Approaching Mean-Variance Efficiency for Large Portfolios

Mengmeng Ao
WISE and SOE, Xiamen University

Yingying Li
Hong Kong University of Science and Technology

Xinghua Zheng
Hong Kong University of Science and Technology

This paper introduces a new approach to constructing optimal mean-variance portfolios. The approach relies on 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, controls for risk and attains the maximum expected return as both the numbers of assets and observations grow. The superior properties of our approach are demonstrated through comprehensive simulation and empirical analysis. Notably, using our strategy, we find that investing in individual stocks, in addition to the Fama-French three-factor portfolios, leads to substantially improved performance. (*JEL* G11, C10)

Received October 6, 2014; editorial decision July 13, 2018 by Editor Andrew Karolyi. Authors have furnished an Internet Appendix, which is available on the Oxford University Press Web site next to the link to the final published paper online.

The groundbreaking mean-variance portfolio theory proposed by Markowitz (1952) continues to play a significant role in research and practice. The mean-variance efficient portfolio has a simple explicit expression that depends only on the expected mean and covariance matrix of asset returns. Because these two parameters are not observed in practice, sample mean and sample covariance matrix are used as proxies, and the resultant "plug-in" portfolio has been widely

We are very grateful to the editors, Andrew Karolyi and Leonid Kogan, and several anonymous referees for their very valuable comments and constructive suggestions, which led to great improvements of this paper. We also appreciate thoughtful comments from Yacine Ait-Sahalia, Torben Andersen, Federico Bandi, Utpal Bhattacharya, Tim Bollerslev, Robert Engle, Jianqing Fan, Eric Ghysels, Campbell Harvey, Joel Hasbrouck, Ravi Jagannathan, Raymond Kan, Steven Lalley, Jia Li, Per Mykland, Deniela Osterrieder, Andrew Patton, Matthew Richardson, Jeffery Russell, George Tauchen, Allan Timmermann, Viktor Tordorov, Ruey Tsay, and Dacheng Xiu and participants of various conferences and seminars. Internet appendix containing supplementary results and proofs can be found on *The Review of Financial Studies* Web site. Send correspondence to Yingying Li, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong SAR, China; telephone: +852 23587744. E-mail: yyli@ust.hk.

doi:10.1093/rfs/hhy105 Advance Access publication September 22, 2018

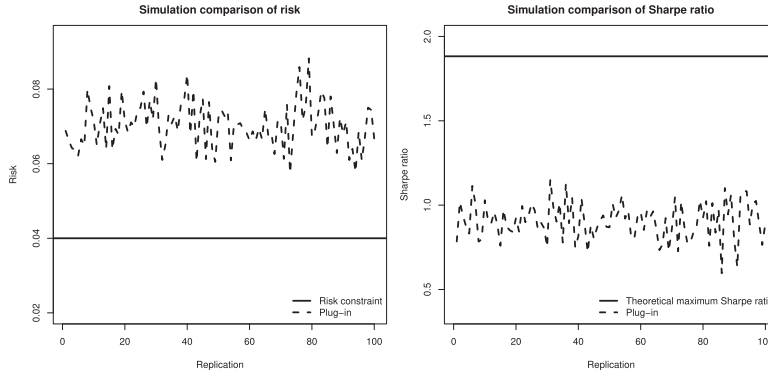

Figure 1
Performance of the plug-in portfolio
This figure compares risks and Sharpe ratios of the plug-in portfolio versus the risk constraint and the theoretical maximum Sharpe ratio, respectively. The portfolios are constructed based on data generated from an i.i.d. multivariate normal distribution with parameters calibrated from real data (see Section 2.2 for details). The left panel plots the portfolio risks, and the right panel plots the Sharpe ratios. The asset pool comprises 100 stocks and three factors, and the number of observations totals 240. The comparison was replicated 100 times.

adopted. Such an approach is justified by classical statistics theory, because the plug-in portfolio is a maximum-likelihood estimator of the optimal portfolio. However, as documented by 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 (see, e.g., Best and Grauer 1991; Green and Hollifield 1992; Chopra and Ziemba 1993; Britten-Jones 1999; Kan and Zhou 2007; Basak et al. 2009; DeMiguel et al. 2009b).

Modern portfolios often involve a large number of assets. This makes the mean-variance problem high-dimensional in nature and poses serious challenges. Consider the plug-in portfolio, for example. As we will see in Figure 1, the risk of the plug-in portfolio can be substantially higher than the prespecified risk level even when the portfolio weights are computed based on simulated independent and identically distributed (i.i.d.) returns. Moreover, this high risk is not adequately compensated by a high return, resulting in a significantly suboptimal Sharpe ratio. The key message is that, even in the ideal situation, when all the assumptions of the mean-variance optimization are satisfied (i.e., normally distributed returns and no time-varying parameters or regime switching), estimating the mean-variance efficient portfolio is still intrinsically challenging.

Figure 1 demonstrates how poor the out-of-sample performance of the plug-in portfolio can be. The asset pool consists of 100 stocks and three factors. The underlying mean and covariance matrix are calibrated from real data (see Section 2.2 for details). We construct the plug-in portfolio based on 20 years of monthly returns generated from an i.i.d. multivariate normal distribution. The simulated returns satisfy all the assumptions of the mean-variance framework.

The out-of-sample risk and Sharpe ratio of the plug-in portfolio are compared with the risk constraint and theoretical maximum Sharpe ratio, respectively. The left panel of Figure 1 illustrates that in all the 100 replications, the plug-in portfolio carries a risk nearly twice the specified level. Meanwhile, as shown in the right panel of Figure 1, the Sharpe ratio of the plug-in portfolio is only approximately 50% of the theoretical maximum Sharpe ratio.

The aforementioned phenomenon has been noted by Kan and Zhou (2007) and later investigated by Bai et al. (2009), El Karoui (2010), and Chen and Yuan (2016). These papers document that the deviation of the plug-in portfolio from the optimal portfolio is systematic, and that the bias is due to the dimension (number of assets) not being small compared with the sample size. More precisely, under the normality assumption on the returns and if the ratio between the number of assets, N, and the sample size, T, satisfies $N/T \to \rho \in (0,1)$, the Sharpe ratio of the plug-in portfolio, SR(plug-in), is asymptotically related to the theoretical maximum Sharpe ratio, SR^*, as follows:

$$\frac{SR(\text{plug-in})}{SR^*} \xrightarrow{P} \sqrt{\frac{1-\rho}{1+\rho/(SR^*)^2}} < \sqrt{1-\rho} < 1, \tag{1}$$

where "$\xrightarrow{P}$" stands for convergence in probability. Internet Appendix provides the proof of (1). For the global minimum variance (GMV) portfolio, Basak et al. (2009) derive a result in a similar spirit, stating that the plug-in GMV portfolio has, on average, a risk that is a larger-than-1 multiple of the true minimum risk, with the multiplier explicitly depending on the number of assets and sample size.

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, researchers seek to improve portfolio performance by plugging in better estimates of the underlying mean and covariance matrix. A widely used alternative estimator for estimating the covariance matrix is the linear shrinkage estimator proposed in Ledoit and Wolf (2003, 2004), who estimate the covariance matrix using a suitable linear combination of sample covariance matrix and a target 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. Regarding estimation of the mean, among other papers, Black and Litterman (1991) propose a quasi-Bayesian approach that combines investors' views with expected returns calculated using the capital asset pricing model. This quasi-Bayesian approach is extended to a fully Bayesian approach by Lai et al. (2011), who consider the mean-variance portfolio problem from a different perspective and aim to maximize a certain utility function. Garlappi et al. (2007) propose to adjust estimates of expected returns using a multiprior approach, also with an aim of maximizing a utility function.

Another direction to improve portfolio performance is to modify the original optimization by imposing constraints on the portfolio weights. Most research in

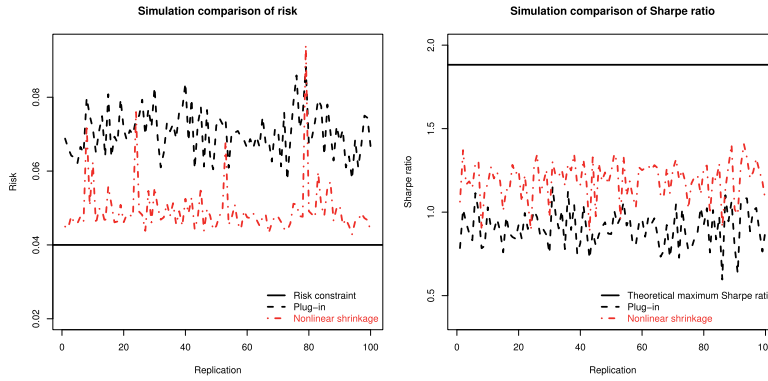

Figure 2
Performance of the nonlinear shrinkage portfolio
This figure compares the plug-in and nonlinear shrinkage portfolios. The portfolios are constructed based on the same data used for Figure 1. The left and right panels plot the portfolio risks and Sharpe ratios, respectively. The asset pool comprises 100 stocks and three factors, and the sample size totals 240. The comparison was replicated 100 times.

this direction focuses on the GMV portfolio. Imposing constraints on weights has been empirically shown to be helpful (see, e.g., Jagannathan and Ma 2003; DeMiguel et al. 2009a; Brodie et al. 2009; Fastrich et al. 2015). Fan et al. (2012b) provide a theoretical justification for the empirical results obtained by Jagannathan and Ma (2003) and also investigate the GMV portfolio with gross-exposure constraints in which some short positions are allowed. Fan et al. (2012a) consider GMV portfolio estimation using high-frequency data under gross-exposure constraints. In addition to these approaches, combinations of different portfolios have been investigated (Kan and Zhou 2007 and Tu and Zhou 2011).

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

In this paper, we propose a new methodology for estimating the mean-variance efficient portfolio, which we call the maximum-Sharpe-ratio estimated and sparse regression (MAXSER) method. MAXSER is a general approach that can be applied to various situations in which the number of assets is not small compared with the sample size. 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 methodology that can achieve these two objectives simultaneously for large portfolio optimization.

Specifically, we make four contributions. Our first contribution is to establish an equivalent unconstrained regression representation of the mean-variance portfolio problem. This special regression representation is novel. Two closely related approaches are those of Britten-Jones (1999), who uses 1 as the response, and Brodie et al. (2009), who use the maximum expected return as the response. The regression in Britten-Jones (1999) requires a challenging scaling to obtain the mean-variance portfolio. Brodie et al. (2009) employ a constrained regression, which induces errors and biases resulting from replacing the constraint with the analogous sample version. By contrast, our regression representation is, on the one hand, *equivalent* to the original optimization problem, and, on the other hand, *unconstrained*; thus, it can be conveniently combined with high-dimensional regression techniques.

