High-Dimensional Portfolio Selection with Cardinality Constraints

Jin-Hong Du, Yifeng Guo, and Xueqin Wang

ABSTRACT
The expanding number of assets offers more opportunities for investors but poses new challenges for modern portfolio management (PM). As a central plank of PM, portfolio selection by expected utility maximization (EUM) faces uncontrollable estimation and optimization errors in ultrahigh-dimensional scenarios. Past strategies for high-dimensional PM mainly concern only large-cap companies and select many stocks, making PM impractical. We propose a sample-average-approximation-based portfolio strategy to tackle the difficulties above with cardinality constraints. Our strategy bypasses the estimation of mean and covariance, the Chinese walls in high-dimensional scenarios. Empirical results on S&P 500 and Russell 2000 show that an appropriate number of carefully chosen assets leads to better out-of-sample mean-variance efficiency. On Russell 2000, our best portfolio profits as much as the equally weighted portfolio but reduces the maximum drawdown and the average number of assets by 10% and 90%, respectively. The flexibility and the stability of incorporating factor signals for augmenting out-of-sample performances are also demonstrated. Our strategy balances the tradeoff among the return, the risk, and the number of assets with cardinality constraints. Therefore, we provide a theoretically sound and computationally efficient strategy to make PM practical in the growing global financial market. Supplementary materials for this article are available online.

1. Introduction
In modern portfolio management (PM), the number of assets grows with economic development and globalization. Taking the United States as an example, there are currently more than 3500 stocks with a capitalization of over 48 trillion dollars, not to mention the financial assets across different countries in global indices. Therefore, the investors face considerable challenges in building feasible strategies to gain substantial returns with controllable risks. Nevertheless, the traditional PM approaches cannot handle such high dimensional problems as the number of assets nowadays can be in the same order or even more extensive than the sample size, bringing both theoretical and empirical challenges to PM.

As a widely used portfolio strategy, the mean-variance (MV) framework introduced by Markowitz (1952) aims at approximating the expected utility maximization (EUM) problems, which laid the foundation of modern portfolio theory. The modern portfolio theory emphasizes the diversified portfolio construction with a tradeoff between return and risk. Although MV strategy provides reasonable estimates of expected utility in his more recent paper (Markowitz 2014), its limitations are also documented in lots of literature, see Michaud (1989), Best and Grauer (1991), Kan and Zhou (2007), and Ao, Yingying, and Zheng (2019). More specifically, as the number of assets increases, the resulting “plug-in” portfolio is susceptible to the estimation error and input uncertainty during the optimization process (F fabozzi et al. 2007) since the expected mean and covariance are hard to be estimated from samples in the high-dimensional scenarios. In contrast, directly maximizing expected utility needs the exact information of distribution which is more intractable in practice.

In addition to the above theoretical challenges, the computational challenges hinder the exploration of large datasets. As the S&P 500 index is the leading U.S. equities benchmark and usually outperforms other indices, it has been used to test high-dimensional PM approaches, see, for example, Fan, Zhang, and Yu (2012), Hautsch and Voigt (2019), Ao, Yingying, and Zheng (2019), Kremer et al. (2020), and Ding, Li, and Zheng (2021). However, whether these approaches can effectively handle datasets with much more stocks remains unknown. Considering mid- and small-cap companies that may offer the investors more opportunities, Pun and Wong (2019) extend the analysis to the incomplete Russell 2000 index constituents where the stocks with volatility greater than 0.04 are discarded, involving up to 1420 stocks in which most existing PM approaches fail. The optimality of their methods can only be established when some technical assumptions are satisfied. Also, their methods suffer from the instability of covariance estimation, which leads to a high turnover rate and transaction fee in their cases.

In reality, any rational investor may choose a reasonable number of assets to hold when considering the transaction
fees, the management fees, the budget constraint, and even the mental cost of looking into too many assets (Gao and Li 2013). The portfolio strategies must be feasible so the investors can efficiently operate and manage the portfolios. On the other hand, it is crucial to maintain portfolio diversification and reduce overall risk profiles by selecting a suitable number of assets. Such a number may depend on the correlations among their prices (Fieldsend, Matatko, and Peng 2004). Overall, it is reasonable to impose cardinality constraints on EUM. However, the high dimensionality of the assets also brings computational challenges of existing methods on solving cardinality-constrained problems, which are generally NP-hard because a combinatorial number of searches is required to obtain exact solutions (Zhang et al. 2019), though many relaxation methods are proposed to approximate it within a reasonable time (Chang et al. 2000; Fieldsend, Matatko, and Peng 2004; Gao and Li 2013).

As the global financial market gets expanding, efficient portfolio strategies for high-dimensional PM are of practical relevance. For example, the Russell 2000 dataset containing around 1900 stocks is more challenging and less predictable because of its inherently ultrahigh dimensionality, higher volatility and lower signal-to-noise ratio, which will be further investigated in Section 2. Faced with the above theoretical and empirical challenges for the ultrahigh-dimensional PM, we aim to answer the following question:

**How can we design an efficient algorithm to deal with the tradeoff among the return, the risk, and the number of holding assets for high-dimensional portfolio management?**

To this end, we design a general portfolio strategy induced from the sample average approximation (SAA) of the EUM problem with cardinality and probability simplex constraints for a variety of risk-averse utility functions. Compared with MV-based approaches, our proposed strategy avoids the estimation error for moments, making it applicable in high-dimensional scenarios. We convert the original problem to a $l_1$-regularized SAA problem with theoretical guarantees and propose an efficient proximal algorithm with acceleration by safe screening techniques. Our strategy is effective and efficient even when the number of assets is far more than the sample size. In the meantime, the investors easily identify just a few valuable assets that can form diversified portfolios. Last but not least, our strategy highlights new perspectives that an appropriate number of carefully chosen assets leads to better out-of-sample mean-variance efficiency from extensive case studies.

### 2. Data Characteristics

In this article, we consider the following two well-known stock market indices. (i) **S&P 500**. As one of the most widely followed equity indices, the S&P 500 measures the performance of 500 large companies that are publicly traded. We collect daily data from the CRSP database, including the open prices, the close prices, and the (holding period) returns from 2010 to 2020. Our stock pool is formed by: (a) Initialization with the S&P 500 index components listed from 2010 to 2020; (b) If mergers and acquisitions (M&A) occur from 2010 to 2020, then the records before and after M&A are concatenated to form only one stock, and the return on the date M&A happened is filled with the growth rate on that date, which results in a pool of 714 stocks in total; (c) For each rebalancing date, we only consider stocks that were listed in the S&P 500 index by that date and had complete records of historical data in the previous year. (ii) **Russell 2000**. Rather than focusing on large-cap companies like the S&P 500, the Russell 2000 index seeks to represent the overall performance of small- and mid-cap companies on the stock market in the United States. Therefore, the smallest 2000 stocks in the Russell 3000 index are included to form the Russell 2000 index. First, we select a group of 5555 stocks listed in Russell 2000 from 2004 to 2020 following the same procedure mentioned above and collect daily stock data accordingly. As discussed earlier, most high-dimensional portfolio selection literature involves empirical studies on the S&P 500 index, while only a few analyze their strategies on Russell 2000. Dealing with the latter requires computational efficiency and numerical stability on high-dimensional datasets, which is more challenging.

