\theta}_s - N}{T} + \frac{2(\widehat{\theta}_s)^{N/2}(1+\widehat{\theta}_s)^{-(T-2)/2}}{TB_{\widehat{\theta}_s/(1+\widehat{\theta}_s)}(N/2,(T-N)/2)}, \tag{2.32}$$

where, recall that, $\widehat{\theta}_s$ is the plug-in estimators of θ, and

$$B_x(a,b) = \int_0^x y^{a-1}(1-y)^{b-1}dy.$$

Under the factor model setting, we adopt the following adjustments of $\widehat{\theta}_f$ and $\widehat{\theta}_{all}$:

$$\widehat{\theta}_{fa} = \frac{(T-K-2)\widehat{\theta}_{sf} - K}{T} + \frac{2(\widehat{\theta}_{sf})^{K/2}(1+\widehat{\theta}_{sf})^{-(T-2)/2}}{TB_{\widehat{\theta}_{sf}/(1+\widehat{\theta}_{sf})}(K/2,(T-K)/2)}, \tag{2.33}$$

$$\widehat{\theta}_{all,a} = \frac{(T-N-K-2)\widehat{\theta}_{sa} - N - K}{T} + \frac{2(\widehat{\theta}_{sa})^{(N+K)/2}(1+\widehat{\theta}_{sa})^{-(T-2)/2}}{TB_{\widehat{\theta}_{sa}/(1+\widehat{\theta}_{sa})}((N+K)/2,(T-N-K)/2)}, \tag{2.34}$$

where, recall that, $\widehat{\theta_{sf}}$ and $\widehat{\theta_{sa}}$ are the plug-in estimators of θ_f and θ_{all}. The adjusted $\widehat{\theta}_u$ is $\widehat{\theta}_{ua} := \widehat{\theta}_{all,a} - \widehat{\theta}_{fa}$.

2.4.3 Implementation Steps

When there is no factor structure

To sum up, our method consists of the following steps:

Step 1 Compute the estimates of the square of the maximum Sharpe ratios $\widehat{\theta}$ by (2.32), and compute the response $\widehat{r_c}$;

Step 2 Choose λ by cross-validation according to the procedure described in Section 2.4.1. Denote the chosen value by $\widehat{\lambda}$;

Step 3 Set λ in (2.12) to be $\widehat{\lambda}$ and solve for $\widehat{\boldsymbol{w}^*}$, the MAXimum-Sharpe-ratio Estimated sparse Regression (MAXSER) portfolio.

When factor structure presents

Under the factor model setting, MAXSER is implemented as follows:

Step 1 Perform OLS regressions of observed asset returns $\boldsymbol{X}$ on observed factor returns $\boldsymbol{F}$ to obtain $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{U}}$;

Step 2 Compute the estimates of the square of the maximum Sharpe ratios $\widehat{\theta}_f$, $\widehat{\theta}_{all}$ and $\widehat{\theta}_u$, and compute the response $\widehat{r_c}$;

Step 3 Choose λ by cross-validation according to the procedure described in Section 2.4.1 (based on risk $\widehat{\boldsymbol{w}_{all}}'\boldsymbol{\Sigma}_{all}\widehat{\boldsymbol{w}_{all}}$). Denote the chosen value by $\widehat{\lambda}$;

Step 4 Set λ in (2.26) to be $\widehat{\lambda}$ and solve for $\widehat{\boldsymbol{w}_u^*}$;

Step 5 Compute $\widehat{\boldsymbol{w}_f^*}$ and combine the estimates from the previous steps to obtain the MAXSER portfolio weight $\widehat{\boldsymbol{w}_{all}}$.

3 SIMULATION STUDIES

3.1 Methods to be Compared with

In addition to the plug-in and nonlinear shrinkage portfolios that we discussed in the Introduction, we include several other strategies in our simulation comparisons. The complete list is given in Table 1.

Table 1

List of portfolios under comparison and their abbreviations. "MV" stands for mean-variance portfolio, and "GMV" stands for global minimum variance portfolio.

Portfolio	Abbreviation
Plug-in MV on factors	Factor
Three-fund portfolio by Kan and Zhou (2007)	KZ
MV/GMV with estimated covariance matrix inputs	
MV with sample cov	MV-P
MV with linear shrinkage cov	MV-LS
MV with nonlinear shrinkage cov	MV-NLS
Nonlinear shrinkage adjusted for factor models by Ledoit and Wolf (2017)	NLSF
GMV with linear shrinkage cov	GMV-LS
GMV with nonlinear shrinkage cov	GMV-NLS
MV with no-short-sale constraint	
MV with sample cov & no-short-sale constraint	MV-P-NSS
MV with linear shrinkage cov & no-short-sale constraint	MV-LS-NSS
MV with nonlinear shrinkage cov & no-short-sale constraint	MV-NLS-NSS
MV with short-sale constraint & cross-validation	
MV with sample cov & short-sale-CV	MV-P-SSCV
MV with linear shrinkage cov & short-sale-CV	MV-LS-SSCV
MV with nonlinear shrinkage cov & short-sale-CV	MV-NLS-SSCV

Among the portfolios under comparison, a special one is the portfolio Factor, which is the Markowitz portfolio on factors. Specifically, suppose $\widehat{\boldsymbol{\mu}_f}$ and $\widehat{\boldsymbol{\Sigma}_f}$ are the sample mean and sample covariance matrix computed from the observed factor returns, then the Factor portfolio has the following explicit form:

$$\widehat{\boldsymbol{w}}_{Fac} := \frac{\sigma}{\sqrt{\widehat{\boldsymbol{\mu}_f}'\widehat{\boldsymbol{\Sigma}_f}^{-1}\widehat{\boldsymbol{\mu}_f}}}\widehat{\boldsymbol{\Sigma}_f}^{-1}\widehat{\boldsymbol{\mu}_f}. \tag{3.1}$$

This portfolio is special in the sense that it only involves a small number of assets (3 in our case). Consequently, the plug-in formula (3.1) indeed gives a nearly optimal portfolio. Including such a portfolio in the comparison would reveal whether there is benefit to invest in idiosyncratic components.