Our second contribution is the estimation of the optimal portfolio when the asset pool comprises only individual assets. Our approach is built on (1) consistent estimation of the response in our regression representation, which directly links to consistent estimation of the maximum Sharpe ratio achieved by the tangency portfolio, and (2) rigorous analysis of a sparse regression, which possesses unique features. We theoretically prove the convergence of the expected return and risk of our estimated portfolio toward the theoretical maximum expected return and risk constraint, respectively.

Our third contribution is the estimation of the optimal portfolio when factors are also included in the asset pool. We establish a framework that decomposes the optimal portfolio on all assets into the optimal portfolio on factors and that on individual assets, on the basis of which we develop our estimator of the optimal full portfolio. Asymptotic consistencies of the expected return and risk of our portfolio are also established in this case.

Our fourth contribution lies in simulation and empirical evidence. We compare our strategy with various other methods including the plug-in portfolio, the equally weighted portfolio (DeMiguel et al. 2009b), the linear and nonlinear shrinkage portfolios of Ledoit and Wolf (2004) and Ledoit and Wolf (2017), and several other variations of mean-variance and GMV portfolios with constraints on portfolio weights. Figure 3 compares the plug-in, nonlinear shrinkage, and the MAXSER portfolios. The added dashed lines represent our MAXSER portfolio. We see that the MAXSER portfolio effectively controls risk. More importantly, the Sharpe ratio comparisons show that our MAXSER portfolio nearly achieves mean-variance efficiency and significantly outperforms the others.

Empirically, both with and without consideration of transaction costs, we find that MAXSER effectively controls the risk and yields high Sharpe ratios. Notably, comparing the Sharpe ratios of the MAXSER portfolio and the mean-variance portfolio on the Fama-French three factors indicates that additionally investing in stocks using our MAXSER strategy yields significant gains. Moreover, the decisive advantage of the MAXSER portfolio over various constrained portfolios confirms that the advantage of MAXSER is

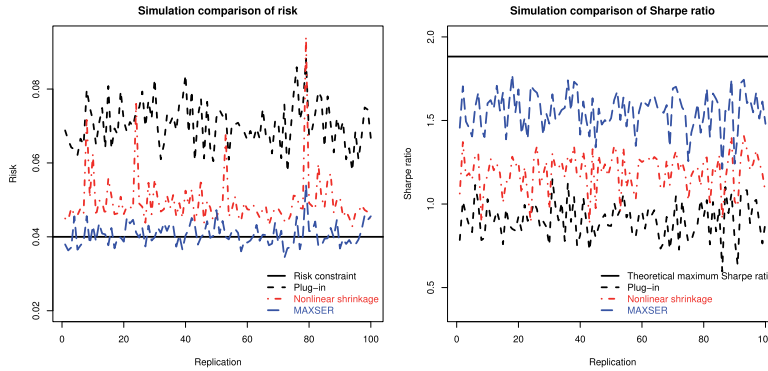

Figure 3
Performance of the MAXSER portfolio
This figure compares our MAXSER portfolio with the plug-in and nonlinear shrinkage portfolios, based on the same simulated data used in Figures 1 and 2. The added dashed lines represent the MAXSER portfolio. The comparison was replicated 100 times.

fundamentally due to its methodology, rather than solely the imposition of constraints.

1. The MAXSER Methodology

1.1 An unconstrained regression representation

Suppose that we have a pool of N risky assets. Denote their (random excess) returns by $\boldsymbol{r}=(r_1,r_2,\ldots,r_N)'$, where for any vector $\boldsymbol{v}$, $\boldsymbol{v}'$ is 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 assets. For a given risk constraint σ, the mean-variance portfolio problem is

$$\arg\max_{\boldsymbol{w}} E(\boldsymbol{w}'\boldsymbol{r})=\boldsymbol{w}'\boldsymbol{\mu} \quad \text{subject to} \quad \mathrm{Var}(\boldsymbol{w}'\boldsymbol{r})=\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}\le\sigma^2. \tag{1.1}$$

If we denote $\theta=\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ as the square of the maximum Sharpe ratio of the optimal portfolio, then the optimization problem (1.1) can be represented in its dual form with a return constraint:

$$\arg\min_{\boldsymbol{w}} \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{w}'\boldsymbol{\mu}\ge r^*:=\sigma\sqrt{\theta}. \tag{1.2}$$

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

$$\boldsymbol{w}^*=\frac{\sigma}{\sqrt{\theta}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}. \tag{1.3}$$

Because of the difficulty of estimating the mean and covariance matrix in high-dimensional situations, instead of working with the formula (1.3) and plugging in estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, we propose a novel *equivalent* and *unconstrained* regression representation of the mean-variance optimization.

Proposition 1. The unconstrained regression

$$\underset{\boldsymbol{w}}{\operatorname{argmin}} E(r_c - \boldsymbol{w}'\boldsymbol{r})^2, \quad \text{where} \quad r_c := \frac{1+\theta}{\theta} r^* \equiv \sigma \frac{1+\theta}{\sqrt{\theta}}, \tag{1.4}$$

is equivalent to the mean-variance optimizations (1.1) and (1.2).

A couple of comments are in order. First, Brodie et al. (2009) use a constrained regression representation of the mean-variance optimization (1.2):

$$\underset{\boldsymbol{w}}{\operatorname{argmin}} E(r^* - \boldsymbol{w}'\boldsymbol{r})^2 \quad \text{subject to} \quad \boldsymbol{w}'\boldsymbol{\mu} = r^*.$$

Such a constrained regression is theoretically difficult to analyze and complicated to solve numerically. To implement the regression, Brodie et al. (2009) replace the constraint $\boldsymbol{w}'\boldsymbol{\mu} = r^*$ with its sample counterpart $\boldsymbol{w}'\overline{\boldsymbol{r}} = r^*$, where $\overline{\boldsymbol{r}}$ is the sample mean vector. However, in the high-dimensional case, $\overline{\boldsymbol{r}}$ is *not* a consistent estimator of $\boldsymbol{\mu}$ (in the sense that $||\overline{\boldsymbol{r}} - \boldsymbol{\mu}||_2 \nrightarrow 0$, where $||\cdot||_2$ denotes the ℓ_2-norm). Second, Britten-Jones (1999) connects the estimation of the tangency portfolio with least squares regression with response 1. The regression yields a multiple of the plug-in tangency portfolio. Scaling the weights so that the weights add to 1 recovers the plug-in tangency portfolio. Britten-Jones (1999) does not consider the mean-variance portfolio problem (1.1) or (1.2). To find the mean-variance portfolio for a given risk constraint σ, recall that the plug-in tangency portfolio will be suboptimal in high dimensions (see Equation (1) for the precise statement). Additionally, even if one knew the true tangency portfolio, further scaling is required because the portfolio has a risk of $\sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/(\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^2}$, and such scaling is challenging because the estimations of $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ and $\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ both involve high-dimensional parameters.

We emphasize that our regression representation (1.4) is intrinsically different from existing regression representations for the mean-variance portfolio estimation. Our representation is *unconstrained* and is *equivalent* to the mean-variance portfolio problem. The elimination of the constraint is particularly helpful for large portfolio selection, in which crucial techniques, such as sparse regression, become directly applicable.

1.2 Sparse regression

Proposition 1 translates the original mean-variance portfolio problems (1.1) and (1.2) into an equivalent unconstrained regression problem. The next step is to utilize this regression and estimate the optimal portfolio by replacing $\underset{\boldsymbol{w}}{\operatorname{argmin}} E(r_c - \boldsymbol{w}'\boldsymbol{r})^2$ with its sample version

$$\underset{\boldsymbol{w}}{\operatorname{argmin}} \frac{1}{T}\sum_{t=1}^{T}\left(r_c - \boldsymbol{w}'\boldsymbol{R}_t\right)^2, \tag{1.5}$$

where $\boldsymbol{R}_t = (R_{t1}, \ldots, R_{tN})'$, $t = 1, \ldots, T$, are T i.i.d. copies of the return vector $\boldsymbol{r}$. When the number of assets N is not small compared with T, equation

(1.5) is a high-dimensional regression problem. It is now well known that consistently estimating the coefficients in high-dimensional regression is in general impossible, and some type of sparsity is required to make it possible. The most widely used sparsity assumption is boundedness on the ℓ_1-norm of the regression coefficients. In our case, this corresponds to the assumption that $||\boldsymbol{w}^*||_1$ is bounded, where $||\boldsymbol{v}||_1 = \sum_{i=1}^{N} |v_i|$ for any $\boldsymbol{v} = (v_1, \ldots, v_N)'$. We adopt the sparse regression technique least absolute shrinkage and selection operator (LASSO; Tibshirani 1996), which leads to the following estimation of the optimal portfolio $\boldsymbol{w}^*$:

$$\boldsymbol{w}(r_c) := \arg\min_{\boldsymbol{w}} \frac{1}{T} \sum_{t=1}^{T} \left(r_c - \boldsymbol{w}' \boldsymbol{R}_t\right)^2 \quad \text{subject to} \quad ||\boldsymbol{w}||_1 \leq \lambda, \tag{1.6}$$

where λ is an ℓ_1-norm constraint tuning parameter. Note that the solution (1.6) is *infeasible* because the response r_c is unknown. In the next subsection, we will develop a feasible solution with a consistent estimator of r_c. For clarity, let us suppose for the moment that r_c is known.

The solution (1.6) involves a tuning parameter λ. It can be easily seen that if λ is chosen to be larger than the ℓ_1-norm of the ordinary least squares (OLS) solution (1.5), then one always gets the OLS solution. On the other hand, as λ varies from 0 to the ℓ_1-norm of the OLS solution, one gets a so-called "solution path," with each solution corresponding to an estimated portfolio. An algorithm named the "least angle regression" (LARS) (Efron et al. 2004) was developed to efficiently solve for the entire solution path. We employ this algorithm in the implementation of our method.

Figure 4 shows the risks and Sharpe ratios of all the estimated portfolios on two solution paths based on simulated i.i.d. returns following a multivariate normal distribution. The mean and covariance matrix are calibrated from idiosyncratic returns[1] obtained by fitting the Fama-French three-factor model to 100 randomly chosen constituents of the S&P 500 index (see Section 2.2 for details).