The differences between the two datasets are illustrated in Figure 1. It is evident that the prices in the S&P 500 dataset exhibit higher correlations than prices in the other dataset. To maintain a more diversified portfolio, one needs to include more assets in Russell 2000 since the available stocks are less correlated. Another difficulty for portfolio management arising in the high-dimensional scenarios is the increasing volatility. Although the high volatility may offer more opportunities for day traders by large swings, it hinders steady gains over holding periods. In general, stocks of companies with smaller market capitalization will exhibit higher volatility, as shown by the histograms in Figure 1(b). The distribution of the volatility of the Russell 2000 dataset has heavier tails than the one of the S&P 500 dataset. The heavy-tailedness also hinders the estimation of the sample covariance matrix, especially when the number of assets can be larger than the number of observations (Ke et al. 2019). Overall, the lower correlations and higher volatility make it tricky to identify valuable assets from large portfolios.

As a result, data-driven portfolio strategies will include lots of assets to reduce potential risks, which sacrifices the out-of-sample mean-variance efficiency and is also impractical for high-dimensional PM. One natural strategy to tackle the previous issues is holding a reasonable number of assets while maintaining diversified portfolios.

With such motivation, our portfolio strategy aims to impose cardinality constraints on the number of holding assets while producing satisfying mean-variance efficiency. Our approach provides new perspectives to the return-risk tradeoffs, where an appropriate selection of $s$ assets can inform us of a better tradeoff between return and risk with more stability. As summarized in Figure 2, each point represents one portfolio with a different value of regularization parameters, demonstrating the tradeoff between the three factors—risk, return, and $s$ of both the MV approach and our proposed method with logarithm utility. The asset pool consists of 3680 stocks based on NYSE data (See Section 4.2). The left panel of Figure 2 is selected from the in-sample efficient frontier of the MV approach. The MV approach and our method show a similar pattern: the better tradeoff can be achieved with a relatively small number of assets than the whole asset pool. A reasonable amount of diversification contributes to better out-of-sample mean-variance efficiency in this example, indicating that our strategies are applicable for high-dimensional PM.
Figure 1. Visualization of the datasets.

(a) Heatmaps of correlations among stocks in S&P 500 (left) and Russell 2000 (right).

(b) Proportional histogram of stock volatility in the two datasets.

Figure 2. In-sample returns and risks against the number of assets with cubic regression lines plotted. The shadow regions denote the estimated 95% confidence intervals via bootstrapping. The portfolios are constructed based on NYSE data with 3680 stocks.

3. Statistical Methodology

3.1. Sparsity-Induced Portfolios

Suppose we have \( d \) assets with price relatives \( X_1, \ldots, X_d \) to be managed, that is, random variables \( X_j \)'s are the nonnegative ratios of the price at the end of the day to the price at the beginning of the day for asset \( j \). For the rest of the article, we assume that \( X_j \geq \eta_{\text{min}} > 0 \) and \( \mathbb{E}[X_j] < \infty \) for all \( j \). Let \( X = (X_1, \ldots, X_d)^\top \in \mathbb{R}^d \) be the price relative vector and \( w \) be its portfolio allocation vector. The goal of portfolio selection is to maximize the expected utility of wealth:

\[
\min_{w \in C} -\mathbb{E}[u(w^\top X)]
\]

where the expectation is taken with respect to \( X \) and \( C \subseteq \mathbb{R}^d \) is a feasible set. The usual choice of \( C \) for portfolio selections can be: (a) The subspace \( \mathbb{R}_+ \) or \( \mathbb{R} \); (b) The simplex of dimension \( d - 1 \), \( S_d = \{ w \in \mathbb{R}^d : w^\top 1 = 1 \} \); (c) The probability simplex of dimension \( d - 1 \), \( \Delta_d = \{ w \in \mathbb{R}^d : w \succeq 0, w^\top 1 = 1 \} \). For the first choice of \( C \), extra norm constraints or regularizers on \( w \) may be needed to guarantee that the entries of solutions are all finite. We will mainly focus on the case when \( C = \Delta_d \). In the expected utility theory, the utility function \( u \) quantitatively
measures the outcome value to an investor, and some common utility functions are listed in Table 1. We formally define it as follows.

**Definition 1 (Utility Functions).** A function \( u : \mathbb{R}_+ \rightarrow \mathbb{R} \) is a utility function if \( u \) is increasing and concave on its domain.

For the following section, we use \( f \) to denote a general objective function. To impose a sparsity structure on the estimated portfolio allocations, one would add a regularization term \( \Omega \), resulting in an optimization problem:

\[
\min_{w \in C} f(w) + \lambda \Omega(w),
\]

where \( \lambda > 0 \) is a regularization parameter that controls the sparsity of the allocation and can be picked from the Lasso-type path, and \( \Omega \) is a sparsity-induced norm. Alternatively, one can consider the constrained problem:

\[
\min_{w \in C} f(w) \quad \text{s.t.} \quad \Omega(w) \leq s,
\]

where \( s > 0 \) is the threshold parameter. The two formulations are mathematically equivalent when the objective function and \( \Omega \) are convex to \( w \), and the global minimum of the constrained problem is strictly feasible. Let \( \Sigma_d = \{ w \in \mathbb{R}^d : \|w\|_0 \leq s \} \). Considering the cardinality constraints with \( C = \Delta_d \), we formulate the problem as

\[
\min_{w \in \Delta_d \cap \Sigma_d} f(w),
\]

which is unfortunately NP-hard even if the expectation can be evaluated efficiently (Zhang et al. 2019). When \( f \) is a quadratic function for a mean-variance portfolio, the problem can be cast into a mixed-integer quadratic program (MIQP) by introducing \( d \) extra variables. However, it is still computationally intractable and memory-consuming when \( d \) is large and the covariance matrix is dense. Though many convex relaxation algorithms are proposed for MIQPs, numerical results show that they may provide poor lower bounds and even be incapable of solving medium-sized problems within a reasonable time. For example, it may take up to 40 min to run only one instance when \( d = 458 \) and \( s = 7 \) (Zheng, Sun, and Li 2014). One may be interested in whether we can first solve the \( l_1 \)-regularized problem (5):

\[
\min_{w \in \mathbb{R}^d_+} f(w) + \lambda \|w\|_1,
\]

with the non-negativity constraint and then project the estimated allocation onto the \( l_1 \)-norm unit sphere. Such an intuitive approach indeed works under mild assumptions. To see this, we first introduce the single crossing property, formally defined as:

**Definition 2 (Single Crossing).** Suppose that \( Z_1 \) and \( Z_2 \) are two random variables. We say that \( Z_1 \) single crosses \( Z_2 \) from below if there is a crossing point \( c \in \mathbb{R} \) such that

\[
\begin{align*}
F_{Z_1}(t) &\geq F_{Z_2}(t), & t \leq c \\
F_{Z_1}(t) &\leq F_{Z_2}(t), & t > c,
\end{align*}
\]

where \( F_{Z_1}(t) = \mathbb{P}(Z_1 > t) \) and \( F_{Z_2}(t) = \mathbb{P}(Z_2 > t) \) are the survival functions of \( Z_1 \) and \( Z_2 \).

