

# Portfolio Selection via Subset Resampling

Weiwei Shen,<sup>†,‡</sup> Jun Wang<sup>†</sup>

<sup>†</sup>School of Computer Science and Software Engineering  
East China Normal University, Shanghai, China

<sup>‡</sup>GE Global Research Center, Niskayuna, NY, USA,  
realsww@gmail.com, wongjun@gmail.com

## Abstract

As the cornerstone of the modern portfolio theory, Markowitz’s mean-variance optimization is a major model adopted in portfolio management. However, the estimation errors in its input parameters substantially deteriorate its performance in practice. Specifically, loss could be huge when the number of assets for investment is not much smaller than the sample size of historical data. To hasten the applicability of Markowitz’s portfolio optimization to large portfolios, in this paper, we propose a new portfolio strategy via subset resampling. Through resampling subsets of the original large universe of assets, we construct the associated subset portfolios with more accurately estimated parameters without requiring additional data. By aggregating a number of constructed subset portfolios, we attain a well-diversified portfolio of all assets. To investigate its performance, we first analyze its corresponding efficient frontiers by simulation, provide analysis on the hyperparameter selection, and then empirically compare its out-of-sample performance with those of various competing strategies on diversified datasets. Experimental results corroborate that the proposed portfolio strategy has marked superiority in extensive evaluation criteria.

## 1 Introduction

Portfolio selection has taken on increasing significance in finance for managing a wide range of assets, such as mutual funds, pension funds and university endowments (Brandt 2010). After more than half a century since the seminal work by (Markowitz 1952), the mean-variance framework remains prevalent and represents the most broadly chosen approach in both industry and academia for portfolio selection (Kolm, Tütüncü, and Fabozzi 2014). Its popularity has called for research on complete comprehension and practical implementation. Briefly, the mean-variance framework formulates the risk-return tradeoff of assets via the means and the covariance matrix of asset returns. It intuitively implies that among available portfolios of assets that can achieve a return target, investors should choose the portfolio with the lowest volatility. However, the classical mean-variance portfolio often performs poorly out of sample, as its input parameters, the first two moments of asset returns, are difficult to estimate accurately (Best and Grauer 1991). Moreover, such a situation

worsens when historical return data are limited. In particular, realized loss could be unbounded when the number of assets is larger than the sample size of data (Tu and Zhou 2011). On the other hand, stationarity of the model parameters is another main concern (Broadie 1993). Return data from thirty years ago might have little bearing on returns this year, so parameters are unlikely to be stationary over a long period of time. While the estimation errors decrease when more data are used, more data generally require a longer time horizon. The dilemma between stationarity of parameters and estimation errors exaggerates the challenges in dealing with large portfolios.

One well-known attempt to overcome the estimation problem without demanding more data is the resampled efficient portfolio proposed in (Michaud 1989). Specifically, the basic concept of Michaud’s resampled efficient portfolio comprises of generating resamples of asset returns by a parametric bootstrap procedure, computing the mean-variance portfolio for each resample, and finally averaging over the obtained portfolios to account for the parameter uncertainty. It aims to lessen the impact of estimation risk on portfolio weights and to obtain a more balanced asset allocation, thereby improving the portfolio performance (Michaud and Michaud 2008). However, due to its mixed and ambiguous testing results, its effectiveness has been continually called into question (Markowitz and Usmen 2003; Harvey et al. 2010; Wolf 2013). Mostly, its generalization performance shows limited or no improvement over Markowitz’s mean-variance portfolio (Scherer 2002).

While the development of new resampling based portfolio strategies has relatively stagnated since Michaud’s work (Becker, Gürtler, and Hibbeln 2015), ensemble methods have achieved remarkable success in advancing performance of existing algorithms and meanwhile enriching the methodology on its own (Dietterich 2000; Zhou 2012). As the resampled efficient portfolio could be viewed as an application of the bootstrap aggregating algorithm (Bagging) in order to enhance stability and mitigate variance (Breiman 1996), exploring new ensemble methods and leveraging them into portfolio selection problems deserves a fresh attempt.

The illuminating papers by (Ho 1998) and (Kleiner et al. 2014) have provided a pathway to applying new ensemble methods into portfolio problems. The former presents the random subspace method (RSM) in a classification context.