On the other hand, the MV/GMV portfolios are constructed by replacing covariance matrix with the sample/linear shrinkage/nonlinear shrinkage (adjusted for factor model) estimators in the formulas of MV/GMV portfolio weights. Details about portfolios "KZ"[3] and "NLSF" can be found in Kan and Zhou (2007) and Ledoit and Wolf (2017), respectively.

In addition, we construct portfolios with either no-short-sale or short-sale constraints on portfolio weights. The "MV-P-NSS", "MV-LS-NSS" and "MV-NLS-NSS" portfolios are with no-short-sale constraints, and are using the sample/linear shrinkage/nonlinear shrinkage covariance matrix, respectively. More generally, the MV portfolios with *short-sale* constraints[4], "MV-P-SSCV", "MV-LS-SSCV" and "MV-NLS-SSCV", are having short position thresholds determined by the cross-validation method that we proposed in Section 2.4.1. In such a way, these portfolios enjoy the same benefit as our MAXSER portfolio in terms of risk control. We include these portfolios to demonstrate the effectiveness of our cross-validation procedure, and that the advantage of MAXSER is not only due to the ℓ_1-norm constraints, but rather, more fundamentally, due to its methodology.

[3] Following Kan and Zhou (2007), the risk aversion is set to be 3.

[4] Specifically, suppose that $\widehat{\boldsymbol{\mu}}_{all}$ is the sample mean of returns on stocks and factors, and $\widehat{\boldsymbol{\Sigma}}_{all}$ is an estimate of the covariance matrix, which can be the sample/linear shrinkage/nonlinear shrinkage covariance matrix. Then the MV portfolio with short-sale constraint is solved by

$$\boldsymbol{w}_{SSCV} = \arg\max_{\boldsymbol{w}} \boldsymbol{w}'\widehat{\boldsymbol{\mu}}_{all}, \quad \text{subject to} \quad \boldsymbol{w}'\widehat{\boldsymbol{\Sigma}}_{all}\boldsymbol{w} \leq \sigma^2 \text{ and } w_i > -\lambda_{SS} \text{ for all } i.$$

Here $\lambda_{SS} > 0$ is the short position threshold determined via a 10-fold cross-validation as follow: split the sample into 10 groups of validation sets, with the rest of the observations being the corresponding training set. For each training set, we solve the optimization for a sequence of λ_{SS} to get a solution path, and find the value of λ_{SS} such that the difference between the risk on validation sets and the given constraint is minimized.

3.2 Parameter Setting

We simulate data from a three-factor model with the parameters calibrated from real data. Specifically, out of the stocks that stayed in S&P 500 index during the period of 2007 – 2016, we randomly pick 100 of them. We then regress the monthly excess returns on the 100 stocks over the Fama-French three factor (FF3) returns, and set the resulting slopes to be the β_i's; the α_i's in (2.16) are obtained by hard thresholding the estimated intercepts with a threshold of 2 standard errors. The covariance matrix of idiosyncratic returns, $\mathbf{\Sigma}_u$, is obtained by applying the soft-thresholding method proposed in Rothman (2012)[5] to the sample covariance matrix of the residuals in the regression above. For the parameters of factors, the mean and covariance matrix are taken to be the sample mean and the sample covariance matrix of the Fama-French three factor returns, respectively.

3.3 Simulation Comparisons

3.3.1 When returns are normally distributed

We first show simulation results for data generated under multivariate normal distribution. Returns of 100 stocks and 3 factors are generated using the parameters described in Section 3.2. The level of risk constraint is fixed to be $\sigma = 0.04$.

We run 1,000 replications to evaluate the performance of the portfolios in terms of the risk and (annualized) Sharpe ratio. The comparison results for sample sizes $T = 120$ and 240 are summarized in Tables 2 and 3.

[5]The soft-thresholding method can be implemented in R by the package "PDSCE". In this case the penalty parameter "lam" is set to be 0.5.

Table 2

Simulation comparison of risks and Sharpe ratios of the portfolios under comparison based on 1,000 replications where returns are generated from multivariate normal distribution. "True" stands for the theoretical optimal portfolio. The risk constraint is set to be 0.04. Both average value and standard deviation (in brackets) of each performance measure are reported.

Normal Distribution	$\sigma = 0.04$	$T = 120$
Portfolio	Risk	Sharpe Ratio
True	0.040	1.882
Factor	0.041 (0.003)	0.401 (0.169)
KZ	0.052 (0.040)	0.329 (0.184)
MAXSER	0.041 (0.005)	**1.183** (0.276)
MV/GMV with estimated covariance matrix inputs		
MV-P	0.296 (0.072)	0.367 (0.168)
MV-LS	0.082 (0.006)	0.697 (0.160)
MV-NLS	0.054 (0.017)	0.945 (0.183)
NLSF	0.044 (0.002)	0.837 (0.139)
GMV-LS	0.013 (0.001)	0.438 (0.132)
GMV-NLS	0.015 (0.003)	0.553 (0.148)
MV with no-short-sale constraint		
MV-P-NSS	0.044 (0.003)	0.399 (0.040)
MV-LS-NSS	0.044 (0.003)	0.409 (0.036)
MV-NLS-NSS	0.043 (0.003)	0.416 (0.035)
MV with short-sale constraint & cross-validation		
MV-P-SSCV	0.044 (0.003)	0.399 (0.040)
MV-LS-SSCV	0.044 (0.003)	0.409 (0.036)
MV-NLS-SSCV	0.044 (0.004)	0.501 (0.169)

Table 3

Simulation comparison of risks and Sharpe ratios of the portfolios under comparison based on 1,000 replications where returns are generated from multivariate normal distribution. "True" stands for the theoretical optimal portfolio. The risk constraint is set to be 0.04. Both average value and standard deviation (in brackets) of each performance measure are reported.