The single crossing property is widely used in the economics literature; see, for example, Jewitt (1987), Athey (2001), and Ballester, Apesteguia, and Lu (2017). This sign change property holds when (a) one of the two decisions leads to a degenerate outcome, such as in the comparative statics problems, or (b) the decisions are made from some distribution family, such as in the single risky asset portfolio problem. An important implication of single crossing is that more risk-averse investors, in the sense of Arrow and Pratt (Arrow 1971), will hold a larger share of their wealth of the safe asset than the single risky asset (Jewitt 1987). From the statistical perspective, the single crossing property holds when two random variables come from common distributions, such as normal, lognormal, and truncated normal, under different conditions, as discussed in (Levy and Robinson 2006, chap. 6). An intuitive example is given in Figure 3 when the two univariate random variables \( Z_1 = w_1^\top \lambda' \) and \( Z_2 = w_2^\top \lambda' \) are normally distributed. We note that this is only a particular case, and no explicit distribution for \( \lambda' \) is assumed in this article. When the decisions \( w_1 \) and \( w_2 \) are estimated from data, \( Z_1 \) and \( Z_2 \) are from two families of distribution functions, with each corresponding pair of realizations crossing no more than once, resulting in the preservation of the risk aversion preference. The following theorem shows that one can solve the \( l_0 \) constrained problem (4), by solving the \( l_1 \) regularized problem (5) under such conditions.

**Theorem 1 (Equivalence for Expected Utility Maximization).** Suppose that the objective function in both optimization problems (5) and (4) is \( f(w) = -\mathbb{E}[u(w^\top \lambda')] \) with \( \lambda'_j \geq \min_{\lambda_j} > 0 \) and \( \mathbb{E}[\lambda'_j] < \infty \) for all \( j \). Let \( w_1 \) and \( w_2 \) be the solutions to problems (5) and (4), respectively with \( s = \|w_1\|_0 > 0 \), and define \( \tilde{w}_1 = w_1/\|w_1\|_1, \tilde{w}_2 = w_2/\|w_1\|_1, Z_i = w_i^\top \lambda' \) and \( \tilde{Z}_i = \tilde{w}_i^\top \lambda' \) for \( i = 1, 2 \). Suppose that (i) \( u \) is strictly increasing and strictly concave, and (ii) \( \tilde{Z}_1 \) and \( \tilde{Z}_2 \) cross for at most one time, then \( \tilde{w}_1 \) is an solution to problem (4).

The single crossing property in Theorem 1 states that the better portfolio (in terms of linear utility) can be distinguishable at different levels of random (linear) utility. Furthermore, an indifference curve of a better portfolio (in the space of utility levels) crosses that of the other portfolio at most once. This assumption captures the idea that a better portfolio is preferred for more random utility. As a result, it is more likely to profit through investing in the better portfolio at higher levels of random utility, thus, making it possible to separate the two portfolios. In the proof of Theorem 1, we also show that such preferences can be preserved for other utility functions. With such preferences, Theorem 1 guarantees that the two portfolios \( \tilde{w}_1 \) and \( w_2 \) obtain the same expected utility under constraints of (4).
3.2. Sample Average Approximation

Since the optimization over expectation is generally intractable and impractical, a natural way to solve the optimization problem (1) is to minimize the empirical sample average objective:

$$\min_{w \in \mathcal{C}} h(w) = -\frac{1}{n} \sum_{i=1}^{n} u(w^\top X_i),$$

(6)

where $X_i$'s are independent and identically distributed as $\mathcal{X}$, which is known as the sample average approximation (SAA) method in the stochastic optimization literature (Shapiro, Dentcheva, and Ruszczynski 2014). The mean-variance portfolio problems are generally solved implicitly by the SAA method, where the mean and variance of $\mathcal{X}$ are estimated from samples. One of the connections between the utility theory and the MV portfolio theory is that the latter’s objective is the utility function’s moment approximation (MA). Define $U_w(x) = u(w^\top x)$ and consider the second-order approximation of $-\mathbb{E}[U_w(X)]$:

$$g_{\delta_0}(w, \mu, \Sigma) = -U_w(x_0) - \nabla_x U_w(x_0)^\top (\mu - x_0) - \frac{1}{2} \text{tr} \left( \nabla_x^2 U_w(x_0) (\Sigma + \mu\mu^\top) \right) + x_0^\top \nabla_x U_w(x_0) \mu - \frac{1}{2} x_0^\top \nabla_x^2 U_w(x_0) x_0,$$

(7)

where $\mu = \mathbb{E}(\mathcal{X})$ and $\Sigma = \text{var}(\mathcal{X})$. If $x_0 = \mu$ and the utility function is quadratic, then Equation (7) turns out to be objective of MV portfolios and the MA methods aim to solve:

$$\min_{w \in \mathcal{C}} g_{\delta_0}(w, \hat{\mu}, \hat{\Sigma}),$$

(8)

where $\hat{\mu}$ and $\hat{\Sigma}$ based on samples $X_1, \ldots, X_n$ of $\mathcal{X}$. Solutions to the problem (8) also refer to the plug-in estimators of $w$ (Cai et al. 2020; Ding, Li, and Zheng 2021). While such an approach is justified by classical statistics theory because the plug-in portfolio is an MLE of the optimal portfolio, its out-of-sample performance is generally poor (Ao, Yingying, and Zheng 2019).

Combining optimization problem (6) with regularization, we next derive the nonasymptotic error bound for coefficients estimation in the ultra-high-dimensional settings. To begin with, we consider the constrained variants of population and empirical minimization problems:

$$\min_{w \in \mathbb{R}^d_+} \mathbb{E}[h(w)] \quad \text{subject to} \quad \|w\|_1 \leq R,$$

(9)

$$\min_{w \in \mathbb{R}^d_+} h(w) + \lambda \|w\|_1 \quad \text{subject to} \quad \|w\|_1 \leq R,$$

(10)

The constraint in the above problems guarantees that the estimator lies in a bounded set, which is helpful for our statistical analysis. One can also view problem (5) with objective functions $f(w)$ being $\mathbb{E}[h_w(w)]$ as the Lagrangian formulation of the problem (9). For problem (10), the constraint $\|w\|_1 \leq R$ will have no effect when the regularization parameter $\lambda$ is large enough. We denote the solutions to the optimization problems (9) and (10) by $w^*$ and $\hat{w}_n$, respectively. Before we present our results, we first introduce some matrix notations. We use $X = [X_1, \ldots, X_n]^\top \in \mathbb{R}^{n \times d}$ to denote the sample matrix and $X_j = [X_{ij}, \ldots, X_{nj}]^\top \in \mathbb{R}^n$ to denote its $j$th column for $j = 1, \ldots, d$. We make the following assumptions:

**Assumption 1.** We assume the following holds:

(a) The utility function $u$ is $L_u$-Lipschitz continuous, differentiable, and concave.