Weak classifiers are constructed in random subspaces of the data feature space. Those classifiers are subsequently combined by simple majority voting in the final decision rule. The latter proposes a method called Bag of Little Bootstraps (BLB) for massive data to assess the quality of estimators. BLB first randomly selects subsets of the data, and then performs a bootstrap on each subset by constructing weighted resamples of the subset. Such two methods respectively show how subset resampling attacks the challenges embedded in the two sides of a typical data matrix, i.e., the small sample problem when data are high-dimensional and the big data problem when data are massive.

Inspired by the aforementioned two subset resampling based methods, to facilitate the applicability of Markowitz's portfolio optimization to large portfolios, in this paper, we propose a new portfolio selection strategy via subset resampling. In particular, we focus on the problem when the number of assets is large and the number of return data is critical. We resample subsets of the whole portfolio so that the sizes of data for estimating the corresponding covariance matrices have been relatively increased. We then can construct the associated subset optimal portfolios with more accurate covariances. Finally, we attain a well-diversified portfolio of all assets through aggregating a number of prolonged subset optimal portfolios. We offer analysis on the hyperparameter selection and conduct extensive empirical comparisons over ten peer strategies on four representative real-world datasets. The experimental results lucidly demonstrate the superiority of the proposed portfolio strategy.

## 2 Background and Related Work

In this section, we review portfolio selection problems with emphasis on the mean-variance portfolio, and then rehash motivating ensemble methods in machine learning.

The mean-variance portfolio assumes that investment decisions on getting a diversified portfolio mainly depend on the means and the covariances of asset returns. In practice, investors need to estimate both input parameters and plug them into an optimization routine to get estimated optimal portfolio weights. However, as the estimation errors in parameters are amplified by optimization and then propagate into the solution of the optimization, extreme portfolio weights and a lack of diversification are commonly observed. This phenomenon that has eventually ruined the out-of-sample performance of Markowitz's portfolio is coined as error maximization by (Michaud 1989).

Tremendous efforts have been expended to handle the estimation risk on the parameter uncertainty. Among them, as the return means are extremely difficult to estimate accurately (Merton 1980), (Scherer 2011) emphasizes that without incorporating the expected return the yield minimum-variance portfolio often performs better out of sample. (Jorin 1986) suggests using the Bayes-Stein shrinkage estimator for the return mean estimation. (Ledoit and Wolf 2004) propose a robust and effective shrinkage estimator for the covariance matrix estimation. (Fan, Zhang, and Yu 2012) and (Shen, Wang, and Ma 2014) show superior portfolio performances when various types of norm regularities are incorporated into the mean-variance framework. (Kan and Zhou 2007) and (Tu

and Zhou 2011) propose three-fund and four-fund blending portfolios to further improve the models based on the Bayes-Stein shrinkage estimator, respectively. More comprehensive reviews may be referred to (Brandt 2010).

Meanwhile, the impressive records of applying machine learning algorithms into numerous regimes spark their applications and adoptions in finance. Specifically, over years machine learning researchers have made significant contributions in designing portfolio selection strategies from many novel aspects. Among them, (Cover 1991) and (Blum and Kalai 1999) propose and analyze the constant-rebalanced portfolio with or without transaction costs. (Borodin, El-Yaniv, and Gogan 2004) consider learning the best asset by exploiting the market volatility and the statistical relationship between assets. (Agarwal et al. 2006) use a Newton step based method to compute the portfolio for the next iteration in the universal portfolio context. (Li et al. 2012) focus on a market timing portfolio of determining a passive or an aggressive trading strategy. (Shen et al. 2015) and (Shen and Wang 2016) propose to employ the bandit learning framework to attack portfolio problems. Illustration over a wide range of portfolio strategies from machine learning may be found in the survey by (Li and Hoi 2014).