Normal Distribution	$\sigma = 0.04$	$T = 240$
Portfolio	Risk	Sharpe Ratio
True	0.040	1.882
Factor	0.041 (0.002)	0.467 (0.108)
KZ	0.091 (0.031)	0.909 (0.130)
MAXSER	0.040 (0.003)	**1.508** (0.154)
MV/GMV with estimated covariance matrix inputs		
MV-P	0.070 (0.005)	0.911 (0.123)
MV-LS	0.061 (0.004)	0.943 (0.117)
MV-NLS	0.049 (0.004)	1.199 (0.117)
NLSF	0.042 (0.001)	1.068 (0.104)
GMV-LS	0.009 (0.000)	0.450 (0.102)
GMV-NLS	0.009 (0.001)	0.539 (0.167)
MV with no-short-sale constraint		
MV-P-NSS	0.042 (0.002)	0.415 (0.032)
MV-LS-NSS	0.042 (0.002)	0.420 (0.031)
MV-NLS-NSS	0.041 (0.002)	0.427 (0.030)
MV with short-sale constraint & cross-validation		
MV-P-SSCV	0.042 (0.002)	0.415 (0.032)
MV-LS-SSCV	0.042 (0.002)	0.420 (0.031)
MV-NLS-SSCV	0.042 (0.003)	0.468 (0.137)

From Tables 2 and 3, we observe that

- In terms of *risk control*,
  - The risk of our MAXSER portfolio is close to the given constraint, whereas MV portfolios with covariance matrix estimated by sample, linear shrinkage and nonlinear shrinkage estimators violate the risk constraint by about 640%, 105%

and 35% when $T = 120$, respectively.

  - The risks of Factor, NLSF, and MV portfolios with (no-)short-sale constraints are similar to the risk of MAXSER.

- In terms of *Sharpe ratio*,

  - MAXSER achieves the highest Sharpe ratio among all portfolios under comparison. In the $T = 240$ case, MAXSER achieves about 80% of the theoretical maximum Sharpe ratio on average, whereas the Sharpe ratio of the MV-NLS portfolio, the second highest among all portfolios, is about 64% of the theoretical maximum value on average.
  - For the $T = 240$ case, the 95% confidence interval of the mean Sharpe ratio of MAXSER is $[1.498, 1.518]$. In terms of comparison, the 95% confidence interval for the difference between the mean Sharpe ratios of MAXSER and MV-NLS is $[0.301, 0.317]$. In particular, we conclude that the higher Sharpe ratio of MAXSER is statistically significant.
  - Compared with MV-P-SSCV, MV-LS-SSCV and MV-NLS-SSCV portfolios, the Sharpe ratio of MAXSER is substantially higher, indicating that the outstanding performance of MAXSER is fundamental and does not rely on cross-validation only.

- In summary,

  - Our MAXSER portfolio effectively controls risk, and is significantly more mean-variance efficient than the other portfolios.
  - The comparison with Factor portfolio suggests that investing in idiosyncratic components, or equivalently, individual stocks, helps improve the mean-variance efficiency.

3.3.2 When returns are heavy-tailed

Given the empirical evidence that financial returns tend to be heavy-tailed, in the following we conduct a simulation study for data with heavy-tails. More specifically, we shall let the factor and idiosyncratic returns be all *Student-t* distributed with 6 degrees of freedom. The mean and covariance matrix parameters are taken to be the same as in Section 3.2.

Table 4

Simulation comparison of risks and Sharpe ratios of the portfolios under comparison based on 1,000 replications where returns are generated from t-distribution with 6 degrees of freedom. The underlying mean and covariance matrix are the same as in Section 3.3.1. "True" stands for the theoretical optimal portfolio. The risk constraint is set to be 0.04. Both average value and standard deviation (in brackets) of each performance measure are reported.

$t(6)$ Distribution	$\sigma = 0.04$	$T = 120$
Portfolio	Risk	Sharpe Ratio
True	0.040	1.882
Factor	0.034 (0.003)	0.350 (0.202)
KZ	0.039 (0.031)	0.288 (0.191)
MAXSER	0.033 (0.005)	**1.035** (0.281)
MV/GMV with estimated covariance matrix inputs		
MV-P	0.246 (0.060)	0.321 (0.174)
MV-LS	0.062 (0.005)	0.635 (0.169)
MV-NLS	0.042 (0.009)	0.845 (0.179)
NLSF	0.036 (0.002)	0.716 (0.150)
GMV-LS	0.013 (0.001)	0.459 (0.130)
GMV-NLS	0.014 (0.003)	0.572 (0.125)
MV with no-short-sale constraint		
MV-P-NSS	0.036 (0.003)	0.394 (0.039)
MV-LS-NSS	0.036 (0.003)	0.406 (0.035)
MV-NLS-NSS	0.035 (0.003)	0.411 (0.036)
MV with short-sale constraint & cross-validation		
MV-P-SSCV	0.036 (0.003)	0.394 (0.039)
MV-LS-SSCV	0.036 (0.003)	0.406 (0.035)
MV-NLS-SSCV	0.037 (0.004)	0.531 (0.196)

Table 5

Simulation comparison of risks and Sharpe ratios of the portfolios under comparison based on 1,000 replications where returns are generated from t-distribution with 6 degrees of freedom. The underlying mean and covariance matrix are the same as in Section 3.3.1. "True" stands for the theoretical optimal portfolio. The risk constraint is set to be 0.04. Both average value and standard deviation (in brackets) of each performance measure are reported.

$t(6)$ Distribution	$\sigma = 0.04$	$T = 240$
Portfolio	Risk	Sharpe Ratio
True	0.040	1.882
Factor	0.033 (0.002)	0.427 (0.141)
KZ	0.059 (0.023)	0.802 (0.154)
MAXSER	0.033 (0.003)	**1.374** (0.203)
MV/GMV with estimated covariance matrix inputs		
MV-P	0.058 (0.004)	0.807 (0.140)
MV-LS	0.048 (0.003)	0.847 (0.133)
MV-NLS	0.040 (0.004)	1.071 (0.138)
NLSF	0.034 (0.001)	0.931 (0.117)
GMV-LS	0.010 (0.000)	0.469 (0.107)
GMV-NLS	0.010 (0.001)	0.538 (0.182)
MV with no-short-sale constraint		
MV-P-NSS	0.034 (0.002)	0.406 (0.035)
MV-LS-NSS	0.034 (0.002)	0.412 (0.033)
MV-NLS-NSS	0.034 (0.002)	0.418 (0.032)
MV with short-sale constraint & cross-validation		
MV-P-SSCV	0.034 (0.002)	0.406 (0.035)
MV-LS-SSCV	0.034 (0.002)	0.414 (0.046)
MV-NLS-SSCV	0.035 (0.003)	0.529 (0.221)

Tables 4 and 5 show that MAXSER portfolio continues to clearly outperform other portfolios. Another observation is that, if we compare Tables 4 and 5 with Tables 2 and 3 for the normal case, we see that heavy-tailedness does to some extent hurt all the strategies in terms of Sharpe ratios.