(b) The theoretical minimizer $w^*$ is unique and denote its cardinality by $s = \|w^*\|$.

(c) There exists $\nu > 0$ such that $(X_{ij} - \nu)$’s are sub-Gaussian with parameter $\sigma$, $\mathbb{E}[(\exp(t(X_{ij} - \nu))) \leq \exp(t^2\sigma^2/2), \forall t > 0, i \in [n], j \in [d]$. ($\sigma$) (Restricted strong convexity) There exists $\sigma > 2$ and $\alpha$ such that almost surely

$$0 < \kappa \leq \inf_{|S| \geq s} \inf_{w \in C(S, \gamma_1, \gamma_2)} \frac{\nu^\top \nabla^2_w \mathbb{E}[h(w^*)] \nu}{\|w\|_2^2},$$

where $\gamma_1 = \frac{\alpha}{\nu - \gamma_1}$, $\gamma_2 = \frac{\alpha}{\nu - \gamma_2}$, and $C(S, \gamma_1, \gamma_2) = \{v \in \mathbb{R}^d | \|v_{S^c} \|_1 \leq \gamma_1 \|v_S \|_1 + \gamma_2 \|v_S \|_2\}$.

From Definition 1, the concavity in Assumption 1(a) naturally holds for all utility functions. Most of the utility functions in Table 1 such as exponential utility and logarithmic utility with $\eta > 0$, are Lipschitz-continuous and differentiable. Then it also follows that $H(z) = -\sum_{i=1}^n u(z_i)/n$ is $L_z$-Lipschitz continuous with $L_z = L_u/n$. Assumption 1(c) is commonly used in the literature of nonasymptotic high-dimensional statistics (Wainwright 2019). Especially, if $\mathcal{X}$ is a bounded set, then Hoefding's lemma guarantees that Assumption 1(c) holds. Assumption 1(d) presumes good curvature for the population cost function within certain region of the parameter space. It is weaker than the one used in (Wainwright 2019, sec. 9.8), since we only need the restricted strong convexity property to hold in the intersection of the unit $l_2$-norm ball and a cone. We present our main results as follows.

**Theorem 2 (Coefficients Estimation Error).** Suppose Assumption 1 holds. For any $\delta \in (0, 1)$ and $\lambda \geq \alpha(\nu + 1) L_u \sigma$...
\[ \sqrt{\log(d/\delta)/n}, \text{with probability at least } 1 - \delta, \text{it holds that} \]

\[ \| \hat{w}_n - w^* \|_2 \lesssim O(\kappa^{-1}(v \lor 1)L_0\sigma\sqrt{\log(d/\delta)/n}). \]

Theorem 2 says that if the sparsity level and the dimension satisfy that \( s \log d = o(n) \), the coefficients estimation error vanishes when both \( n \) and \( d \) go to infinity. Our error rate \( O(\sqrt{s \log d}/n) \) also matches the rate of Lasso estimator for linear regression. As an immediate consequence of Theorem 2 by Cauchy-Schwartz inequality, the portfolio performances of the SAA solutions also have finite-sample error bounds as stated in Corollary 3.

**Corollary 3 (SAA Performance).** Under the same conditions as in Theorem 2, with probability at least \( 1 - \delta \), it holds that

\[ \| \hat{w}_n - w^* \|_2 \lesssim O(\| \mu \|_2\kappa^{-1}(v \lor 1)L_0\sigma\sqrt{\log(d/\delta)/n}) \]

\[ \| \hat{w}_n \Sigma - w^* \Sigma w^* \|_2 \lesssim O(R(\| \Sigma \|_2\kappa^{-1}(v \lor 1)L_0\sigma\sqrt{\log(d/\delta)/n})). \]

Combining the equivalence of \( l_0 \) constraints and \( l_1 \) regularizers with the asymptotic optimality of the latter, solving the \( l_1 \)-regularized SAA problem is provably equivalent to solving the \( l_0 \)-constrained EUM problem. It is beneficial in two ways. First, we no longer deal with expectation as in the EUM problem, which requires the knowledge of the distribution of \( X \). Second, we work on a much easier convex optimization problem instead of a nonconvex problem. Thus, proximal gradient descent (Section 3.4) can be efficiently employed and safe screening (Section 3.3) can be used to filter unimportant assets and speed up the process.

### 3.3. Feature Screening

Turning the nonconvex optimization problem (4) into the convex optimization problem (5) is beneficial for computation. However, it is still computationally challenging when the dimension \( d \) is large. Fortunately, the screening techniques developed for the convex optimization problem can be used to alleviate this problem. Building on the idea of Dantas, Soubies, and Févotte (2021), we propose safe screening methods for the problem (5) and further derive efficient primal updates. We only need the following common assumption about the objective function \( h(w) \) for the remaining context to establish our theoretical results.

**Assumption 2 (Convexity and Smoothness).** The objective function \( h : \mathbb{R}^d \to \mathbb{R} \) is proper convex and has Lipschitz continuous gradients.

Similar to Lasso problems, the maximum regularization parameter \( \lambda_{max} \triangleq \| V h(0, \cdot) \|_\infty \) makes the solution to the optimization problem (5) all zeros. Hence, we would focus on the case when \( \lambda < \lambda_{max} \). The convexity property naturally yields the primal and dual formulations of the convex optimization problem.

**Theorem 4 (Fenchel-Rockafellar Duality).** The primal and dual formulations of the optimization problem (5) is given by

\[ w^* \in \arg \min_{w \in \mathcal{C}_P} P_\lambda (w) \triangleq -\frac{1}{n} \sum_{j=1}^{n} u(w^T X_j) + \lambda \| w \|_1, \]

\[ \theta^* \in \arg \max_{\theta \in \mathcal{C}_D} D_\lambda (\theta) \triangleq -\frac{1}{n} \sum_{j=1}^{n} u^*(\nu \lambda \theta_j), \]

where \( u^* \) is the convex conjugate of \( u \), \( \mathcal{C}_P = \text{dom}(P_\lambda) \cap \mathbb{R}^d \) and \( \mathcal{C}_D = \text{dom}(D_\lambda) \cap \{ \theta \in \mathbb{R}^n \| \phi(X^T \theta) \|_\infty \leq 1 \} \) are the primal and dual feasible sets and, the function \( \phi(x) = \max \{ x, 0 \} \) and inequality relation \( \leq \) are applied element-wisely. For any primal-dual feasible point \( (w^*, \theta^*) \in \mathcal{C}_P \times \mathcal{C}_D \), the optimality condition reads that,

\[ \text{(Primal-Dual Link)} \quad - \lambda \theta^* = \nabla_x H(X w^*), \quad (11) \]

\[ \text{(Subdifferential Inclusion)} \quad \left\{ \begin{array}{l} \| \phi(X^T \theta^*) \|_\infty \leq 1, \quad w^*_j = 0 \smallsetminus 0, \\ \| \phi(X^T \theta^*) \|_\infty = 1, \quad \theta^*_j = 1, \\ \theta^*_j = 1, \quad w^*_j = 0, \text{ with probability at least } 1 - \delta. \end{array} \right. \quad (12) \]