On the other hand, ensemble methods in machine learning virtually share the same theme as portfolio selection: namely, diversification (Derbeko, El-Yaniv, and Meir 2002; Zhou 2012). The level of diversity of weak learners determines the generalization quality of the aggregated learner. Similar to making investment decisions in assets, if we had access to a learner with perfect generalization performance, then there would be no necessity to appeal to ensemble techniques (Dietterich 2000; Rokach 2010). Among ensemble methods, bagging and the random subspace method are particularly efficacious in improving weak learners when training sample sizes are small (Polikar 2006). Also, it is known in machine learning that bagging is useless for learners having a decreasing learning curve, i.e., that the generalization error of the base learner decreases with an increase in the training sample size (Skurichina and Duin 2002). In addition, (Karoui and Purdom 2016) recently show that the bootstrap method cannot improve the estimation accuracy of the covariance matrix when the sample size is critical. Thus, it is unsurprising to see the comments made by researchers in finance on the ineffectiveness of Michaud's bagging based resampled efficient portfolio (Scherer 2002; Harvey et al. 2010). In contrast, the random subspace method is known to be instrumental in weak learners obtained on small and critical training sample sizes when the learning curve is decreasing. It bolsters the performance of learners which suffer from the curse of dimensionality (Ho 1998; Skurichina and Duin 2002). Hence, to overcome the error maximization in portfolio selection due to a small training sample size, we resort to investigating algorithms based on subset resampling rather than bootstrapping.

## 3 Methodology

In this section, we first introduce the notations and finance terms used in this paper. Then we describe the proposed portfolio selection strategy via subset resampling. Finally we

discuss the properties and behaviors of the new method in constructing efficient frontiers by a simulation study.

### Notation

In a frictionless, self-financing, discrete-time and finite horizon investment environment, we denote a series of trading periods as  $t_k = k\Delta t$ ,  $k = 0, \dots, m$ , where  $\Delta t$  represents one day, one week or one month, depending on the rebalance interval. For simplicity, we use  $k$  for short as the index to indicate the trading period at time  $t_k$  hereafter. From time  $t_{k-1}$  to  $t_k$  the gross return vector of  $n$  risky assets accessible to investors is  $\mathbf{R}_k = (R_{k,1}, \dots, R_{k,n})^\top$ . The gross return  $R_{k,i}$  for the  $i$ -th asset is computed as  $R_{k,i} = S_{k,i}/S_{k-1,i}$ , where  $S_{k,i}$  and  $S_{k-1,i}$  represent the prices of the  $i$ -th asset at time  $t_k$  and  $t_{k-1}$ , respectively. Let us respectively denote by  $\boldsymbol{\mu}_k$  and  $\boldsymbol{\Sigma}_k$  the vector of the means and the covariance matrix of the  $n$  asset returns at time  $t_k$ .

Denote  $\boldsymbol{\omega}_k = (\omega_{k,1}, \dots, \omega_{k,n})^\top$  as the vector of the portfolio weights reflecting the investment decision at time  $t_k$ . The  $i$ -th element of  $\boldsymbol{\omega}_k$  specifies the invested percentage of wealth in the  $i$ -th asset. The sum of the portfolio weights equals one, i.e.,  $\boldsymbol{\omega}_k^\top \mathbf{1} = 1$ , where  $\mathbf{1}$  stands for the  $n \times 1$  vector of ones.  $\omega_{k,i} > 0$  means that investors take a long position of the  $i$ -th asset. In contrast,  $\omega_{k,i} < 0$  indicates a short sale of the  $i$ -th asset, where investors liquidate the borrowed  $i$ -th asset for investment in other assets. If the price of the borrowed asset bounces back, investors, who need to buy back and return the borrowed asset, will suffer from a loss. The maximum loss for a long position is the total amount of invested wealth and the maximum loss of a short sale position could theoretically be infinite. Given a gross return  $\mathbf{R}_k$  and a portfolio weight  $\boldsymbol{\omega}_{k-1}$ , the realized portfolio net return  $r_k$  from time  $t_{k-1}$  to  $t_k$  is computed as  $r_k = \mathbf{R}_k^\top \boldsymbol{\omega}_{k-1} - 1$ .

### Subset Resampling Portfolio

One common formulation of the classical long-short mean-variance optimization can be written as

$$\min_{\boldsymbol{\omega}_k^\top \mathbf{1} = 1} \boldsymbol{\omega}_k^\top \boldsymbol{\Sigma}_k \boldsymbol{\omega}_k \text{ s.t. } \boldsymbol{\mu}_k^\top \boldsymbol{\omega}_k \geq \bar{R}_k, \quad (1)$$

where investors attempt to minimize the risk represented by the total variance of the portfolio while achieving a return target  $\bar{R}_k$ . In practice, both the covariance matrix  $\boldsymbol{\Sigma}_k$  and the means of returns  $\boldsymbol{\mu}_k$  are unknown, thereby requiring estimates. In our study, we apply a subset resampling method to reduce the impact of the estimation errors for large portfolios with a short history of return data.