In Figure 4, we present the results based on two LASSO solution paths, one with response r_c (setting $\sigma = 1$) and the other with response 1. The x-axis represents the ratio of the ℓ_1-norm of solution $\boldsymbol{w}$ on the solution path relative to the ℓ_1-norm of the OLS solution:

$$\zeta = \frac{||\boldsymbol{w}||_1}{||\boldsymbol{w}_{ols}||_1}. \tag{1.7}$$

For the entire range of ζ from 0 to 1, the left panel shows the risks of all estimated portfolios on the solution paths based on the two response values,

[1] Here, to be consistent with our simulation and empirical analysis, we use simulated idiosyncratic returns to demonstrate the LASSO solution paths. See Sections 1.4.3 and 1.5.4 for more details.

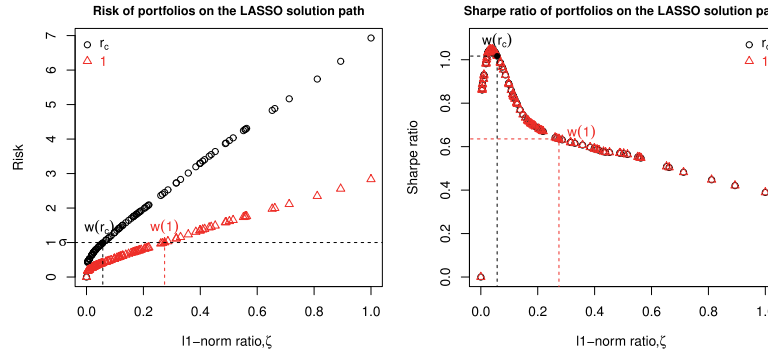

Figure 4
Risks and Sharpe ratios of portfolios on LASSO solution paths
This figure shows simulation comparisons of the risks and Sharpe ratios of all estimated portfolios on two LASSO solution paths: one with response r_c (setting $\sigma = 1$) and the other with response 1. The x-axis stands for the ratio between the ℓ_1-norms of a LASSO solution and the OLS solution. The risk and Sharpe ratio are true values based on the underlying mean and covariance matrix.

whereas the right panel shows the corresponding Sharpe ratios. We observe the following:

(1) For each solution path, the risk increases with respect to the ℓ_1-norm ratio ζ. Consider the solution path with the correct response r_c, for example. When ζ is small, the risk is lower than the risk constraint level; as ζ increases, the risk increases and eventually exceeds the constraint level. This feature shows the importance of appropriately choosing the ℓ_1-norm ratio ζ, or equivalently, the tuning parameter λ.

(2) To identify the pursued portfolio from the solution path, we look for the portfolio with a risk closest to the risk constraint.[2] This leads to portfolios $\boldsymbol{w}(r_c)$ and $\boldsymbol{w}(1)$. The left panel of Figure 4 shows that they correspond to very different ℓ_1-norm ratios.

(3) Regarding the Sharpe ratio comparison, we see that the Sharpe ratio paths coincide. However, because the portfolios $\boldsymbol{w}(r_c)$ and $\boldsymbol{w}(1)$ correspond to different values of ζ, their Sharpe ratios differ considerably. The portfolio with the correct response r_c achieves almost the highest Sharpe ratio on the path, whereas the other portfolio with response 1 yields a much lower Sharpe ratio.

In summary, this illustration demonstrates the importance of (1) our corrected response r_c and (2) the choice of tuning parameter. We discuss the selection of the tuning parameter in more detail in Section 1.5.1.

[2] Alternatively, one may consider using the Sharpe ratio to identify the "optimal" ℓ_1-norm ratio ζ. Notice, however, that the estimation of the Sharpe ratio is less accurate than that of risk. Also, if the response is not r_c, even if the portfolio with the highest Sharpe ratio on the solution path could be identified, additional scaling is needed to obtain the mean-variance portfolio with risk σ, and this scaling is even more challenging than what is discussed after Proposition 1, because the risk of such a portfolio does not have an analytical formula.

1.3 Scenario 1: The asset pool comprises individual assets only

1.3.1 Estimating the maximum Sharpe ratio and the regression response.

In our regression representation, the response r_c is unknown[3] and needs to be estimated. The estimation of the response r_c is closely related to the estimation of the maximum Sharpe ratio, which is considered in Kan and Zhou (2007). It is shown that the square of the plug-in Sharpe ratio has a noncentralized F-distribution (see equation (49) of Kan and Zhou 2007). The result implies that the plug-in Sharpe ratio is heavily biased when the number of assets, N, is not negligible compared with the sample size T. Suppose that $\boldsymbol{R_t} = (R_{t1}, \ldots, R_{tN})', t = 1, \ldots, T$, are T i.i.d. copies of the (excess) return. Let $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ be the sample mean and sample covariance matrix, respectively. The following unbiased estimator is proposed in Kan and Zhou (2007):

$$\widehat{\theta} := \frac{(T-N-2)\widehat{\theta}_s - N}{T}, \tag{1.8}$$

where $\widehat{\theta}_s := \widehat{\boldsymbol{\mu}}' \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\mu}}$ is the sample estimate of $\theta = \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$. Assuming $N/T \to \rho \in (0, 1)$, we establish the following

Proposition 2. Suppose that the maximum Sharpe ratio is bounded. Under normality assumption on returns and assuming that $N/T \to \rho \in (0, 1)$, we have

$$|\widehat{\theta} - \theta| \xrightarrow{P} 0. \tag{1.9}$$

Consequently,

$$\widehat{r_c} := \sigma \frac{1 + \widehat{\theta}}{\sqrt{\widehat{\theta}}} \tag{1.10}$$

satisfies that

$$|\widehat{r_c} - r_c| \xrightarrow{P} 0. \tag{1.11}$$

We emphasize that our estimation of $\theta = \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$ is *not* obtained through consistently estimating $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, which would be *impossible* without imposing strong structural assumptions on them. The challenge is due to error accumulation. Consider the estimation of $\boldsymbol{\mu}$, for example. Each entry of $\boldsymbol{\mu}$ can be estimated with an error of order $1/\sqrt{T}$. However, because the dimension of $\boldsymbol{\mu}$ is N, the total estimation error under ℓ_2-norm is of order $\sqrt{N/T}$, which is $O(1)$ under our assumption that $N/T \to \rho \in (0, 1)$. By contrast, we estimate

[3] In the dual form of the mean-variance portfolio problem, in which the return constraint r^* is given, the response is still unknown because it also depends on θ, the square of the maximum Sharpe ratio.

θ directly. The estimator $\widehat{\theta}$ is consistent, and actually, as can be seen from the proof, the error $|\widehat{\theta}-\theta|$ is of order $1/\sqrt{T}$.

1.3.2 A LASSO-type estimator. With the consistent estimation of r_c in Proposition 2, following (1.6) we construct our feasible LASSO-type estimator $\widehat{\boldsymbol{w}^*}=\left(\widehat{w_1^*},\ldots,\widehat{w_N^*}\right)'$ as follows:

$$\widehat{\boldsymbol{w}^*}=\arg\min_{\boldsymbol{w}}\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{r_c}-\boldsymbol{w}'\boldsymbol{R_t}\right)^2 \quad \text{subject to} \quad ||\boldsymbol{w}||_1\le\lambda. \tag{1.12}$$

Before we give the theoretical properties of $\widehat{\boldsymbol{w}^*}$, we list the assumptions that will be needed.

Assumption:

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

A2 There exists $M<\infty$ such that $\max\left(\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu},\max\limits_{1\le j\le N}|\mu_j|\right)\le M$;

A3 There exists $L<\infty$ such that $\max\limits_{1\le i\le N}|\sigma_{ii}|\le L$, where $\boldsymbol{\Sigma}=:(\sigma_{ij})_{1\le i,j\le N}$;

A4 $||\boldsymbol{w}^*||_1\le\lambda$ for some constant λ;

A5 The number of assets N and the sample size T satisfy that $\rho_T:=N/T\to\rho\in(0,1)$.

We have the following comments about the assumptions above. About Assumption A1, we focus on the most fundamental form of the mean-variance portfolio problem, so we assume normal distribution of returns. Numerically, we find that our proposed method works well even when heavy tails are present.

Assumption A2 is equivalent to the common belief that the maximum Sharpe ratio and asset expected returns are bounded. We emphasize that we do not require the eigenvalues of $\boldsymbol{\Sigma}$ to be bounded, and consequently factor structure in returns is allowed. Nor do we require the eigenvalues of $\boldsymbol{\Sigma}$ to be bounded from below.

Assumption A4, the boundedness of $||\boldsymbol{w}^*||_1$, is our sparsity requirement on the optimal portfolio. Note that Assumption A4 does not require most weights to be zero. For example, it does not rule out value-weighted portfolios. In addition, we will see in Section 1.4 that when factor investing is allowed, the sparsity requirement will be significantly relaxed and only imposed on the optimal portfolio on idiosyncratic components.

Assumption A5 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 only because we need to take inverse of the sample covariance matrix in estimating the maximum Sharpe ratio; see Proposition 2.

1.3.3 Main result 1: MAXSER without factor investing. We now state our first main result, which establishes the asymptotic optimality of our MAXSER portfolio when the asset pool comprises only individual assets.

Theorem 1. Under Assumptions A1–A5, the MAXSER portfolio $\widehat{\boldsymbol{w}^*}$ defined in (1.12) with $\widehat{r_c}$ given by (1.10) satisfies that, as $N \to \infty$,

$$|\boldsymbol{\mu}'\widehat{\boldsymbol{w}^*} - r^*| \xrightarrow{P} 0, \tag{1.13}$$

and

$$\left|\sqrt{\widehat{\boldsymbol{w}^*}'\boldsymbol{\Sigma}\widehat{\boldsymbol{w}^*}} - \sigma\right| \xrightarrow{P} 0. \tag{1.14}$$

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 $\widehat{\boldsymbol{w}^*}$ approaches the mean-variance efficiency.

1.4 Scenario 2: When factor investing is allowed

1.4.1 The optimal portfolio: A factor-idiosyncratic component separation. In addition to individual assets, factors are often included in an investment asset pool. Motivated by this consideration, we now illustrate the implementation of MAXSER when factor investing is allowed. Consider the following model:

$$r_i = \alpha_i + \sum_{j=1}^{K} \beta_{ij} f_j + e_i =: \sum_{j=1}^{K} \beta_{ij} f_j + u_i, \qquad i = 1, \ldots, N, \tag{1.15}$$

where f_j's are factor (excess) returns, β_{ij}'s are individual stock sensitivities to the factors, and e_i's are the remaining errors in the model that are independent of the factor returns (f_j). Unlike the approximate factor model, in which the "idiosyncratic returns" ($u_i = \alpha_i + e_i$) are assumed to have no factor structure, here, in (1.15), we allow (u_i) to still admit factor structures, and, subsequently, we refer to the (u_i) as the idiosyncratic returns. The factors can be any well-recognized factors, such as Fama-French three factors (FF3), or other factors identified in the large literature on asset pricing (see, e.g., Jegadeesh and Titman 2001 and Korajczyk and Sadka 2008). They also can be statistical factors (identified on the basis of a separate set of historical returns of a larger pool of assets).