By Theorem 4, we can compute the duality gap \( \text{Gap}_\lambda (w, \theta) \triangleq P_\lambda (w) - D_\lambda (\theta) \), which is nonnegative due to the weak duality property and quantifies the suboptimality of the feasible primal-dual pair for the optimization problem (5). Moreover, the duality gap can be used to construct a safe region that contains the optimal dual variables, as shown in Theorem 5. By (Ndiaye et al., 2017, Theorem 6), \( \mathcal{D}_\delta \) is strongly concave if \( u_i \)'s are differentiable with Lipschitz gradients. Thus, under Assumption 2, the constructed sphere is guaranteed to be safe and called gap safe sphere. To obtain feasible dual variables, we can apply dual scaling \( \Xi(z) = z / \max \{ \| \phi(X^T z) \|_\infty, 1 \} \) to \( \theta^* \) so that a meaningful safe region can be computed.

**Theorem 5 (Safe Region).** For any primal-dual feasible point \( (w, \theta) \in \mathcal{C}_P \times \mathcal{C}_D \), we have that \( \theta^* \in \mathcal{B}(\theta, r) \) if \( \mathcal{D}_\delta \) is \( \alpha \)-strongly concave. Furthermore, the safe screening rule for each feature \( j \in [d] \) is given by \( \| \phi(X^T \theta^*) \|_\infty \leq 1 \rightarrow w^*_j = 0 \).

### 3.4. Algorithm

Our main algorithm is described in Algorithm 1. For any utility function \( u \) satisfying Assumption 2 and data matrix \( X \), we first define the primal and dual objectives as on line 1 of Algorithm 1. We perform the primal and dual updates in each iteration, followed by the gap safe screening. Recall that the proximal operator is defined as \( \text{prox}_\lambda (g(x)) \triangleq \arg \min_v g(v) + \| x - y \|_2^2/2 \) for any function \( g \). On line 4, we perform one-step gradient descent on \( w^{(l)} \) followed by applying the proximal operator of \( \Omega(v) = \| v \|_1 + \mathbb{1}_{\mathcal{A}^{(l)}}(v) \), where \( \mathcal{A}^{(l)} \) is the active set of stocks in iteration \( l \) and \( \mathbb{1}_C(v) \) is the indicator function on a set \( C \) that takes value 0 when \( v \in C \) and infinity otherwise. Noted that the proximal update is not the same as soft-thresholding for Lasso because of the non-negativity constraint. Although there is no closed-form for the proximal operator, we can still evaluate it efficiently, as discussed in supplement materials. On line 5 of Algorithm 1, we compute the dual variable defined in (11) and apply the scaling function \( \Xi(z) = z / \max \{ \| \phi(X^T z) \|_\infty, 1 \} \) and called...
Algorithm 1 Portfolio Selection on $\Delta_d$ with Feature Screening.

Input: The utility function $u$, the relative price matrix $X \in \mathbb{R}^{n \times d}$, the regularization parameter $\lambda$, the step size $\iota$, and the initial value of $w^{(0)}$.

1. Initialize the step number $t = 1$, the screening set $S^{(0)} = \emptyset$ and the active set $A^{(0)} = [d]$. Form the primal and dual objective functions

$$P_\lambda(w) = h(w) + \lambda \|w\|_1, \quad D_\lambda(\theta) = h^*(\theta), \quad \forall (w, \theta) \in \mathcal{C}_P \times \mathcal{C}_D,$$

where $h(w) = H(Xw)$ and the active set $A^{(0)} = [d]$. Form the primal and dual objective functions

2. while not terminate do
3. Allocate memory for $w^{(t)}$ (e.g., modified in place of $w^{(t-1)}$).
4. Primal Update
5. $w^{(t)}_{A^{(t)}} = \text{prox}_{\lambda \Omega}[w^{(t-1)} - \iota \nabla_w h(w^{(t-1)}))]$ where $\Omega(v) = \|v\|_1 + \iota \mathbb{1}_{A^{(t)}}(v)$.
6. Dual Evaluate And Scaling
7. $\theta^{(t)} = \mathbb{E}(-\nabla_w H(Xw^{(t)}/\lambda)$.
8. Screening
9. Gap
10. $r^{(t)} = \sqrt{2 \text{Gap}_{\lambda}^{(t)})/\alpha}$.
11. $S^{(t)} = S^{(t-1)} \cup \{j \in A^{(t-1)} : \phi(X_j^{\lambda,\theta^{(t)}} + r^{(t)}) \|X_j\|_2 < 1\}$.
12. $A^{(t)} = [d] \setminus S^{(t)}$.
13. $w_{S^{(t)}} = 0_{\{S^{(t)}\}}$.
14. $t = t + 1$.

end while

Output: The portfolio allocation $\hat{w}$.

so that the dual variable is feasible. Finally, we perform gap-safe screening and update the set of active coordinates. For the next iteration, we just need to repeat the above process for the subvector $w_{A^{(t)}}$ since any coordinate of $w$ in the complement of $A^{(t)}$ is guaranteed to be zero. Though Algorithm 1 is general and inclusive, we will specialize it to logarithmic utility and exponential utility in Section 4.1, and some valuable ingredients are summarized in supplement materials.

Combining the proximal gradient descent algorithm with the screening rules, the convergence guarantee as in Theorem 6 can also be established for Algorithm 1. This implies that Algorithm 1 has a convergence rate $O(1/t)$ or $O(1/\epsilon)$ where $t$ is the number of iterations and $\epsilon$ is the tolerance. We also noted that it is not necessary to perform screening at each iteration in Algorithm 1. So we further add an option in our implementation to run screening for every $i_j$, iteration, where $i_j$ is a parameter. In general, $i_j$ should not scale linearly with the maximum number of iterations so that the time consumption on screening will not be significant compared to proximal gradient descent steps.

Theorem 6 (Convergence). Suppose that Assumption 2 holds and $w^*$ is defined in Theorem 4. If the step size $\iota \leq 1/L\nabla h$, then the estimate $w^{(t)}$ at the $t$th step of Algorithm 1 satisfies

$$|P_\lambda(w^{(t)}) - P_\lambda(w^*)| \leq \frac{\|w^{(0)} - w^*\|_2^2}{2t}.$$
Figure 4. Simulation results on NYSE historical data when \( n = 24 \) and \( d = 3680 \).

(a) Screening ratio against iterations for \( \log_{10}(\lambda/\lambda_{\text{max}}) \in [-3, 0] \) for LOG and EXP-1.00 utility functions (The lighter, the more screened features).

(b) The relative execution time (the ratio of screened time to unscreened time) against \( \log_{10}(\lambda/\lambda_{\text{max}}) \), and the convergence rate (duality gap) against execution time when \( \lambda/\lambda_{\text{max}} = 5e-1 \), for LOG and EXP-1.00 utility functions.