4 EMPIRICAL STUDIES

We investigate the performance of our strategy through two types of empirical studies:

- *Practical* evaluation:

  The asset pool containing the constituents of the DJIA 30 index is considered. Under a rolling-window scheme to be specified below, at each rebalancing time point, only the constituents at that time are considered to invest in. Portfolio performances are compared on the basis of both raw returns and returns net of transaction costs; see Section 4.2 for the details.

- *General statistical* evaluation:

  We compare the performance of the compared strategies using 100 random datasets, in which the stocks are randomly picked historical constituents of the S&P 500 index. See Section 4.3 for the details.

4.1 Portfolios Under Comparison in Empirical Study

In addition to the strategies we compared in simulation studies, we further investigate the performance of five more portfolios in our empirical study, including the index, the equally weighted portfolio ("1/N" rule), and three ℓ_1-norm constrained mean-variance portfolios. Based on different covariance matrices estimated by the sample estimator, the linear shrinkage estimator (Ledoit and Wolf (2004)), and the nonlinear shrinkage estimator (Ledoit and Wolf (2017)), we construct the portfolios "MV-P-l1CV", "MV-LS-l1CV" and "MV-NLS-l1CV" by imposing the ℓ_1-norm constraint[6] for which the tuning parameter λ is determined by the cross-validation method we proposed in Section 2.4.1. We include these ℓ_1-norm constrained portfolios to examine the effect of imposing ℓ_1-norm constraint, and, more importantly, to demonstrate that the advantage of MAXSER is not only due to the constraint and cross-validation, but rather more due to its methodology.

[6] Due to the time cost of solving ℓ_1-norm constrained mean-variance optimizations, the latter three ℓ_1-norm constrained methods are only applied to the DJIA data with around 30 stocks.

4.2 Practical Evaluation

4.2.1 Data & Investment Rolling–Window Scheme

We first evaluate our proposed portfolio, MAXSER, based on the stock universe of DJIA 30 index constituents. We obtain the lists of DJIA 30 index constituents from COMPUSTAT and CRSP. Fama-French three factors are also included in our asset pools. We evaluate the portfolios based on a practical rolling-window scheme. More specifically, at the beginning of each month, one asset pool is formed by including the current constituents of DJIA 30 index and the Fama-French three factors. The portfolios are constructed using the monthly excess returns during the past T months, where T is the sample size to be specified. If a stock has missing data in the T-month training period, it is excluded from the asset pool. As a consequence, the number of stocks would vary over time and can be slightly smaller than the total number of constituents. The risk constraint is fixed to be the standard deviation of the DJIA 30 index returns during the first training period. The portfolios are held for one month, and the corresponding returns are recorded. We then evaluate the performance of the portfolios under comparison based on the out-of-sample monthly portfolio returns.

4.2.2 Performance Summary

We evaluate the performance of MAXSER and other competing portfolios in terms of risk, annualized Sharpe ratio[7] and statistical test of Sharpe ratio. We also investigate the effect of transaction costs, and demonstrate the comparisons based on portfolio returns net of transaction costs.

Without transaction costs

For the DJIA data set with around 30 stocks in each investment pool, we use a sample size $T = 60$, which means that each training set contains 5 years returns. The testing period is February 1967 – December 2016, which results in, for each strategy, 599 out-of-sample monthly returns. In addition to comparing out-of-sample risks and Sharpe ratios, to show the significance of the advantage of our portfolio MAXSER, we conduct hypothesis tests about the Sharpe ratio. More specifically, we test

$$H_0 : SR_{MAXSER} \leqslant SR_0 \quad vs \quad H_a : SR_{MAXSER} > SR_0, \tag{4.1}$$

where SR_{MAXSER} denotes the Sharpe ratio of MAXSER portfolio, and SR_0 denotes the Sharpe

[7]For computing Sharpe ratios, we obtain the risk-free rate r_f from Fama/French Data Library.

ratio of the portfolio under comparison. To conduct such a test, we adopt Memmel (2003)'s corrected version of Jobson and Korkie (1981)'s test. The summary without considering transaction costs is reported in Table 6, which shows the risk, Sharpe ratio, and the p-value of test (4.1) for each competing portfolio.

Table 6

Summary of risk, Sharpe ratio and p-value of Sharpe ratio test of the portfolios under comparison on DJIA 30 index constituents and Fama-French three factors. The testing period is February 1967 – December 2016. The risk constraint $\sigma = 0.0366$ *is the standard deviation of the excess returns on DJIA 30 index during February 1962 – January 1967.*

DJIA 30 & FF3	σ=0.0366	$T = 60$	
Portfolio	Risk	Sharpe Ratio	p-value
Index	0.044	0.102	0.000
Equally weighted	0.043	0.064	0.000
Factor	0.040	0.425	0.000
KZ	0.124	0.535	0.000
MAXSER	0.044	**0.701**	–
MV/GMV with estimated covariance matrix inputs			
MV-P	0.089	0.595	0.008
MV-LS	0.053	0.303	0.000
MV-NLS	0.053	0.457	0.000
NLSF	0.051	0.524	0.000
GMV-LS	0.017	0.397	0.000
GMV-NLS	0.017	0.337	0.000
MV with no-short-sale constraint			
MV-P-NSS	0.040	0.463	0.000
MV-LS-NSS	0.035	0.431	0.000
MV-NLS-NSS	0.035	0.399	0.000
MV with short-sale constraint & cross-validation			
MV-P-SSCV	0.042	0.543	0.000
MV-LS-SSCV	0.047	0.371	0.000
MV-NLS-SSCV	0.041	0.294	0.000
MV with ℓ_1-norm constraint & cross-validation			
MV-P-l1CV	0.038	0.357	0.000
MV-LS-l1CV	0.045	0.440	0.000
MV-NLS-l1CV	0.046	0.386	0.000

From Table 6, one can observe the following:

- In terms of *risk control*,
  - The risk of MAXSER is close to those of the index, the equally weighted, the Factor, the short-sale constrained and ℓ_1-norm constrained portfolios.
  - The risk of the plug-in ("MV-P") portfolio is more than twice of that of our MAXSER portfolio, which is far beyond the risk constraint level and hardly bearable for investors.
- In terms of *Sharpe ratio*,
  - Our MAXSER portfolio yields the highest Sharpe ratio.
  - Compared with portfolios with similar risks to ours, the Sharpe ratio of MAXSER is 29% higher than that of MV-P-SSCV portfolio, which performs better than other portfolios under comparison at the same risk level.
  - The NLSF portfolio, which also takes factor structure into account, yields a 16% higher risk and a 25% lower Sharpe ratio compared with our MAXSER portfolio.
  - The three ℓ_1-norm constrained MV portfolios (MV-P-l1CV, MV-LS-l1CV and MV-NLS-l1CV) also possess the sparsity property that only a small portion of stocks are invested in. However, they yield much lower Sharpe ratio than MAXSER. Such a comparison shows that the outstanding performance of MAXSER is largely due to its methodology rather than solely imposing ℓ_1-norm constraint.
  - The small p-values of Sharpe ratio tests against all the other portfolios demonstrate the statistical significance of the advantage of MAXSER.
  - The comparison between the Sharpe ratios of MAXSER (0.7) and Factor (0.4) portfolios indicates that investing on individual stocks in addition to factors using our strategy MAXSER can substantially improve the performance.

In summary, our MAXSER strategy effectively controls out-of-sample portfolio risk, and dominates the competing portfolios in terms of mean-variance efficiency.

With transaction costs

Next, we take the transaction cost into account and compute the returns of each portfolio net of transaction costs. Here we adopt a simple and widely-used formula of transaction costs, which is closely related to the portfolio turnover. The turnover is defined as

$$\text{Turnover}(t) := \sum_{j=1}^{N} |w_j(t+1) - w_j(t+)|, \tag{4.2}$$

where $w_j(t+1)$ is the weight on asset j at the beginning of period $t+1$, and $w_j(t+)$ is the weight of the same asset at the end of period t. The transaction cost of the portfolio at time t is proportional to Turnover(t) and a cost level c_0, which measures transaction cost per dollar traded. It can be derived that the portfolio return net of transaction cost in period t, $r_{net}(t)$, has the following relation with the total portfolio return $r(t)$:

$$r_{net}(t) = (1 - c_0 \text{Turnover}(t))\,(1 + r(t)) - 1. \tag{4.3}$$

In Engle et al. (2012), it is found that the average cost level for NYSE stocks is around 0.088%. In the following analysis we adopt $c_0 = 0.1\%$. Table 7 shows the risk and Sharpe ratio net of transaction costs.

Table 7

The summary of risk, Sharpe ratio and p-value of Sharpe ratio test based on returns net of transaction costs of the portfolios on DJIA 30 constituents and Fama-French three factors. The out-of-sample testing period is February 1967 – December 2016.

DJIA & FF3	$c_0 = 0.1\%$	σ=0.0366	
Portfolio	Risk	Sharpe Ratio	p-value
Equally weighted	0.043	0.059	0.000
Factor	0.040	0.403	0.000
KZ	0.123	0.440	0.000
MAXSER	0.044	**0.630**	–
MV/GMV with estimated covariance matrix inputs			
MV-P	0.089	0.516	0.004
MV-LS	0.053	0.251	0.000
MV-NLS	0.053	0.379	0.000
NLSF	0.051	0.466	0.000
GMV-LS	0.017	0.368	0.000
GMV-NLS	0.017	0.266	0.000
MV with no-short-sale constraint			
MV-P-NSS	0.040	0.437	0.000
MV-LS-NSS	0.035	0.409	0.000
MV-NLS-NSS	0.035	0.373	0.000
MV with short-sale constraint & cross-validation			
MV-P-SSCV	0.042	0.449	0.000
MV-LS-SSCV	0.047	0.282	0.000
MV-NLS-SSCV	0.041	0.150	0.000
MV with ℓ_1-norm constraint & cross-validation			
MV-P-l1CV	0.038	0.269	0.000
MV-LS-l1CV	0.045	0.339	0.000
MV-NLS-l1CV	0.045	0.273	0.000

Table 7 again shows the clear advantage of our MAXSER portfolio. With transaction costs deducted, all methods have a lower Sharpe ratio, and MAXSER still yields a Sharpe ratio that is significantly higher than the other strategies.

4.3 General Statistical Evaluation

The comparisons in Section 4.2.2 are from a practical viewpoint, where for each period the stock pools are updated to include all index constituents. In this section, we further evaluate the portfolio performances from a more statistical point of view, and base our investigation on 100 random stock pools formed by historical constituents of S&P 500 index. Specifically, each stock pool consists of 100 stocks randomly chosen form the stock universe, which contains 369 stocks that have been included into S&P 500 index from 1964 to 2016, and also have complete price data during the period of January 1992 – December 2016, the whole study period. The stock pool is then kept fixed throughout the study period. In such a way, we reduce the effect of inclusion/exclusion of stocks. We will make overall evaluations based on the 100 randomizations.

4.3.1 An Overall Comparison

The following results are again based on the same rolling-window scheme as described in Section 4.2.1. The sample size is $T = 120$, and we still include the Fama-French factors into our investment pools. The means and standard deviations of portfolio risks and Sharpe ratios are shown in Table 8.

Table 8

Summary of risks and Sharpe ratios of the portfolios under comparison for 100 random datasets, each containing 100 stocks randomly selected from S&P 500 index historical constituents. The testing period is January 2002 – December 2016. The risk constraint is taken to be the standard deviation of the index excess returns during January 1992 – December 2001, and is fixed over time and for all asset pools. Both average value and standard deviation (in brackets) of each performance measure are reported.