Model (1.15) can be written in a compact form as

$$\boldsymbol{r} = \boldsymbol{\beta}\boldsymbol{f} + \boldsymbol{u}, \tag{1.16}$$

where $\boldsymbol{\beta} = (\beta_{ij})_{N\times K}$, $\boldsymbol{f} = (f_1, \ldots, f_K)'$, and $\boldsymbol{u} = (u_1, \ldots, u_N)'$. Let $\boldsymbol{\mu}_f$ be the mean of factor returns, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)'$, 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 $\boldsymbol{\Sigma}$:

$$\boldsymbol{\mu}=\boldsymbol{\beta}\boldsymbol{\mu}_f+\boldsymbol{\alpha}, \qquad \boldsymbol{\Sigma}=\boldsymbol{\beta}\boldsymbol{\Sigma}_f\boldsymbol{\beta}'+\boldsymbol{\Sigma}_u. \tag{1.17}$$

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,

$$\boldsymbol{\mu}_{all}=\begin{pmatrix}\boldsymbol{\mu}_f\\ \boldsymbol{\mu}\end{pmatrix}, \quad \boldsymbol{\Sigma}_{all}=\begin{pmatrix}\boldsymbol{\Sigma}_f & \boldsymbol{\beta}'\boldsymbol{\Sigma}_f\\ \boldsymbol{\Sigma}_f\boldsymbol{\beta} & \boldsymbol{\Sigma}\end{pmatrix}. \tag{1.18}$$

We make the following assumptions:

Assumption:

B1 The number of factors included, K, is bounded;

B2 There exists $M<\infty$ such that $\max\left(\boldsymbol{\alpha}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\alpha}, \max\limits_{1\le j\le N}|\alpha_j|\right)\le M$;

B3 There exists $L<\infty$ such that $\max\limits_{1\le i\le N}|\sigma_{ii,u}|\le L$, where $\boldsymbol{\Sigma}_u=:(\sigma_{ij,u})_{1\le i,j\le N}$.

Similar to Assumption A2, for $\boldsymbol{\Sigma}_u$ in Assumption B2, we do not require its eigenvalues to be bounded, in other words, we do not require $(f_1,\ldots,f_K)'$ to be the full set of factors, and $(u_1,\ldots,u_N)'$ can still have factor structure. Our recommendation is to only include a small number of strong factors.

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

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

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

(iii) suitably combine these two portfolios.

Specifically, we have the following:

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

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

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 expressions:

$$\boldsymbol{w}_f^*=\frac{1}{\sqrt{\theta_f}}\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\mu}_f, \quad \boldsymbol{w}_u^*=\frac{1}{\sqrt{\theta_u}}\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\alpha}. \tag{1.19}$$

We have two remarks about Proposition 3. First, the decomposition above is similar in spirit to the decomposition in Theorem 4.1 of Uppal and Zaffaroni (2016). There the factors are taken to be latent factors, whereas here the factors are taken to be observable and are treated as extra investable assets. Second, in the case in which $\alpha_i = 0$ for all $i = 1, \ldots, N$, we have $\theta_u = 0$ and the optimal allocation is given by $(\sigma \boldsymbol{w}_f^*, \boldsymbol{0})$. In other words, the optimal portfolio will be fully invested in factors. However, in practice, this is unlikely, especially when one wants to only include a small number of strong factors. If, on the other hand, one intends to incorporate a large number of factors, then estimating the optimal portfolio on the factors will be of high-dimensional nature, in which case our strategy in Section 1.3 can be applied.

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

1.4.2 Estimating the maximum Sharpe ratios. Suppose that $\boldsymbol{R_t} = (R_{t1}, \ldots, R_{tN})'$ and $\boldsymbol{F_t} = (F_{t1}, \ldots, F_{tK})'$, $t = 1, \ldots, T$, 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 $\boldsymbol{F} = (\boldsymbol{F_1}, \ldots, \boldsymbol{F_T})'$. We estimate the coefficient $\boldsymbol{\beta}$ in (1.16) by regressing $\boldsymbol{R}$ on $\boldsymbol{F}$. Denote the estimated beta matrix by $\widehat{\boldsymbol{\beta}}$. Correspondingly, let $\widehat{\boldsymbol{U}} = (\widehat{\boldsymbol{U}_1}, \ldots, \widehat{\boldsymbol{U}_T})' = \boldsymbol{R} - \boldsymbol{F}\widehat{\boldsymbol{\beta}}$ be the estimator of $\boldsymbol{U} = \boldsymbol{R} - \boldsymbol{F}\boldsymbol{\beta}$. Let $\widehat{\boldsymbol{\mu}}_f$ and $\widehat{\boldsymbol{\Sigma}}_f$ denote the sample mean and sample covariance matrix of the factor return $\boldsymbol{F}$, respectively.

Three Sharpe ratios need to be estimated in the current scenario: $\sqrt{\theta_f}$, $\sqrt{\theta_u}$ and $\sqrt{\theta_{all}}$, which are the maximum Sharpe ratios on factors, idiosyncratic components and all assets, respectively. Because the number of factors included is bounded, $\sqrt{\theta_f}$ can be consistently estimated by its plug-in estimator:

$$\left| \sqrt{\widehat{\theta}_f} - \sqrt{\theta_f} \right| \xrightarrow{P} 0 \quad \text{as} \quad T \to \infty, \quad \text{where} \quad \widehat{\theta}_f = \widehat{\boldsymbol{\mu}}_f' \widehat{\boldsymbol{\Sigma}}_f^{-1} \widehat{\boldsymbol{\mu}}_f. \tag{1.20}$$

The remaining two Sharpe ratios, $\sqrt{\theta_u}$ and $\sqrt{\theta_{all}}$, involve a large number of assets, and we need to adjust for the bias in the plug-in estimator. In parallel with Proposition 2, we have the following:

Proposition 4. Define the following estimator of θ_{all}:

$$\widehat{\theta}_{all} := \frac{(T - N - K - 2)\widehat{\theta}_{s,all} - N - K}{T}, \tag{1.21}$$

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

$$\left| \sqrt{\widehat{\theta}_{all}} - \sqrt{\theta_{all}} \right| \xrightarrow{P} 0.$$

One last Sharpe ratio needs to be estimated, that is, $\sqrt{\theta_u}$, 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 (1.21) applied to $\widehat{\boldsymbol{U}}$ will be biased.

The solution to the aforementioned difficulty lies in the relationships among θ_f, θ_u and θ_{all}. Based on model (1.16), one can show that[4]

$$\theta_{all} = \theta_f + \theta_u. \tag{1.22}$$

By (1.20) and 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 $|\sqrt{\widehat{\theta}_u} - \sqrt{\theta_u}| \xrightarrow{P} 0$. Therefore, for $r_c := (1+\theta_u)/\sqrt{\theta_u}$,

$$\widehat{r_c} := \frac{1+\widehat{\theta}_u}{\sqrt{\widehat{\theta}_u}} \tag{1.23}$$

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

1.4.3 Estimating the optimal portfolio on idiosyncratic components. The optimal portfolio on the idiosyncratic components, $\boldsymbol{w}_u^*$, solves the following mean-variance portfolio problem:

$$\arg\max_{\boldsymbol{w}} \boldsymbol{\alpha}'\boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{w}'\boldsymbol{\Sigma}_u\boldsymbol{w} \le 1. \tag{1.24}$$

The optimal portfolio yields an expected return of

$$r_u^* := \sqrt{\theta_u}, \tag{1.25}$$

which equals the maximum Sharpe ratio of portfolios on (u_i). Following our regression representation (1.4) in Section 1.1, we estimate $\boldsymbol{w}_u^*$ based on the following:

$$\arg\min_{\boldsymbol{w}} E\left(r_c - \boldsymbol{u}'\boldsymbol{w}\right)^2, \quad \text{where } r_c := \frac{1+\theta_u}{\theta_u} r_u^* = \frac{1+\theta_u}{\sqrt{\theta_u}}. \tag{1.26}$$

One major difference here is that the idiosyncratic returns are not observable, and we have to rely on the estimated idiosyncratic return $\widehat{\boldsymbol{U}}$. More explicitly, based on estimated idiosyncratic return $\widehat{\boldsymbol{U}}$ and estimated $\widehat{r_c}$ in (1.23), we define our estimator of $\boldsymbol{w}_u^*$ analogously to (1.12) as the following

$$\widehat{\boldsymbol{w}_u^*} = \arg\min_{\boldsymbol{w}} \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{r_c} - \boldsymbol{w}'\widehat{\boldsymbol{U}}_t\right)^2 \quad \text{subject to} \quad ||\boldsymbol{w}||_1 \le \lambda, \tag{1.27}$$

where λ is the ℓ_1-norm tuning parameter. To give the theoretical properties of $\widehat{\boldsymbol{w}_u^*}$, the following assumptions will be needed.

[4] Relationship (1.22) is also used in Kan and Wang (2017).

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 $\rho_T := N/T \to \rho \in (0,1)$.

Assumption C2 is our sparsity requirement in this setting when factor investing is allowed. We emphasize that the sparsity assumption is put on the optimal portfolio on the **idiosyncratic components**. This amounts to say that, the mean-variance efficiency is achieved by an addition of a sparse stock portfolio to the optimal portfolio on factors.

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–B3 and C1–C3, we have as $N \to \infty$,

$$|\boldsymbol{\alpha}' \widehat{\boldsymbol{w}_u^*} - \boldsymbol{\alpha}' \boldsymbol{w}_u^*| \xrightarrow{P} 0, \tag{1.28}$$

and

$$\left| \sqrt{\widehat{\boldsymbol{w}_u^*}' \boldsymbol{\Sigma}_u \widehat{\boldsymbol{w}_u^*}} - 1 \right| \xrightarrow{P} 0. \tag{1.29}$$

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 (1 in this case).