4.3. Portfolio Selection on S&P 500 Index Constituents

We first investigate the performance of different strategies on the S&P 500 index constituents.

4.3.1. Comparing with Benchmark Portfolios

We compare our methods with the following three types of portfolio strategies: (a) Equally weighted (EW) portfolio. (b) Global minimum variance (GMV) portfolio using sample covariance matrix (GMV-P), the linear shrinkage estimation of the covariance matrix \( \Sigma \) (GMV-LS), and nonlinear shrinkage estimation of \( \Sigma \) (GMV-NLS). (c) Mean-variance (MV) portfolios with the covariance estimators analogously (MV-P, MV-LS, and MV-NLS).

We perform the out-of-sample portfolio evaluation, beginning in 2011 and rebalancing per quarter (63 trading days, denote \( n_{\text{hold}} \)). The new model parameters are estimated from 120 training samples before the rebalancing trade day. The resulting sample matrices are of shape \( 120 \times d \), where \( d \) varies in different periods (median:436, maximum:454), which is naturally challenging for estimation and optimization because of its high dimensionality (the number of features is much larger than the number of samples). We use 5-fold time-series cross-validation to select the hyperparameters, analogous to \( K \)-fold
cross-validation without shuffling but only with training samples prior to the evaluation set in each fold. The regularization parameter $\lambda$ from the set of 100 points spaced evenly on a log scale from $\log_{10}(\lambda_{\text{max}}) - 2$ to $\log_{10}(\lambda_{\text{max}})$ for our models, and $\lambda_{\text{MV}}$ from the set of 100 points uniformly distributed in spaced evenly on a log scale from $-3$ to $2$ for MV portfolios.

For our models, we set the maximum iterations to be $10^4$ and the convergence tolerance to be $10^{-5}$ during cross-validation. After the parameter $\lambda$ is selected from cross-validation, we set them to $10^5$ and $10^{-8}$, respectively. We perform screening every 30 iterations for all the experiments in this section. For GMV and MV portfolios, we use solver Cvxopt (Andersen, Dahl, and Vandenberghe 2021) to solve the corresponding convex optimization problems (GMV Portfolio) and (MV Portfolio) by quadratic programming. The parameters for the solver are specified as abstol=$10^{-12}$, reltol=$10^{-11}$, maxiter=$10^4$, feastol=$10^{-16}$.

We measure the out-of-sample performance of each model by four indicators on the daily returns: (a) the accumulated return (RET),$^1$ measuring how much wealth the portfolio strategy earns; (b) the maximum drawdown (MDD), the worst possible risk the investors may face when adopting the strategy; (c) the annualized Sharpe ratio (SR), describing how much excess returns relative to its volatility the portfolio offers and (d) the annualized Sortino ratio (SoR), which replaces the overall volatility in SR with the downside deviation (Foster and Stutzer 2003).

The results of the S&P 500 datasets are summarized in Table 2. The baseline portfolio strategy EW invests every available asset with the same amount. We see that EW has a return of 2.3231, which is 0.99 more than the excess return of the S&P index (whose open and close prices in the evaluation period are 1257.62 and 3756.07, respectively) because we have filtered stocks not listed in the index or having missing records in the preceding year. However, the EW strategy has the most significant drawdown among all methods, indicating the validity of both MV-based strategies and our methods.

The fine-grained plots of monthly returns show that the impacts of the market on the effectiveness of all strategies are similar. For example, all three methods failed to profit from June to July 2017 and from January to February 2020, when the U.S. stock market faced downside risks. From the plots, we can also see that the LOG strategy is less risk-averse because it gains enormous profits for some months and loses more for others.

In the long run, however, it earns 50% more than MV-NLS in 10 years.

To consider the transaction fees, we define the net returns as $\tilde{c}(t) = \left(1 - \sum_{j=1}^{d} c|w_j^{(t)} - w_j^{(t-1)}| - \sum_{j=1}^{d} \tilde{c} R_{|w_j^{(t)}| < s} \right) (1 + r(t)) - 1$, where $w_j^{(t)}$ and $r(t)$ are the allocation vector and the return during the $t$th trading period, and the constant rate $c$ is chosen to be 0.1% as suggested by Robert, Robert, and Jeffrey (2012) and $\tilde{c}$ is chosen to be 0.001% to account for trading activity fees. As shown in Table 3, the LOG strategy and most EXP strategies outperform benchmark methods when considering transaction fees.

| Method      | Return  | Maximum drawdown | Sharpe ratio | Sortino ratio | Avg. Num. of assets |
|-------------|---------|------------------|--------------|---------------|---------------------|
| Benchmark   |         |                  |              |               |                     |
| EW          | 2.3231  | 0.3941           | 0.7084       | 0.9848        | 437                 |
| GMV-P       | 2.1061  | 0.2483           | 0.9487       | 1.4924        | 281                 |
| GMV-LS      | 2.4796  | 0.2929           | 0.9800       | 1.4648        | 258                 |
| GMV-NLS     | 2.3775  | 0.3223           | 0.9410       | 1.3577        | 323                 |
| MV-P        | 1.5481  | 0.2343           | 0.5287       | 0.7654        | 211                 |
| MV-LS       | 1.9540  | 0.2363           | 0.5981       | 0.8735        | 170                 |
| MV-NLS      | 4.0390  | 0.2341           | 0.8681       | 1.2980        | 118                 |

| Our methods | Return  | Maximum drawdown | Sharpe ratio | Sortino ratio | Avg. Num. of assets |
|-------------|---------|------------------|--------------|---------------|---------------------|
| LOG-1.00    | 8.9922  | 0.3194           | 0.9953       | 1.5743        | 20                  |
| EXP-0.05    | 6.2313  | 0.3504           | 0.8310       | 1.4563        | 37                  |
| EXP-0.10    | 5.7691  | 0.3027           | 0.8729       | 1.3894        | 23                  |
| EXP-0.50    | 8.5086  | 0.3558           | 0.9509       | 1.5421        | 7                   |
| EXP-1.00    | 8.7293  | 0.3558           | 0.9777       | 1.5679        | 6                   |
| EXP-1.50    | 8.1633  | 0.3558           | 0.9637       | 1.5077        | 5                   |

4.3.2. Portfolio with Cardinality Constraints

For real-life asset allocations, investors have to face more significant systematic risks to have potentially higher returns, but they can reduce the exposure to unsystematic risks by diversifying their investments. While there is no single answer to what is a reasonable number, one can always explore possible numbers through historical data. We evaluate our methods on the same dataset in Section 4.3.1 but with restrictions on the asset allocations. Formally, we aim to solve the problem (4) given different numbers of assets $s$. Due to the equivalence relationship, we can portray the Lasso-type path efficiently for the problem (5).