The subset resampling portfolio (SSR) presented in Algorithm 1 is straightforward to implement. At time  $t_\tau$ , let us assume the available historical asset returns are  $\{\mathbf{R}_k\}_{k=1}^\tau$  with  $\tau \approx n$  and the return target of investors is  $\bar{R}_\tau$ . Instead of taking bootstrap samples of the returns of all assets and computing the estimated optimal portfolio weights for all assets simultaneously as (Michaud and Michaud 2008), SSR averages over multiple estimated optimal weights for small portfolios. More formally, given a subset size  $b < n$ , SSR uniformly at random samples  $s$  subsets of size  $b$  from the original  $n$  assets. Denote by  $\mathcal{I}_j \subset \{1, \dots, n\}$  the corresponding index set with  $|\mathcal{I}_j| = b$  for  $j = 1, \dots, s$ . Denote by

---

### Algorithm 1 Subset Resampling Portfolio

---

- 1: **Inputs:**  $\tau$ : number of periods for estimation;  $\{\mathbf{R}_k\}_{k=1}^\tau$ : historical return data;  $\bar{R}_{\tau+1}$ : one out of sample return;  $n$ : number of assets;  $b$ : subset size;  $s$ : number of sampled subsets;  $\bar{R}_\tau$ : return target;
  - 2: **for**  $j = 1 \rightarrow s$  **do**
  - 3: Randomly sample a set  $\mathcal{I}_j$  of  $b$  indices from  $\{1, \dots, n\}$  without replacement;
  - 4: Select the associated return data as  $\{\mathbf{R}_k^{j,b}\}_{k=1}^\tau$ ;
  - 5: Compute the sample covariance matrix as  $\hat{\boldsymbol{\Sigma}}_\tau^{j,b}$ ;
  - 6: Compute the sample means of returns as  $\hat{\boldsymbol{\mu}}_\tau^{j,b}$ ;
  - 7: Compute the optimal subset portfolio weights  $\hat{\boldsymbol{\omega}}_\tau^{j,b}$  by solving the mean-variance optimization (1) based on the estimated parameters  $\hat{\boldsymbol{\Sigma}}_\tau^{j,b}$  and  $\hat{\boldsymbol{\mu}}_\tau^{j,b}$ ;
  - 8: Construct the weights for the whole portfolio  $\hat{\boldsymbol{\omega}}_\tau^j$ ;
  - 9: **for**  $i = 1 \rightarrow n$  **do**
  - 10:  $\hat{\omega}_{\tau,i}^j = \hat{\omega}_{\tau,i}^{j,b} \mathbb{I}\{i \in \mathcal{I}_j\}$ ;
  - 11: Aggregate the constructed portfolio weights based on  $s$  resamples as  $\hat{\boldsymbol{\omega}}_\tau = s^{-1} \sum_{j=1}^s \hat{\boldsymbol{\omega}}_\tau^j$ ;
  - 12: Compute the realized out-of-sample portfolio net return  $\hat{r}_{\tau+1} = \mathbf{R}_{\tau+1}^\top \hat{\boldsymbol{\omega}}_\tau - 1$ ;
  - 13: **Outputs:** The vector of portfolio weights  $\hat{\boldsymbol{\omega}}_\tau$  and the realized out-of-sample portfolio return  $\hat{r}_{\tau+1}$ .
- 