1.4.4 Main result 2: MAXSER with factor investing. So far, we have achieved consistency in estimating θ_f, θ_u, θ_{all} and $\boldsymbol{w}_u^*$. One final item needs to be estimated: $\boldsymbol{w}_f^*$, the optimal portfolio on the factors (with risk equal to 1). This is straightforward because the number of factors included is bounded, and the simple "plug-in" estimator works. Specifically, $\widehat{\boldsymbol{w}_f^*} := \frac{1}{\sqrt{\theta_f}} \widehat{\boldsymbol{\Sigma}}_f^{-1} \widehat{\boldsymbol{\mu}}_f$ will be a consistent estimator of $\boldsymbol{w}_f^*$. 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. Define our estimator of the optimal full portfolio $\boldsymbol{w}_{all}$ as

$$\widehat{\boldsymbol{w}_{all}} := (\widehat{\boldsymbol{w}_f}, \widehat{\boldsymbol{w}}) = \sigma \left(\sqrt{\frac{\widehat{\theta}_f}{\widehat{\theta}_{all}}} \widehat{\boldsymbol{w}_f^*} - \sqrt{\frac{\widehat{\theta}_u}{\widehat{\theta}_{all}}} \widehat{\boldsymbol{\beta}}' \widehat{\boldsymbol{w}_u^*}, \sqrt{\frac{\widehat{\theta}_u}{\widehat{\theta}_{all}}} \widehat{\boldsymbol{w}_u^*} \right). \tag{1.30}$$

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

$$|\widehat{\boldsymbol{w}_{all}}' \boldsymbol{\mu}_{all} - r^*| \xrightarrow{P} 0, \quad \text{and} \quad |\widehat{\boldsymbol{w}_{all}}' \boldsymbol{\Sigma}_{all} \widehat{\boldsymbol{w}_{all}} - \sigma^2| \xrightarrow{P} 0. \tag{1.31}$$

Theorem 2 guarantees that our MAXSER portfolio with factor investing can again asymptotically yield the maximum expected return, satisfy the risk constraint, and, consequently, achieve mean-variance efficiency.

1.5 Practical implementation of MAXSER

1.5.1 Choosing λ in (1.12) or (1.27). To implement MAXSER, selecting an appropriate λ in (1.12) or (1.27), or, equivalently, the ℓ_1-norm ratio ζ defined in (1.7), is vital. Because one of our goals is to meet the risk constraint, also in light of Figure 4, we select ζ such that the estimated portfolio has a risk close to the given risk constraint.

In practice, the underlying covariance matrix $\mathbf{\Sigma}$ or $\mathbf{\Sigma}_{all}$ is unknown. To circumvent this difficulty, we use a cross-validation method. 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 in the sample. Next, for each such training set i, obtain the whole solution path $(\widehat{\boldsymbol{w}^*_\zeta})_{0\le\zeta\le 1}$ ($(\widehat{\boldsymbol{w}^*_{all,\zeta}})_{0\le\zeta\le 1}$ in scenario 2), and calculate the value $\zeta(i)$ corresponding to the portfolio that minimizes the difference between the risk computed using the validation set and the given risk constraint. The ultimate $\widehat{\zeta}$ is taken to be the average of $(\zeta(i),\ i=1,\ldots,10)$.

The method above is similar to the cross-validation procedure used by other norm-constrained portfolio optimization approaches (see, e.g., DeMiguel et al. 2009a). Our simulation and empirical analysis demonstrates that cross-validation effectively helps control out-of-sample risk.

1.5.2 Adjustment of $\widehat{\theta}$, $\widehat{\theta}_u$, and $\widehat{\theta}_{all}$. Kan and Zhou (2007) observe that $\widehat{\theta}$ in (1.8), the unbiased estimator of the square of maximum Sharpe ratio, often takes negative values. They propose an adjusted estimator that improves over the unbiased estimator:

$$\widehat{\theta}_{adj}=\frac{(T-N-2)\widehat{\theta}_s-N}{T}+\frac{2(\widehat{\theta}_s)^{N/2}(1+\widehat{\theta}_s)^{-(T-2)/2}}{T\,B_{\widehat{\theta}_s/(1+\widehat{\theta}_s)}(N/2,(T-N)/2)}, \tag{1.32}$$

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 second setting that allows for factor investing, we adopt the following adjustment of $\widehat{\theta}_{all}$:

$$\widehat{\theta}_{all,adj}=\frac{(T-N-K-2)\widehat{\theta}_{s,all}-N-K}{T}+\frac{2(\widehat{\theta}_{s,all})^{\frac{N+K}{2}}(1+\widehat{\theta}_{s,all})^{-\frac{T-2}{2}}}{T\,B_{\frac{\widehat{\theta}_{s,all}}{(1+\widehat{\theta}_{s,all})}}\left(\frac{N+K}{2},\frac{T-N-K}{2}\right)}, \tag{1.33}$$

where, recall that, $\widehat{\theta}_{s,all}$ is the plug-in estimator of θ_{all}. The adjusted $\widehat{\theta}_u$ is $\widehat{\theta}_{u,adj}:=\widehat{\theta}_{all,adj}-\widehat{\theta}_f$.

1.5.3 Subpool selection. When the number of assets is large, we suggest a simple subpool selection procedure to reduce the number of assets to which MAXSER is applied. The motivation for this procedure is twofold.

First, the variable selection literature reveals that screening the variables before implementing LASSO can largely improve computational efficiency and practical performance (see, e.g., Fan and Lv 2008 and Tibshirani et al. 2012). Second, the convergence rates of the expected return and risk of the MAXSER portfolio depend on the dimension N and sample size T; for a given sample size T, the smaller the dimension N, the faster the convergence.

Specifically, given a pool of N individual assets, we select a subpool containing $N^- < N$ assets in the following manner. Consider the second scenario in which factor investing is included. Proposition 3 provides decompositions of the whole portfolio into portfolios on factors and idiosyncratic components, whose respective contributions to the Sharpe ratio have a simple relationship as in Equation (1.22). Such decompositions provide an important insight, which is that the gain of investing additionally in individual assets originates from θ_u; the higher the θ_u, the larger the gain. Therefore, we select the subpool that offers a high θ_u, or equivalently, a high θ_{all} when combined with the factors. To find such a subpool, we randomly form a large number of subpools—for example, 1,000 subpools—each containing N^- assets from the original pool, combine each of the subpools with the factors, and then calculate the corresponding squared maximum Sharpe ratios $\widehat{\theta}_{all,adj}$. The subpool for constructing the MAXSER portfolio can then be naturally selected as a subpool with a high $\widehat{\theta}_{all,adj}$. For example, one can select the subpool with $\widehat{\theta}_{all,adj}$ corresponding to the 95% quantile[5] of all $\widehat{\theta}_{all,adj}$. Analogously, for the first scenario (no factor investing), one may form 1,000 subpools of randomly picked N^- assets, calculate the squared maximum Sharpe ratio $\widehat{\theta}_{adj}$ for each subpool, and select the subpool that corresponds to the 95% quantile of all $\widehat{\theta}_{adj}$. In both scenarios, the only computation involved is the estimation of the squared maximum Sharpe ratio ($\widehat{\theta}_{adj}$ in (1.32) for the first scenario; $\widehat{\theta}_{all,adj}$ in (1.33) for the second scenario). The entire subpool selection procedure can be completed within seconds for N's in the order of hundreds.

1.5.4 Implementation steps.

Scenario 1: The asset pool comprises individual assets only

In this case, our method consists of the following steps:

Step 0 If the number of assets is large, conduct the subpool selection procedure in Section 1.5.3;

Step 1 Estimate the square of the maximum Sharpe ratio by $\widehat{\theta}_{adj}$ in (1.32), and compute the response $\widehat{r_c}$ in (1.10);

Step 2 Select λ through cross-validation according to the procedure in Section 1.5.1. Denote the selected value by $\widehat{\lambda}$;

[5] It may not be desirable to select the subpool with the highest $\widehat{\theta}_{all,adj}$ because of potential abnormal outliers. We conduct the selection with 95% quantile in our numerical studies.

Step 3 Set λ in (1.12) to be $\widehat{\lambda}$ and solve for the MAXSER portfolio $\widehat{\boldsymbol{w}^*}$ in (1.12).

Scenario 2: When factors are included in the asset pool

When incorporating factor investing, MAXSER is implemented as follows:

Step 0 If the number of assets is large, conduct the subpool selection procedure in Section 1.5.3;

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

Step 2 Compute the estimates of the square of the maximum Sharpe ratios $\widehat{\theta}_f$, $\widehat{\theta}_{all,adj}$ and $\widehat{\theta}_{u,adj}$, and compute the response $\widehat{r_c}$ in (1.23);

Step 3 Select λ through cross-validation according to the procedure in Section 1.5.1. Denote the chosen value by $\widehat{\lambda}$;

Step 4 Set λ in (1.27) to be $\widehat{\lambda}$ and solve for $\widehat{\boldsymbol{w}_u^*}$;

Step 5 Compute $\widehat{\boldsymbol{w}_f^*}$ and plug in the estimates from the previous steps into (1.30) to obtain the MAXSER portfolio $\widehat{\boldsymbol{w}_{all}}$.

2. Simulation Analysis

In this section, we examine and compare the performance of our MAXSER portfolio with that of various benchmark strategies. The simulated asset pool includes both stocks and factors. Parameters for generating returns are calibrated from S&P 500 index constituents and Fama-French three factors (see Section 2.2).

2.1 Benchmark portfolios

In addition to the plug-in and nonlinear shrinkage portfolios that we discussed in the Introduction, we include various other strategies in our comparisons. Table 1 gives the complete list.

Among the portfolios under comparison, a special one is the portfolio "Factor," which is the mean-variance portfolio on factors. Specifically, recall that $\widehat{\boldsymbol{\mu}}_f$ and $\widehat{\boldsymbol{\Sigma}}_f$ are the sample mean and sample covariance matrix of the factor returns, respectively. The Factor portfolio is defined as

$$\widehat{\boldsymbol{w}}_{Fac} = \frac{\sigma}{\sqrt{\widehat{\boldsymbol{\mu}}_f' \widehat{\boldsymbol{\Sigma}}_f^{-1} \widehat{\boldsymbol{\mu}}_f}} \widehat{\boldsymbol{\Sigma}}_f^{-1} \widehat{\boldsymbol{\mu}}_f \left(= \sigma \widehat{\boldsymbol{w}_f^*} \right). \tag{2.1}$$

This portfolio is special in the sense that it only involves a small number of assets (three in our case). Consequently, the plug-in formula (2.1) yields a nearly optimal portfolio. Including such a portfolio in the comparison would reveal whether additional investment in individual assets is beneficial.