Figure 6 shows the portfolio results constrained on the maximum number of assets. We can see that both logarithm and exponential utility functions have similar performance when restricting the maximum number of assets to be invested. For most strategies, the annualized SRs are more than 0.9 when $s$ is larger than 20 and become relatively stable as $s$ further increases. From the plots of the average number of assets, we can also see

---

$^1$We obtain the risk-free rate from the Kenneth French Data Library to compute excess returns.
Figure 5. Annual and monthly returns on S&P 500 from 2011 to 2020 with \( n = 120, d \leq 454, n_{\text{hold}} = 63 \) for MV-NLS, LOG and EXP-1.00. The dash lines denote the mean annual returns.

Table 3. Out-of-sample results (with transaction fees) on S&P 500 from 2011 to 2020.

| Method | Return | Maximum drawdown | Sharpe ratio | Sortino ratio | Avg. Num. of assets |
|--------|--------|------------------|--------------|---------------|---------------------|
| Benchmark | | | | | |
| EW | 1.8648 | 0.3941 | 0.6283 | 0.8708 | 437 |
| GMV-P | 1.5428 | 0.2483 | 0.7805 | 1.2172 | 281 |
| GMV-LS | 1.8424 | 0.2929 | 0.8220 | 1.2208 | 258 |
| GMV-NLS | 1.7499 | 0.3223 | 0.7838 | 1.1240 | 323 |
| MV-P | 1.0943 | 0.2415 | 0.4336 | 0.6252 | 211 |
| MV-LS | 1.4417 | 0.2550 | 0.5062 | 0.7363 | 170 |
| MV-NLS | 3.3271 | 0.2341 | 0.7924 | 1.1809 | 118 |

| Our methods | | | | | |
| LOG-1.00 | 8.1604 | 0.3240 | 0.9613 | 1.5183 | 20 |
| EXP-0.05 | 5.5768 | 0.3626 | 0.7961 | 1.3927 | 37 |
| EXP-0.10 | 5.2082 | 0.3073 | 0.8375 | 1.3311 | 23 |
| EXP-0.50 | 7.7953 | 0.3558 | 0.9215 | 1.4922 | 7 |
| EXP-1.00 | 8.0122 | 0.3558 | 0.9481 | 1.5184 | 6 |
| EXP-1.50 | 7.4773 | 0.3558 | 0.9333 | 1.4584 | 5 |

Table 4. Out-of-sample results (without transaction fees) on Russell 2000 from 2005 to 2020.

| Method | Return | Maximum drawdown | Sharpe ratio | Sortino ratio | Avg. Num. of assets |
|--------|--------|------------------|--------------|---------------|---------------------|
| Benchmark | | | | | |
| EW | 3.1023 | 0.6125 | 0.4176 | 0.5948 | 1640 |
| GMV-P | 1.2725 | 0.2677 | 0.4122 | 0.5948 | 838 |
| GMV-LS | 1.7997 | 0.3725 | 0.4402 | 0.6251 | 877 |
| GMV-NLS | 1.5940 | 0.4502 | 0.3687 | 0.5247 | 820 |
| MV-P | -0.7531 | 0.9398 | -0.0408 | -0.0554 | 302 |
| MV-LS | -0.8415 | 0.9530 | -0.0879 | -0.1195 | 351 |
| MV-NLS | -0.8987 | 0.9722 | -0.1180 | -0.1590 | 406 |

| Our methods | | | | | |
| LOG | 3.4483 | 0.5473 | 0.4514 | 0.6330 | 160 |
| EXP-0.05 | 1.8137 | 0.4942 | 0.3738 | 0.5135 | 307 |
| EXP-0.10 | 1.9293 | 0.4940 | 0.3866 | 0.5313 | 332 |
| EXP-0.50 | 1.9791 | 0.4868 | 0.3953 | 0.5432 | 314 |
| EXP-1.00 | 1.9514 | 0.4830 | 0.3929 | 0.5397 | 308 |
| EXP-1.50 | 2.0456 | 0.4753 | 0.4076 | 0.5602 | 255 |

4.4. Portfolio Selection on Russell 2000 Index Constituents

When faced with larger stock pools and longer time horizons, we analyze our portfolio techniques in this section to see how they fare. We execute the out-of-sample portfolio evaluation in the same way as in Section 4.3. The subpool of stocks with complete data in the preceding half year at the beginning of January, April, July, and October is selected. The median and the maximum value of the numbers of stocks in the pool are 1626 and 1804, respectively. The allocation is then determined using a 5-fold time-series cross-validation from the prior half year, and the assets are held until the first trade day of the next quarter.

Such a large portfolio problem is very challenging in practice. What is worse, data-driven methods based on historical data tend to produce highly sparse allocations, especially when facing the appearance of outliers. It happens more commonly in small-cap stocks, which fluctuate much more violently than large-cap stocks. To deal with this issue, we clip the stock relative prices within the top and bottom 2.5% quantiles of the overall empirical distribution to exclude extreme values, which we find effective for all algorithms. We restrict the number of selected assets to be no less than a prespecified threshold \( n_{\text{min}} = 100 \).

Their out-of-sample performances are again compared with benchmark portfolios as in Section 4.3.1. As shown in Table 4, the MV methods fail to profit for the long run, and the EW and the GMV methods obtain good mean-variance efficiency.

how the risk-aversion parameter of the exponential utility influences the portfolio allocations. As the risk-aversion parameter increases, the portfolio strategy tends to put fewer investments on risky assets, and thus fewer assets will be included when the regularization weakens.
It coincides with DeMiguel, Garlappi, and Uppal (2009) that it is hard to beat the EW portfolio in the long run. Our methods are on par with GMV methods, while we only require half of the assets on average. Furthermore, our methods excel the GMV methods when the transaction fees are considered (Table S2) because the overall composition of the GMV portfolios between rebalancing periods is changing dramatically, indicating that choosing a suitable number of assets is beneficial.

We further explore augmenting portfolio performance with factor signals following Ledoit and Wolf (2017). The factor signals are exposure to the stocks and may imply the research on efficient market as it enhances to explain cross-sectional stock returns. By now, many factors are documented in academia or constructed by the industry. We examine two simple factors, the moving average Sharpe ratio and the moving average relative strength index (RSI). The definitions of the two factors are detailed in supplement materials. The results are summarized in Tables 5 and S4 (without transaction fees) and Tables S3 and S5 (with transaction fees). We observed that the SRs and SoRs of the GMV portfolios considerably deteriorate because they are only concerned about minimizing the variation of the factor of the selected assets. On the contrary, the MV portfolios and our methods gain significant improvements compared with portfolios without factors. In most cases, our methods outperform the EW strategy containing 1640 stocks on average in terms of all metrics.