S&P 500 & FF3	$\sigma = 0.041$	$T = 120$
Portfolio	Risk	Sharpe Ratio
Index	0.042	0.223
Factor	0.041	0.320
Equally weighted	0.050 (0.002)	0.261 (0.041)
KZ	0.072 (0.019)	0.311 (0.234)
MAXSER	0.043 (0.003)	**0.532 (0.185)**
MV/GMV with estimated covariance matrix inputs		
MV-P	0.334 (0.031)	0.325 (0.230)
MV-LS	0.065 (0.004)	0.194 (0.180)
MV-NLS	0.061 (0.004)	0.188 (0.179)
NLSF	0.057 (0.003)	0.353 (0.161)
GMV-LS	0.025 (0.001)	0.420 (0.127)
GMV-NLS	0.025 (0.001)	0.414 (0.123)
MV with no-short-sale constraint		
MV-P-NSS	0.047 (0.002)	0.345 (0.120)
MV-LS-NSS	0.043 (0.002)	0.288 (0.138)
MV-NLS-NSS	0.042 (0.002)	0.285 (0.146)
MV with short-sale constraint & cross-validation		
MV-P-SSCV	0.047 (0.002)	0.345 (0.120)
MV-LS-SSCV	0.043 (0.002)	0.286 (0.139)
MV-NLS-SSCV	0.043 (0.002)	0.286 (0.153)

Table 8 shows that

- The plug-in portfolio in general carries a risk much higher than other portfolios. The KZ portfolio, on average carries a risk 67% higher than the average risk of our MAXSER portfolio, which is close to the risk constraint.

- Among the portfolios with reasonable risk levels, MAXSER achieves the highest Sharpe ratio on average, which is 27% higher than the average Sharpe ratio of GMV-LS portfolio, the second highest among the portfolios under comparison.

Next, we again take transaction costs into account. The returns net of transaction costs are computed by formula (4.3). The transaction cost level is again taken to be 0.1%. The comparisons are summarized in Table 9.

Table 9
Comparison of risks and Sharpe ratios based on portfolio returns net of transaction costs, for 100 random asset pools of size 103 formed by S&P 500 constituents and Fama-French three factors. Both average value and standard deviation (in brackets) of each performance measure net of transaction costs are reported. The length of training period is $T = 120$.

S&P 500 & FF3	$c_0 = 0.1\%$	$\sigma = 0.0247$
Portfolio	Net Risk	Net Sharpe Ratio
Factor	0.041	0.303
Equally weighted	0.050 (0.002)	0.256 (0.041)
Kan	0.072 (0.018)	0.140 (0.234)
MAXSER	0.043 (0.003)	**0.467 (0.185)**
MV/GMV with estimated covariance matrix inputs		
MV-P	0.328 (0.029)	0.101 (0.231)
MV-LS	0.065 (0.004)	0.139 (0.182)
MV-NLS	0.061 (0.004)	0.123 (0.181)
NLSF	0.057 (0.003)	0.300 (0.159)
GMV-LS	0.025 (0.001)	0.388 (0.126)
GMV-NLS	0.025 (0.001)	0.365 (0.122)
MV with no-short-sale constraint		
MV-P-NSS	0.047 (0.002)	0.326 (0.121)
MV-LS-NSS	0.043 (0.002)	0.272 (0.139)
MV-NLS-NSS	0.042 (0.002)	0.269 (0.147)
MV with short-sale constraint & cross-validation		
MV-P-SSCV	0.047 (0.002)	0.326 (0.121)
MV-LS-SSCV	0.043 (0.002)	0.269 (0.140)
MV-NLS-SSCV	0.043 (0.002)	0.265 (0.154)

Table 9 and its comparison with Table 8 reveal the following:

- Transaction costs cause more harm to the portfolios with relatively high risks such as the plug-in and KZ portfolios, whose Sharpe ratios drop from 0.311/0.325 to 0.140/0.101, after the deduction of transaction costs.
- MAXSER portfolio maintains its advantage over other portfolios. Its Sharpe ratio is more than 20% higher than that of GMV-LS, the second best portfolio in terms of Sharpe ratio among the portfolios under comparison.

In summary, both without or with taking transaction costs into account, our MAXSER portfolio generally outperforms other portfolios. Moreover, as we will see in the next section, the comparisons that we observe in terms of Sharpe ratio are statistically significant.

4.3.2 Statistical Tests About Sharpe Ratio

In order to show the statistical significance of the advantage of our MAXSER portfolio in terms of Sharpe ratio, for the 100 random asset pools formed by S&P 500 stocks and Fama-French three factors, we also conduct the Sharpe ratio tests (4.1) based on raw returns and returns net of transaction costs. The histograms of the p-values based on the 100 random investment pools are given in Figures 4$\sim$5.