Regarding the other portfolios, the MV and GMV portfolios are constructed by replacing the covariance matrix with the sample/linear shrinkage

Table 1
List of benchmark portfolios and their abbreviations

Portfolio	Abbreviation
Plug-in MV on factors	Factor
Three-fund portfolio by Kan and Zhou 2007	KZ
MV/GMV with different covariance matrix estimates	
MV with sample cov	MV-P
MV with linear shrinkage cov	MV-LS
MV with nonlinear shrinkage cov	MV-NLS
MV with nonlinear shrinkage cov adjusted for factor models	MV-NLSF
GMV with linear shrinkage cov	GMV-LS
GMV with nonlinear shrinkage cov	GMV-NLS
MV with short-sale constraint and cross-validation	
MV with sample cov and short-sale-CV	MV-P-SSCV
MV with linear shrinkage cov and short-sale-CV	MV-LS-SSCV
MV with nonlinear shrinkage cov and short-sale-CV	MV-NLS-SSCV
MV with ℓ_1-norm constraint and cross-validation	
MV with sample cov and ℓ_1-CV	MV-P-L1CV
MV with linear shrinkage cov and ℓ_1-CV	MV-LS-L1CV
MV with nonlinear shrinkage cov and ℓ_1-CV	MV-NLS-L1CV

"MV" stands for mean-variance portfolio, and "GMV" stands for global minimum variance portfolio.

(Ledoit and Wolf 2004)/nonlinear shrinkage/nonlinear shrinkage adjusted for factor model (Ledoit and Wolf 2017) estimator in the respective MV and GMV portfolio weight formulas. Details regarding the portfolio "KZ"[6] can be found in Kan and Zhou (2007).

In addition, we construct portfolios with either a short-sale or ℓ_1-norm constraint on the portfolio weights.[7] The MV portfolios with a short-sale constraint—"MV-P-SSCV," "MV-LS-SSCV," and "MV-NLS-SSCV"—have a short position constraint added to the mean-variance optimization in which the covariance matrix is estimated by the sample, linear shrinkage, and nonlinear shrinkage estimator. For portfolios "MV-P-L1CV," "MV-LS-L1CV," and "MV-NLS-L1CV," we impose an ℓ_1-norm constraint. In this manner, these portfolios

[6] Following Kan and Zhou (2007), the risk aversion is set to 3. This risk aversion is not designed to meet a given risk constraint; thus, in the following we do not comment on the risk of the portfolio KZ.

[7] When the number of stocks is large, the cost of computing short-sale or ℓ_1-norm-constrained portfolios based on the optimization form (1.1) is high. To reduce the computational costs, we solve for the short-sale and ℓ_1-norm constrained portfolios based on the dual form (1.2), which leads to much more efficient computation. 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. The MV portfolio with a short-sale or ℓ_1-norm constraint is solved by

$$\underset{\boldsymbol{w}}{\operatorname{argmin}}\, \boldsymbol{w}'\widehat{\boldsymbol{\Sigma}}_{all}\boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{w}'\widehat{\boldsymbol{\mu}}_{all} \geq \widehat{r^*} \text{ and } w_i \geq -\lambda_{SS} \text{ for all } i = 1, \ldots, N, \tag{2.2}$$

or

$$\underset{\boldsymbol{w}}{\operatorname{argmin}}\, \boldsymbol{w}'\widehat{\boldsymbol{\Sigma}}_{all}\boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{w}'\widehat{\boldsymbol{\mu}}_{all} \geq \widehat{r^*} \text{ and } ||\boldsymbol{w}||_1 \leq \lambda, \tag{2.3}$$

respectively, where $\widehat{r^*} := \sigma\sqrt{\widehat{\theta}_{all}}$ is the estimated return constraint corresponding to the risk constraint level σ, $\lambda_{SS} > 0$ is the short position threshold, and λ is the ℓ_1-norm constraint. Both λ_{SS} and λ are determined through the same 10-fold cross-validation procedure, which is described in Section 1.5.1.

enjoy the same benefit in terms of risk control as our MAXSER portfolio. We include these portfolios to demonstrate that the advantage of MAXSER is not solely due to the ℓ_1-norm constraint, but rather, more fundamentally, its methodology.

2.2 Parameter setting

We simulate monthly returns from model (1.15) with parameters calibrated from real data. Specifically, for the parameters of factor returns, the mean $\boldsymbol{\mu}_f$ and covariance matrix $\boldsymbol{\Sigma}_f$ are taken to be the sample mean and sample covariance matrix of the FF3 monthly returns from 2007 to 2016. Regarding $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, we randomly select 100 stocks out of those that were in the S&P 500 index for the entire period 2007 to 2016. We then regress their monthly excess returns over the FF3 returns and set the resultant slopes to be the β_i's; the α_i's are obtained through hard-thresholding the estimated intercepts with a threshold of two standard errors. Finally, the covariance matrix of idiosyncratic returns, $\boldsymbol{\Sigma}_u$, is obtained by applying the soft-thresholding method proposed in Rothman (2012)[8] to the sample covariance matrix of the residuals in the regression above. Notably, the idiosyncratic returns generated from the parameters selected as above may still have factor structure.

2.3 Simulation comparisons

Here, we present the results for data generated under multivariate normal distribution. In the Internet Appendix, we provide results based on data generated under heavy-tailed distribution. Returns of 100 stocks and three factors are generated using the parameters described in Section 2.2. The level of risk constraint is set to be $\sigma = 0.04$. Because the number of stocks is large, when implementing MAXSER, we start from step 0 described in Section 1.5.4 and select subpools containing 50 stocks.

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

From Table 2, we observe the following:

In terms of *risk control*, the risk of our MAXSER portfolio is close to the given constraint. The Factor portfolio also has a risk close to the constraint because for such a low-dimensional portfolio, the simple plug-in estimator is sufficient. On the other hand, MV portfolios with the covariance matrix estimated by the sample, linear shrinkage, and nonlinear shrinkage estimators severely violate the risk constraint. When $T = 120$, their risks exceed the constraint by approximately 640%, 105%, and 35%, respectively. By contrast, when either a short-sale or ℓ_1-norm constraint is imposed, the corresponding MV strategies

[8] The soft-thresholding method can be implemented in R using the package "PDSCE." In our setting, the penalty parameter "lam" is set to be 0.5.

Table 2
Summary of risks and Sharpe ratios of the portfolios under comparison

$\sigma = 0.04$	$T = 120$		$T = 240$	
Portfolio	Risk	Sharpe ratio	Risk	Sharpe ratio
Factor	0.041 (0.003)	0.401 (0.169)	0.041 (0.002)	0.467 (0.108)
KZ	0.052 (0.040)	0.329 (0.184)	0.091 (0.031)	0.909 (0.130)
MAXSER	0.043 (0.005)	1.083 (0.302)	0.041 (0.003)	1.422 (0.200)
MV/GMV with different covariance matrix estimates				
MV-P	0.296 (0.072)	0.367 (0.168)	0.070 (0.005)	0.911 (0.123)
MV-LS	0.082 (0.006)	0.697 (0.160)	0.061 (0.004)	0.943 (0.117)
MV-NLS	0.054 (0.017)	0.945 (0.183)	0.049 (0.004)	1.199 (0.117)
MV-NLSF	0.044 (0.002)	0.837 (0.139)	0.042 (0.001)	1.068 (0.104)
GMV-LS	0.013 (0.001)	0.438 (0.132)	0.009 (0.000)	0.450 (0.102)
GMV-NLS	0.015 (0.003)	0.553 (0.148)	0.009 (0.001)	0.539 (0.167)
MV with short-sale constraint and cross-validation				
MV-P-SSCV	0.057 (0.035)	0.400 (0.112)	0.038 (0.008)	0.754 (0.259)
MV-LS-SSCV	0.039 (0.025)	0.666 (0.177)	0.038 (0.008)	0.744 (0.275)
MV-NLS-SSCV	0.035 (0.023)	0.850 (0.259)	0.038 (0.010)	0.847 (0.396)
MV with ℓ_1-norm constraint and cross-validation				
MV-P-L1CV	0.041 (0.011)	0.539 (0.215)	0.036 (0.006)	1.057 (0.185)
MV-LS-L1CV	0.032 (0.012)	0.726 (0.179)	0.036 (0.005)	1.121 (0.177)
MV-NLS-L1CV	0.029 (0.011)	0.973 (0.171)	0.037 (0.005)	1.207 (0.154)

The summary is based on 1,000 replications, where returns of sample size $T = 120$ or 240 are generated from multivariate normal distribution. The risk constraint is set to 0.04. The theoretical maximum Sharpe ratio is 1.882. The average value and standard deviation (in parentheses) of each performance measure are reported.

lead to portfolios with risk close to the risk constraint level. This indicates that the cross-validation procedure works for these strategies just as effectively as for our own MAXSER portfolio in terms of risk control.

In terms of the *Sharpe ratio*, as the sample size T increases from 120 to 240, most portfolios improve. The MAXSER portfolio has the highest Sharpe ratio among all the portfolios compared. In the $T = 240$ case, MAXSER attains approximately 76% of the theoretical maximum Sharpe ratio on average, whereas the Sharpe ratio of the MV-NLS-L1CV portfolio, the second highest among all portfolios, is approximately 64%. Moreover, the 95% confidence interval of the mean Sharpe ratio of the MAXSER portfolio is $[1.397, 1.447]$. By comparison, the 95% confidence interval for the difference between the mean Sharpe ratios of the MAXSER and MV-NLS-L1CV portfolios is $[0.186, 0.243]$. This range confirms that the improvement in the Sharpe ratio achieved by MAXSER is not only statistically significant but also economically large. In addition, compared with those of the short-sale and ℓ_1-norm constrained MV portfolios, the Sharpe ratio of the MAXSER portfolio is significantly higher, indicating that the outstanding performance of the MAXSER portfolio is fundamentally due to our methodology rather than solely imposing constraint.

In summary, our MAXSER portfolio effectively controls risk and is significantly more mean-variance efficient than the benchmark portfolios. Table A1 in the Internet Appendix provides similar results for heavy-tailed returns.

3. Empirical Analysis

We investigate the performance of our strategy based on two stock universes drawn from the Dow Jones Industrial Average (DJIA) 30 index constituents and the S&P 500 index constituents, together with FF3 factors for the purpose of factor investing.[9] We report the results based on the two stock universes in Sections 3.1 and 3.2, respectively.