In addition to the portfolio performance, we are also interested in the time consumption of MV portfolios and our method when the asset number is much larger than the sample number. We choose the period when 1615 stocks are included in the training set of 128 samples. It takes about 60 sec each fold each \( \lambda_{MV} \) for MV-P to complete 1000 iterations on average,
with the desired number of holding assets. As shown in Figure 7, the ultra-high-dimensional settings.

iterations, showing computational benefits of our methods in different risk tolerance preferences can decide suitable choices more in terms of SRs and SoRs. Hence, different investors with more assets help minimize risks but do not necessarily benefit also controls the maximum number of assets to hold. Overall, seen in Section 4.3.2. For EXP portfolios, the aversion parameter sample performances of our methods have similar patterns as when imposing different cardinality constraints, the out-of-sample results (without transaction fees) on Russell 2000 with SR

| Method    | Return | Maximum drawdown | Sharpe ratio | Sortino ratio | Avg. Num. of assets |
|-----------|--------|------------------|--------------|---------------|---------------------|
| Benchmark |        |                  |              |               |                     |
| EW        | 3.1023 | 0.6125           | 0.4176       | 0.5948        | 1640                |
| GMV-P     | 0.1927 | 0.7264           | 0.1324       | 0.1925        | 813                 |
| GMV-LS    | 0.6516 | 0.6640           | 0.2030       | 0.2906        | 806                 |
| GMV-NLS   | 0.4707 | 0.6798           | 0.1755       | 0.2518        | 657                 |
| MV-P      | 0.3664 | 0.5924           | 0.1728       | 0.2480        | 243                 |
| MV-LS     | 0.6889 | 0.5837           | 0.2202       | 0.3150        | 275                 |
| MV-NLS    | 0.5009 | 0.5765           | 0.1969       | 0.2819        | 330                 |
| EXP-0.00  | 4.2410 | 0.5187           | 0.5009       | 0.7087        | 124                 |
| EXP-0.05  | 3.5765 | 0.5331           | 0.4678       | 0.6609        | 277                 |
| EXP-0.10  | 3.5342 | 0.5383           | 0.4656       | 0.6577        | 284                 |
| EXP-0.50  | 3.0583 | 0.5511           | 0.4382       | 0.6168        | 202                 |
| EXP-1.00  | 2.9818 | 0.5472           | 0.4332       | 0.6098        | 298                 |
| EXP-1.50  | 3.0385 | 0.5339           | 0.4365       | 0.6141        | 245                 |
| Our methods |       |                  |              |               |                     |
| LOG       | 4.2410 | 0.5187           | 0.5009       | 0.7087        | 124                 |
| EXP-0.00  | 3.5765 | 0.5331           | 0.4678       | 0.6609        | 277                 |
| EXP-0.10  | 3.5342 | 0.5383           | 0.4656       | 0.6577        | 284                 |
| EXP-0.50  | 3.0583 | 0.5511           | 0.4382       | 0.6168        | 202                 |
| EXP-1.00  | 2.9818 | 0.5472           | 0.4332       | 0.6098        | 298                 |
| EXP-1.50  | 3.0385 | 0.5339           | 0.4365       | 0.6141        | 245                 |

while only 5 sec each fold each λ for LOG to complete 10,000 iterations, showing computational benefits of our methods in the ultra-high dimensional settings.

Finally, our algorithm is also flexible for producing portfolios with the desired number of holding assets. As shown in Figure 7, when imposing different cardinality constraints, the out-of-sample performances of our methods have similar patterns as seen in Section 4.3.2. For EXP portfolios, the aversion parameter also controls the maximum number of assets to hold. Overall, more assets help minimize risks but do not necessarily benefit more in terms of SRs and SoRs. Hence, different investors with different risk tolerance preferences can decide suitable choices of s for their own based on historical data.

5. Conclusion

Strategies for high-dimensional portfolio management are becoming increasingly important as the global financial market expands. This portfolio management study provides new perspectives on the mean-variance efficiency, where an appropriate number of holding assets can facilitate a better tradeoff between return and risk with more stability. When the number of holding assets is chosen in a purely data-driven manner based on historical data, our portfolio strategies are on par with the best benchmark GMV portfolio strategies in Sharpe ratios and Sortino ratios, while only requiring less than half of assets as them, on both S&P 500 and Russell 2000 datasets. The proposed strategies significantly outperform all benchmark methods when considering transaction costs. When the number of holding stocks is fixed in advance, we find that about 10–30 assets can maintain well-diversified portfolios on S&P 500 datasets. At the same time, 160–300 are needed for Russell 2000 datasets, depending on the utility function one uses. To the best of our knowledge, this is the first case study concerning such a large number of assets in the literature on portfolio management. The results indicate that sparsity-induced portfolios will not ruin the diversification of the constrained portfolio with increasing risks. For example, on Russell 2000, our best portfolio profits as much as the EW portfolio but reduces the maximum drawdown and the average number of assets by 10% and 90%, respectively. Finally, the feasibility and stability of the proposed strategies with factor augmentation are verified.

Focusing on the challenges in the ultrahigh-dimensional scenarios, we propose a general strategy based on the SAA formulation of EUM with cardinality constraints. Our proposed strategy is effective even when the number of assets is far more than the sample size because it precludes the estimation error of the moments, a sore point of the mean-variance approach for large portfolios, and gives consistent solutions to the EUM problem. The strategy is justified by the equivalence of the ℓ₁-regularized problem on the nonnegative orthant and the ℓ₀-constrained problem on probability simplex under mild conditions. By working with this flexible convex optimization problem, we can identify a small subset of diversified assets with superior out-of-sample portfolio performance. Hence, we present a theoretically sound and computationally efficient strategy to make high-dimensional PM reliable and actionable in the rapidly growing financial asset market.

Albeit the present study demonstrates successful examples in constructing diversified portfolios with our SAA-based method, the limitations of our study remain to be addressed. For example, almost all data-driven methods are susceptible to outliers when facing higher volatility. One possible direction is to incorporate robust optimization techniques into our analysis framework or use outlier detection algorithms to remove unwanted samples. Another limitation is that the current work considers long-only portfolios, while allowing short sales may be an important direction for further work. Besides, we may introduce liquidity risk (Acharya and Pedersen 2005), or consider different types of investors.

Supplementary Materials

Supplementary Notes include the proof for all the theorems and extra experiment results. The programs to replicate all our experiments can be obtained from https://github.com/jaydu1/SparsePortfolio/tree/supplement.

Acknowledgments

We thank the editor, the associate editor and two anonymous reviewers for their valuable and constructive comments, which led to significant improvements of this paper. Wang’s research is partially supported by NSFC (72171216, 12231017, 71921001, 71991474), and Science and Technology Program of Guangzhou, China (202002030129).

ORCID

Jin-Hong Du https://orcid.org/0000-0001-9683-4146
Yifeng Guo https://orcid.org/0000-0003-1206-5744
Xueqin Wang https://orcid.org/0000-0001-5205-9950

References

Acharya, V. V., and Pedersen, L. H. (2005), “Asset Pricing with Liquidity Risk,” Journal of Financial Economics, 77, 375–410. [790]
Algoet, P. H., and Cover, T. M. (1988), “Asymptotic Optimality and Asymptotic Equi-partition Properties of Log-Optimum Investment,” The Annals of Probability, 16, 876–898. [785]
Andersen, M., Dahl, J., and Vandenberghe, L. (2021), Cvxopt: Python Software for Convex Optimization, version 1.2.6. Available at https://cvxopt.org. [787]