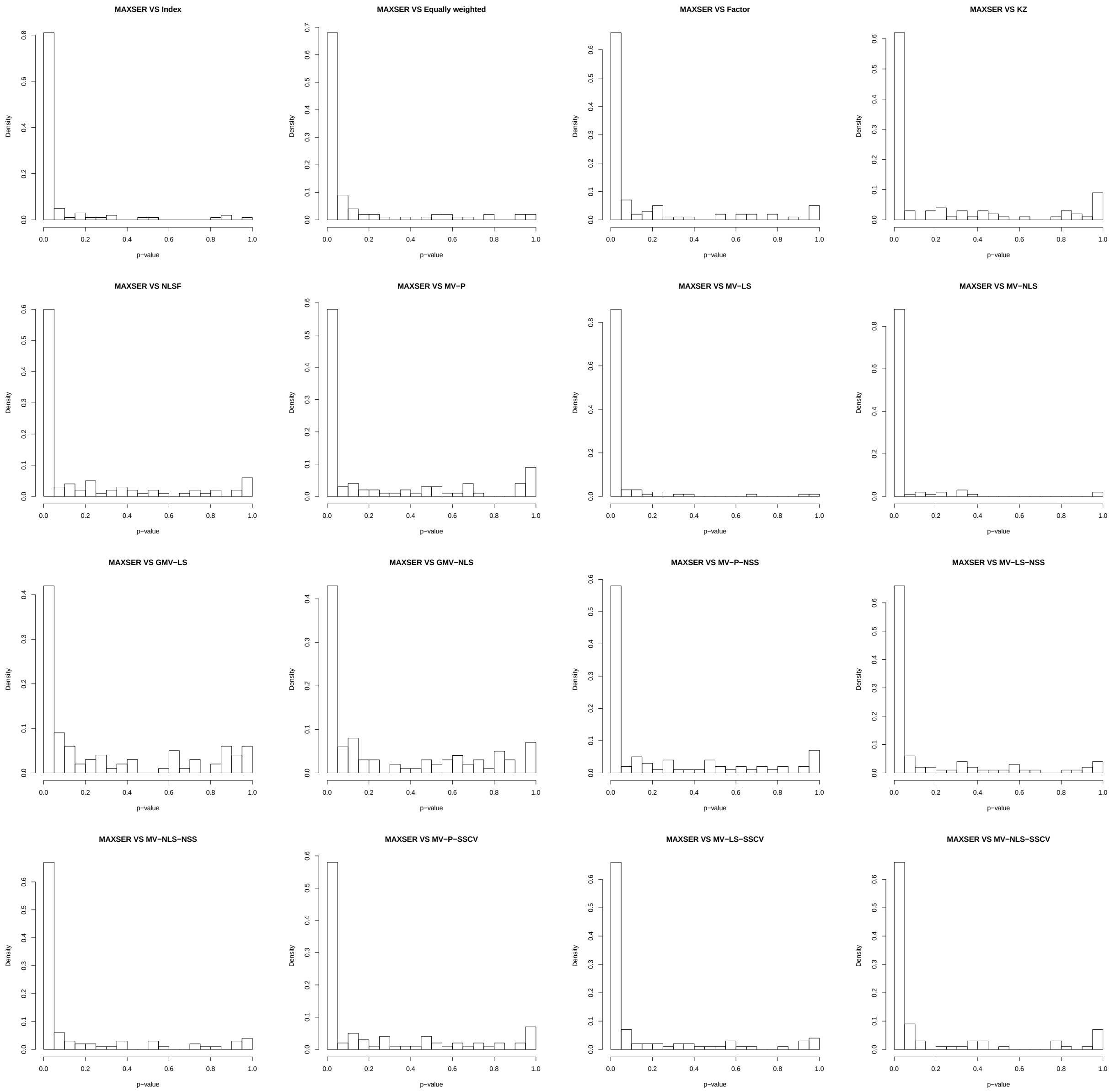

Figure 4. *Histograms of p-values for the Sharpe ratio test* (4.1) *against the portfolios under comparison, based on 100 random investment pools. In this figure transaction costs are not taken into account.*

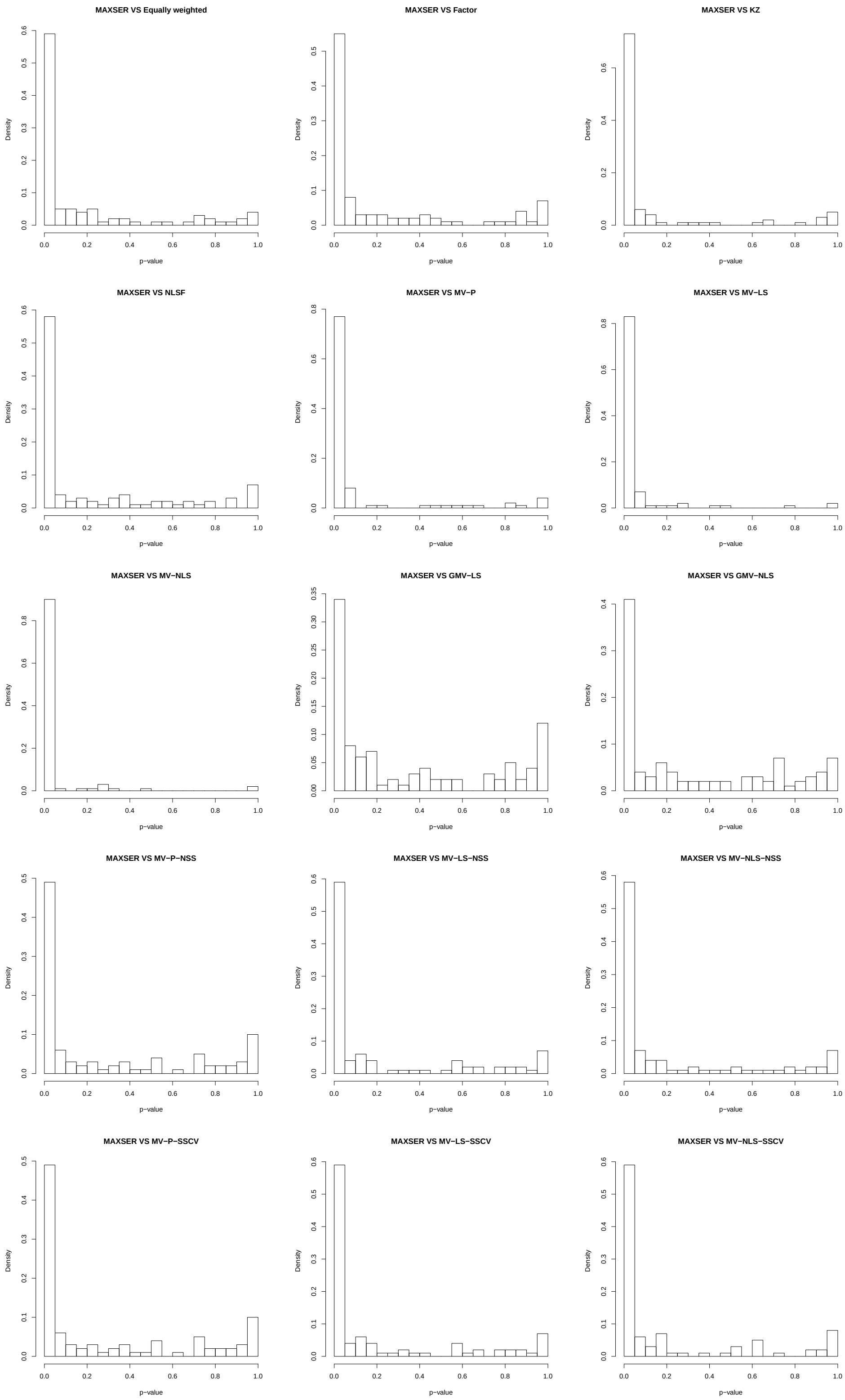

Figure 5. *Histograms of p-values for the Sharpe ratio test* (4.1) *against the portfolios under comparison, based on 100 random investment pools. In this figure transaction costs are deducted from portfolio returns.*

Note that if the null hypothesis H_0 in (4.1) holds, then the p-values would be roughly uniformly distributed. This is obviously not the case here. Furthermore, we observe from Figures 4 and 5 that in all comparisons, the p-values are mostly small. The histograms indicate that, among all portfolios under comparison, MAXSER does in general yield the highest Sharpe ratio.