3.1 DJIA 30

3.1.1 Data and investment rolling window scheme. We first evaluate our proposed portfolio, the MAXSER portfolio, using the stock universe of the DJIA 30 index constituents and FF3 factors. We obtain the lists of DJIA 30 index constituents from Compustat and CRSP. The evaluation is conducted under a rolling-window scheme. Specifically, at the beginning of each month, an asset pool is formed by including the constituents of the DJIA 30 index at that point in time and the FF3 factors. For all strategies to be compared, portfolios are constructed using the monthly excess returns[10] during the prior T months, where T is the sample size to be specified. A stock is excluded from the asset pool if it has missing data in the T-month training period. Therefore, the number of stocks varies over time and can be smaller than the total number of constituents. The risk constraint σ is set to be 0.05, which is the standard deviation of the DJIA 30 index monthly (excess) returns in the first training period.

3.1.2 Performance evaluation. We evaluate the performance of the MAXSER and other benchmark portfolios[11] using both risk and annualized Sharpe ratio computed from their respective (out-of-sample) monthly returns. We also investigate the effect of transaction costs and report the comparisons based on returns net of transaction costs.

For the DJIA data set, which has approximately 30 stocks in each asset pool, we use a sample size of $T = 60$; that is, each training set contains the monthly returns of the previous five years. We consider two testing periods—a 40-year period from 1977 to 2016 and a 20-year period from 1997 to 2016—that result in 480 and 240 out-of-sample monthly returns, respectively, for each strategy.

---

[9] On the basis of our empirical analysis results, for long testing periods, such as the 40 years from 1977 to 2016, comparing the two scenarios without or with factor investing shows that investing in both stocks and factors such as the FF3 factors is beneficial for all strategies. For this reason, in the simulation (Section 2) and empirical analysis, we focus on the setting with factor investing. For the long-short factor portfolios SMB and HML, the weight on a factor is the proportion of the value of long positions, which equals the value of short positions.

[10] For computing excess returns, we obtain the risk-free rate r_f from the Kenneth French Data Library.

[11] In addition to the portfolios included in the simulation analysis, the equally weighted portfolio (the "1/N rule"), which is an important benchmark portfolio as shown in DeMiguel et al. (2009b), is included in our empirical analysis along with the index.

Table 3
Portfolio performance without transaction costs based on DJIA 30 constituents and Fama-French three factors

DJIA 30 constituents and FF3 (*without transaction costs*)				$T=60$		$\sigma=0.05$
Period	1977–2016			1997–2016		
Portfolio	Risk	Sharpe ratio	p-value	Risk	Sharpe ratio	p-value
Index	0.043	0.270	.000	0.043	0.310	.001
Equally weighted	0.042	0.328	.000	0.044	0.307	.001
Factor	0.055	0.427	.000	0.058	0.254	.000
KZ	0.104	0.250	.000	0.097	0.265	.000
MAXSER	0.060	0.556	–	0.064	0.567	–
MV/GMV with different covariance matrix estimates						
MV-P	0.116	0.196	.000	0.132	0.292	.000
MV-LS	0.070	0.132	.000	0.077	0.376	.003
MV-NLS	0.068	0.166	.000	0.073	0.352	.001
MV-NLSF	0.067	0.232	.000	0.070	0.290	.000
GMV-LS	0.016	0.453	.030	0.018	0.307	.000
GMV-NLS	0.016	0.364	.000	0.018	0.274	.000
MV with short-sale constraint and cross-validation						
MV-P-SSCV	0.045	0.407	.001	0.045	0.448	.042
MV-LS-SSCV	0.044	0.376	.000	0.045	0.469	.070
MV-NLS-SSCV	0.044	0.443	.005	0.044	0.473	.072
MV with ℓ_1-norm constraint and cross-validation						
MV-P-L1CV	0.043	0.136	.000	0.043	0.253	.000
MV-LS-L1CV	0.041	0.102	.000	0.040	0.366	.002
MV-NLS-L1CV	0.040	0.131	.000	0.038	0.317	.000

This table summarizes risks, Sharpe ratios, and p-values for the Sharpe ratio test (3.1) for the portfolios under comparison on DJIA 30 index constituents and Fama-French three factors. The two testing periods are 1977–2016 and 1997–2016. The risk constraint is $\sigma=0.05$, which is the standard deviation of the monthly excess returns on DJIA 30 index during the first training period from 1972 to 1976.

In addition to comparing out-of-sample risks and Sharpe ratios, we conduct hypothesis tests regarding the Sharpe ratio to verify the statistical significance of the advantage of our MAXSER portfolio. More specifically, we test

$$H_0: SR_{MAXSER} \leqslant SR_0 \quad vs \quad H_a: SR_{MAXSER} > SR_0, \tag{3.1}$$

where SR_{MAXSER} denotes the Sharpe ratio of the MAXSER portfolio, and SR_0 denotes the Sharpe ratio of the portfolio under comparison. The test is conducted using the correction of Memmel (2003) to the test of Jobson and Korkie (1981).

3.1.3 Comparison summary. A summary comparing the MAXSER and benchmark portfolios based on returns with and without transaction costs follows.

3.1.3.1 Without transaction costs. Table 3, which shows the risk, Sharpe ratio, and p-value for testing (3.1) for each benchmark portfolio, provides a summary without considering transaction costs.

From Table 3, we observe the following:

In terms of *risk control*, the risk of the MAXSER portfolio is close to those of the index, the equally weighted, the Factor, the short-sale constrained, and ℓ_1-norm constrained portfolios. By contrast, the risk of the plug-in ("MV-P") portfolio is nearly twice the risk constraint level and unacceptable for investors.

In terms of *Sharpe ratio*, first of all, our MAXSER portfolio yields the highest Sharpe ratio. Second, the ℓ_1-norm constrained portfolios, MV-P-L1CV, MV-LS-L1CV and MV-NLS-L1CV, yield considerably lower Sharpe ratios than that of the MAXSER portfolio. This again confirms that the advantage of MAXSER is fundamental rather than solely due to imposing the constraint. Moreover, the small p-values of the Sharpe ratio tests confirm that the advantage of MAXSER is not only economically large but also statistically significant. A new observation from this study is that, investing in individual stocks in addition to factors using our strategy MAXSER can yield substantial gain, as demonstrated by the comparison of Sharpe ratios of the MAXSER and Factor portfolios.

In summary, MAXSER effectively controls out-of-sample risk and dominates the benchmark strategies in terms of mean-variance efficiency.

3.1.3.2 With transaction costs. Next, we take transaction costs into account and compute the returns net of transaction costs. Following Engle et al. (2012), we compute the portfolio return net of transaction costs in each period as follows:

$$r_{net}(t)=\left(1-\sum_j c_{t,j}|w_j(t+1)-w_j(t+)|\right)(1+r(t))-1, \tag{3.2}$$

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 that measures the transaction cost per dollar traded for trading asset j, and $r(t)$ is the portfolio return without transaction cost in period t. For the cost level $c_{t,j}$, based on Brandt et al. (2009) and Engle et al. (2012) and considering that we only include large-cap stocks, for each individual stock, we set $c_{t,j}$ to decrease linearly from 0.6% to 0.1% during the period 1977 to 1990, and set it to be a constant 0.1% from 1991 to 2016. Regarding factors, following Table 3 of Novy-Marx and Velikov (2016), the monthly returns of SMB and HML are subtracted by 5 bps to account for the rebalancing costs. Furthermore, the cost level $c_{t,j}$ for trading FF3 factors is set to be triple that for individual stocks. Table 4 presents the risks, Sharpe ratios, and p-values after deducting transaction costs.

Table 4 shows the advantage of our MAXSER portfolio. For the period 1977 to 2016, because of the high transaction cost levels before 1990, all the mean-variance portfolios are severely impaired. Although the MAXSER

Table 4
Portfolio performance with transaction costs based on DJIA 30 constituents and Fama-French three factors

DJIA 30 constituents and FF3 (*with transaction costs*) $T = 60$ $\sigma = 0.05$

Period	1977–2016			1997–2016		
Portfolio	Risk	Sharpe ratio	p-value	Risk	Sharpe ratio	p-value
Index	0.043	0.270	.002	0.043	0.310	.011
Equally weighted	0.042	0.317	.724	0.044	0.300	.108
Factor	0.055	0.265	.273	0.058	0.146	.000
KZ	0.108	−0.134	.000	0.098	0.040	.000
MAXSER	0.061	0.284	–	0.064	0.402	–
MV/GMV with different covariance matrix estimates						
MV-P	0.117	−0.073	.000	0.132	0.101	.000
MV-LS	0.071	−0.014	.000	0.077	0.299	.067
MV-NLS	0.069	−0.077	.000	0.073	0.213	.002
MV-NLSF	0.067	0.045	.000	0.070	0.187	.000
GMV-LS	0.016	0.313	.716	0.018	0.213	.005
GMV-NLS	0.017	0.079	.000	0.018	0.066	.000
MV with short-sale constraint and cross-validation						
MV-P-SSCV	0.046	−0.258	.000	0.045	0.147	.000
MV-LS-SSCV	0.045	−0.042	.000	0.045	0.323	.105
MV-NLS-SSCV	0.045	−0.099	.000	0.044	0.299	.047
MV with ℓ_1-norm constraint and cross-validation						
MV-P-L1CV	0.044	−0.350	.000	0.043	0.011	.000
MV-LS-L1CV	0.042	−0.127	.000	0.040	0.281	.036
MV-NLS-L1CV	0.041	−0.232	.000	0.038	0.172	.000

This table summarizes risks, Sharpe ratios, and p-values of Sharpe ratio test (3.1) based on returns net of transaction costs of the portfolios on DJIA 30 index constituents and Fama-French three factors. The two testing periods are 1977–2016 and 1997–2016. The risk constraint is $\sigma = 0.05$, the standard deviation of the monthly excess returns on DJIA 30 index from 1972 to 1976, the first training period. For the index portfolio, we did not account for its transaction costs, although, in practice, investing in the index comes at an expense, such as tracking costs. Therefore, the actual returns of the index are worse than those reported here, and this is why the Index row is marked in gray color.

portfolio[12] is one of the least affected mean-variance portfolios, its Sharpe ratio becomes lower than those of the equally weighted and GMV-LS portfolios, both of which incur low transaction costs. However, over the more recent 20 years from 1997 to 2016, thanks to the reduction in the transaction cost levels, MAXSER maintains its advantage and delivers a Sharpe ratio that is higher than those of the other strategies, with all p-values being small.

Furthermore, comparing the MAXSER and Factor portfolios shows that, with transaction cost taken into account, MAXSER still enhances the performance.

3.2 S&P 500

We now consider a larger stock universe, S&P 500 index constituents, to form larger portfolios.

[12] We do not incorporate the transaction costs in the optimization; we leave the extension of MAXSER to the settings with transaction cost to future studies.

Table 5
Portfolio performance without transaction costs based on S&P 500 constituents and Fama-French three factors

S&P 500 Constituents and FF3 (*without transaction costs*) $T=120$ $\sigma=0.04$

Period	1977–2016			1997–2016		
Portfolio	Risk	Sharpe ratio	p-value	Risk	Sharpe ratio	p-value
Index	0.043	0.279	.000	0.044	0.302	.000
Equally weighted	0.047	0.332	.000	0.049	0.344	.001
Factor	0.040	0.517	.002	0.045	0.409	.005
KZ	0.081	0.369	.000	0.087	0.331	.001
MAXSER	0.047	0.667	–	0.053	0.591	–
MV/GMV with different covariance matrix estimates						
MV-P	0.347	0.383	.000	0.367	0.257	.000
MV-LS	0.079	0.248	.000	0.078	0.093	.000
MV-NLS	0.061	0.232	.000	0.064	0.091	.000
MV-NLSF	0.054	0.348	.000	0.057	0.141	.000
GMV-LS	0.022	0.277	.000	0.025	0.436	.027
GMV-NLS	0.025	0.271	.000	0.027	0.467	.063
MV with short-sale constraint and cross-validation						
MV-P-SSCV	0.061	0.347	.000	0.067	0.316	.000
MV-LS-SSCV	0.054	0.120	.000	0.058	0.157	.000
MV-NLS-SSCV	0.054	0.096	.000	0.058	0.139	.000
MV with ℓ_1-norm constraint and cross-validation						
MV-P-L1CV	0.047	0.318	.000	0.047	0.128	.000
MV-LS-L1CV	0.044	0.047	.000	0.048	−0.053	.000
MV-NLS-L1CV	0.043	0.060	.000	0.048	−0.059	.000

This table summarizes risks, Sharpe ratios and p-values of Sharpe ratio test (3.1) for the portfolios under comparison on S&P 500 index constituents and Fama-French three factors. The testing periods are 1977–2016 and 1997–2016. The risk constraint is set to be $\sigma=0.04$, which is the standard deviation of the monthly excess returns on S&P 500 index from 1967 to 1976, the first training period. Here, the Sharpe ratios of the factor portfolio are different from those in Table 3 because different lengths of training periods are used.

3.2.1 Data and portfolio construction. We perform the analysis under a rolling-window scheme similar to that described in Section 3.1. All portfolios under comparison are rebalanced monthly, and the out-of-sample returns over the period 1977 to 2016 are recorded. At the beginning of each year, we randomly form pools of 100 stocks from the S&P 500 index constituents that have complete price data for the prior 120 months. We construct portfolios using the excess returns of the 100 stocks and FF3 factors over the same period. The portfolios' weights are updated for every half year (in every July, in the selected 100-stock pool, stocks that have missing data from January to June of that year are excluded from the stock pool for the rest of the year). Finally, because the number of stocks is large in this setting, when implementing MAXSER, we start from step 0 described in Section 1.5.4 and select subpools containing 50 stocks.

3.2.2 Comparison summary. The comparison of the MAXSER and benchmark portfolios in terms of risks, Sharpe ratios, and p-values of the Sharpe ratio tests (3.1), without considering transaction costs, is summarized in Table 5.

Table 6
Portfolio performance with transaction costs based on S&P 500 constituents and Fama-French three factors

S&P 500 Constituents and FF3 (*with transaction costs*) $T = 120$ $\sigma = 0.04$

Period	1977–2016			1997–2016		
Portfolio	Risk	Sharpe ratio	p-value	Risk	Sharpe ratio	p-value
Index	0.043	0.279	.003	0.044	0.302	.012
Equally weighted	0.047	0.307	.012	0.049	0.330	.030
Factor	0.040	0.408	.228	0.045	0.330	.013
KZ	0.082	0.009	.000	0.087	0.160	.000
MAXSER	0.048	0.445	–	0.053	0.483	–
MV/GMV with different covariance matrix estimates						
MV-P	0.349	−0.185	.000	0.357	−0.018	.000
MV-LS	0.079	0.066	.000	0.078	0.011	.000
MV-NLS	0.061	0.099	.000	0.064	0.022	.000
MV-NLSF	0.054	0.175	.000	0.057	0.054	.000
GMV-LS	0.022	0.104	.000	0.025	0.350	.044
GMV-NLS	0.025	0.142	.000	0.027	0.398	.139
MV with short-sale constraint and cross-validation						
MV-P-SSCV	0.062	0.059	.000	0.068	0.174	.000
MV-LS-SSCV	0.054	−0.043	.000	0.059	0.083	.000
MV-NLS-SSCV	0.054	−0.028	.000	0.058	0.075	.000
MV with ℓ_1-norm constraint and cross-validation						
MV-P-L1CV	0.047	0.040	.000	0.047	−0.013	.000
MV-LS-L1CV	0.044	−0.101	.000	0.048	−0.112	.000
MV-NLS-L1CV	0.043	−0.059	.000	0.048	−0.110	.000

This table summarizes risks, Sharpe ratios and p-values of Sharpe ratio test (3.1) based on returns net of transaction costs of the portfolios on S&P 500 index constituents and Fama-French three factors. The testing periods are 1977–2016 and 1997–2016. The risk constraint is $\sigma = 0.04$, the standard deviation of the monthly excess returns on S&P 500 index from 1967 to 1976, the first training period. Here, similar to Table 4, for the index portfolio, we did not account for its transaction costs. Hence, its actual returns are worse than those reported here, and the Index row is marked in gray color.

Table 5 indicates that the MAXSER portfolio again carries a risk close to the risk constraint and has the highest Sharpe ratio. The small p-values reveal that the MAXSER portfolio has a statistically significant advantage over the other portfolios under comparison.

Next, we take transaction costs into account. The returns net of transaction costs are again computed using formula (3.2) and the cost functions described below (3.2). The comparison is summarized in Table 6.

Table 6 shows that MAXSER portfolio maintains its advantage over other portfolios.

In summary, our MAXSER portfolio outperforms benchmark portfolios regardless of whether transaction costs are considered.

4. Conclusion

In this paper, we propose a new approach for estimating the mean-variance efficient portfolio when the number of assets is not small compared with the sample size. The approach builds upon a novel unconstrained regression representation of the mean-variance portfolio problem. Scenarios both without

and with factor investing are considered. In both scenarios, we prove that our strategy, MAXSER, asymptotically achieves the mean-variance efficiency and meanwhile effectively controls the risk. To the best of our knowledge, this is the first method that simultaneously achieves these two goals for large portfolios.

The sound theoretical properties of MAXSER are supported by comprehensive numerical analysis. In both simulation and empirical analysis, MAXSER effectively controls the risk, and more importantly, yields high Sharpe ratios. Our empirical analysis reveals that, investing in individual stocks in addition to the Fama-French three factors can significantly enhance portfolio performance.

References

Bai, Z., H. Liu, and W.-K. Wong. 2009. Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory. *Mathematical Finance* 19:639–67.

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

Best, M. J., and R. R. Grauer. 1991. On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies* 4:315–42.

Black, F., and R. B. Litterman. 1991. Asset allocation: Combining investor views with market equilibrium. *Journal of Fixed Income* 1:7–18.

Brandt, M. W., P. Santa-Clara, and R. Valkanov. 2009. Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies* 22:3411–47.

Britten-Jones, M. 1999. The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance* 54:655–71.

Brodie, J., I. Daubechies, C. De Mol, D. Giannone, and I. Loris. 2009. Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences* 106:12267–72.

Chatterjee, S. 2014. Assumptionless consistency of the Lasso. arXiv, arXiv:1303.5817. https://arxiv.org/abs/1303.5817.

Chen, J., and M. Yuan. 2016. Efficient portfolio selection in a large market. *Journal of Financial Econometrics* 14:496–524.

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

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

DeMiguel, V., L. Garlappi, and R. Uppal. 2009. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies* 22:1915–53.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani. 2004. Least angle regression. *Annals of Statistics* 32:407–99.

El Karoui, N. 2010. High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation. *Annals of Statistics* 38:3487–566.

———. 2013. On the realized risk of high-dimensional Markowitz portfolios. *SIAM Journal on Financial Mathematics* 4:737–83.

Engle, R., R. Ferstenberg, and J. Russell. 2012. Measuring and modeling execution cost and risk. *Journal of Portfolio Management* 38:14–28.

Fan, J., Y. Li, and K. Yu. 2012. Vast volatility matrix estimation using high-frequency data for portfolio selection. *Journal of the American Statistical Association* 107:412–28.

Fan, J., and J. Lv. 2008. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70:849–911.

Fan, J., J. Zhang, and K. Yu. 2012. Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association* 107:592–606.

Fastrich, B., S. Paterlini, and P. Winker. 2015. Constructing optimal sparse portfolios using regularization methods. *Computational Management Science* 12:417–34.

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

Green, R. C., and B. Hollifield. 1992. When will mean-variance efficient portfolios be well diversified? *Journal of Finance* 47:1785–809.

Jagannathan, R., and T. Ma. 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance* 58:1651–84.

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

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

Kan, R., and X. Wang. 2017. On the economic value of alphas. Working Paper.

Kan, R., and G. Zhou. 2007. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis* 42:621–56.

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

Lai, T. L., H. Xing, and Z. Chen. 2011. Mean-variance portfolio optimization when means and covariances are unknown. *Annals of Applied Statistics* 5:798–823.

Ledoit, O., and M. Wolf. 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10:603–21.

———. 2004. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88:365–411.

———. 2017. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks.*Review of Financial Studies* 30:4349–88.

Markowitz, H. 1952. Portfolio selection. *Journal of Finance* 7:77–91.

Memmel, C. 2003. Performance hypothesis testing with the Sharpe ratio. *Finance Letters* 1:21–23.

Michaud, R. O. 1989. The Markowitz optimization enigma: Is “optimized" optimal? *Financial Analysts Journal* 45:31–42.

Novy-Marx, R., and M. Velikov. 2016. A taxonomy of anomalies and their trading costs. *Review of Financial Studies* 29:104–47.

Rothman, A. J. 2012. Positive definite estimators of large covariance matrices. *Biometrika* 99:733–40.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 58:267–88.

Tibshirani, R., J. Bien, J. Friedman, T. Hastie, N. Simon, J. Taylor, and R. J. Tibshirani. 2012. Strong rules for discarding predictors in lasso-type problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74:245–66.

Tu, J., and G. Zhou. 2011. Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics* 99:204–15.

Uppal, R., and P. Zaffaroni. 2016. Portfolio choice with model misspecification: A foundation for alpha and beta portfolios. Working Paper.