5 CONCLUSION

In this paper, we propose a novel approach to construct the mean-variance efficient portfolio when the number of assets in the investment pool is not small compared with sample size. We prove that, under a mild sparsity assumption on the optimal portfolio, the MAXSER portfolio asymptotically achieves the mean-variance efficiency and meanwhile controls the risk effectively. To the best of our knowledge, this is the first time that these two objectives can be simultaneously achieved for large portfolios.

In addition to the sound statistical properties, the MAXSER portfolio possesses an attractive feature for practical implementation. Being a sparse portfolio strategy, MAXSER performs favorably by investing in only a small portion of assets. Moreover, our strategy involves relatively low transaction costs as demonstrated in the empirical studies.

References

Bai, J. and Ng, S. "Determining the number of factors in approximate factor models." *Econometrica*, 70(1):191–221 (2002).

Bai, Z., Li, H., and Wong, W.-K. "The best estimation for high-dimensional Markowitz mean-variance optimization." (2013).

Bai, Z., Liu, H., and Wong, W.-K. "Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory." *Mathematical Finance*, 19(4):639–667 (2009).
URL http://dx.doi.org/10.1111/j.1467-9965.2009.00383.x

Basak, G. K., Jagannathan, R., and Ma, T. "Jackknife estimator for tracking error variance of optimal portfolios." *Management Science*, 55(6):990–1002 (2009).

Best, M. J. and Grauer, R. R. "On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results." *Review of Financial Studies*, 4(2):315–342 (1991).

Black, F. and Litterman, R. B. "Asset allocation: combining investor views with market equilibrium." *The Journal of Fixed Income*, 1(2):7–18 (1991).

Britten-Jones, M. "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights." *The Journal of Finance*, 54(2):655–671 (1999).

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

Chamberlain, G. and Rothschild, M. "Arbitrage, factor structure, and mean-variance analysis on large asset markets." *Econometrica*, 51(5):1281–1304 (1983).
URL `http://dx.doi.org/10.2307/1912275`

Chopra, V. K. and Ziemba, W. T. "The effect of errors in means, variances, and covariances on optimal portfolio choice." *The Journal of Portfolio Management*, 19(2):6–11 (1993).

Connor, G., Goldberg, L. R., and Korajczyk, R. A. *Portfolio risk analysis*. Princeton University Press (2010).

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

El Karoui, N. "High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation." *The Annals of Statistics*, 38(6):3487–3566 (2010).
URL `http://dx.doi.org/10.1214/10-AOS795`

Engle, R., Ferstenberg, R., and Russell, J. "Measuring and Modeling Execution Cost and Risk." *The Journal of Portfolio Management*, 38(2):14–28 (2012).

Fan, J., Li, Y., and Yu, K. "Vast volatility matrix estimation using high-frequency data for portfolio selection." *Journal of the American Statistical Association*, 107(497):412–428

(2012a).
URL http://dx.doi.org/10.1080/01621459.2012.656041

Fan, J., Zhang, J., and Yu, K. "Vast portfolio selection with gross-exposure constraints." *Journal of the American Statistical Association*, 107(498):592–606 (2012b).
URL http://dx.doi.org/10.1080/01621459.2012.682825

Fastrich, B., Paterlini, S., and Winker, P. "Constructing Optimal Sparse Portfolios Using Regularization Methods." *Available at SSRN 2169062* (2012).

Feng, G., Giglio, S., and Xiu, D. "Taming the Factor Zoo." (2017).

Garlappi, L., Uppal, R., and Wang, T. "Portfolio selection with parameter and model uncertainty: A multi-prior approach." *Review of Financial Studies*, 20(1):41–81 (2007).

Jagannathan, R. and Ma, T. "Risk reduction in large portfolios: Why imposing the wrong constraints helps." *The Journal of Finance*, 58(4):1651–1684 (2003).

Jegadeesh, N. and Titman, S. "Profitability of momentum strategies: An evaluation of alternative explanations." *The Journal of Finance*, 56(2):699–720 (2001).

Jobson, J. D. and Korkie, B. M. "Performance hypothesis testing with the Sharpe and Treynor measures." *The Journal of Finance*, 36(4):889–908 (1981).

Kan, R. and Zhou, G. "Optimal portfolio choice with parameter uncertainty." *Journal of Financial and Quantitative Analysis*, 42(3):621 (2007).

Korajczyk, R. A. and Sadka, R. "Pricing the commonality across alternative measures of liquidity." *Journal of Financial Economics*, 87(1):45–72 (2008).

Lai, T. L., Xing, H., and Chen, Z. "Mean–variance portfolio optimization when means and covariances are unknown." *The Annals of Applied Statistics*, 5(2A):798–823 (2011).
URL http://dx.doi.org/10.1214/10-AOAS422

Ledoit, O. and Wolf, M. "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection." *Journal of Empirical Finance*, 10(5):603–621 (2003).

—. "A well-conditioned estimator for large-dimensional covariance matrices." *Journal of Multivariate Analysis*, 88(2):365–411 (2004).
URL http://dx.doi.org/10.1016/S0047-259X(03)00096-4

—. “Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks.” *The Review of Financial Studies* (2017).

Markowitz, H. “Portfolio selection.” *The Journal of Finance*, 7(1):77–91 (1952).

Memmel, C. “Performance hypothesis testing with the Sharpe ratio.” *Finance Letters*, 1(1) (2003).

Michaud, R. O. “The Markowitz optimization enigma: is “optimized” optimal?” *Financial Analysts Journal*, 31–42 (1989).

Rothman, A. J. “Positive definite estimators of large covariance matrices.” *Biometrika*, 99(3):733–740 (2012).

Tibshirani, R. “Regression shrinkage and selection via the lasso.” *Journal of the Royal Statistical Society. Series B. Methodological*, 58(1):267–288 (1996).

Tu, J. and Zhou, G. “Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies.” *Journal of Financial Economics*, 99(1):204–215 (2011).