$\{\mathbf{R}_k^{j,b}\}_{k=1}^\tau$  the associated return data for subset  $j$ . First, for each subset, SSR calculates the associated  $b \times b$  sample covariance matrix as  $\hat{\boldsymbol{\Sigma}}_\tau^{j,b}$  and the  $b \times 1$  sample means of returns as  $\hat{\boldsymbol{\mu}}_\tau^{j,b}$ . Second, SSR computes the optimal subset portfolio weights  $\hat{\boldsymbol{\omega}}_\tau^{j,b}$  by solving the mean-variance optimization (1) based on the estimated mean  $\hat{\boldsymbol{\mu}}_\tau^{j,b}$  and covariance matrix  $\hat{\boldsymbol{\Sigma}}_\tau^{j,b}$ . Third, SSR averages over portfolio weights from all the subsets to generate the weights for the whole portfolio as  $\hat{\boldsymbol{\omega}}_\tau = s^{-1} \sum_{j=1}^s \hat{\boldsymbol{\omega}}_\tau^j$ , where  $\hat{\boldsymbol{\omega}}_\tau^j = (\hat{\omega}_{\tau,1}^j, \dots, \hat{\omega}_{\tau,n}^j)^\top$  is the prolonged  $n \times 1$  vector with  $\hat{\omega}_{\tau,i}^j = \hat{\omega}_{\tau,i}^{j,b} \mathbb{I}\{i \in \mathcal{I}_j\}$  for  $i = 1, \dots, n$ , and the symbol  $\mathbb{I}(\cdot)$  stands for an indicator function. Accordingly, the realized out-of-sample portfolio net return  $\hat{r}_{\tau+1}$  from time  $t_\tau$  to  $t_{\tau+1}$  is  $\hat{r}_{\tau+1} = \mathbf{R}_{\tau+1}^\top \hat{\boldsymbol{\omega}}_\tau - 1$ .

### Discussions

SSR partially sacrifices diversification benefits to alleviate the estimation risk on the parameter uncertainty. The pivotal hyperparameters that determine its performance characteristics are the subset size  $b$  and the number of subsets  $s$ . To enjoy the advantages in mitigating the estimation errors, we need to choose  $b < \min(n, \tau)$ . However, a tradeoff exists when we determine  $b$ . The smaller the subset size is, the more accurate the estimation of the  $b \times b$  sample covariance matrix would be, yet the more diversification benefits would be lost, and vice versa. On the other hand, another tradeoff between computational costs and diversification benefits exists when we determine the number of subsets  $s$ . In principle, a large number of subsets would surely be preferred. However, that could increase the computational burden to be infeasible.

While general results for the optimal hyperparameters are unavailable as they would heavily rely on the underlying return dynamics of assets, our following study offers some insights and suggestions.

In Figure 1, we investigate the empirical performance characteristics of SSR with respect to the two hyperparameters in constructing efficient frontiers by simulation. Briefly, an efficient frontier plots the best risk-return tradeoff curve that a set of assets could possibly achieve by following one particular strategy. An efficient frontier dominates another when the former lies on the upper left of the latter. Following (Broadie 1993), three types of efficient frontiers are illustrated and compared: estimated, true and actual efficient frontiers. An estimated efficient frontier represents in-sample performance of a portfolio strategy, an actual efficient frontier shows out-of-sample performance, and a true efficient frontier stands for the best performance that could be achieved. See the long version of this work for details.

First, Figures 1(a), 1(e) and 1(i) show that the estimated efficient frontiers always become more overoptimistic as the subset size  $b$  increases, and the efficient frontiers directly computed by sample moments are more volatile. Second, Figures 1(b), 1(f) and 1(j) demonstrate that the actual efficient frontiers have a typical U-shape performance due to the tradeoff between diversification benefits and estimation errors, and the results from the subset size  $b = n^{0.8}$ , which is conceptually about 5 data points per asset, are fairly stable. Third, Figures 1(c), 1(d), 1(g), 1(h), 1(k) and 1(l) illustrate that both the estimated and the actual efficient frontiers are less sensitive to the number of subsets  $s$ , and the effects from the subset size  $b$  is more important than the number of subsets  $s$ . In sum, the study suggests that users of SSR should focus on testing its performance by tuning the subset size  $b$  and choose a large number of subsets  $s$  to increase the diversification benefits. Besides cross-validation types of hyperparameter tuning, we suggest two possible guidelines of electing  $b$ : Users could start with  $b$  that gives no fewer than 5 data points per asset and combine this rule with the classical result that 30 assets often form a well-diversified portfolio.

## 4 Experiments

In this section, we first introduce the tested datasets, the competing portfolios and the evaluation metrics. Then we conduct comparison studies and report the experimental results.

### Data

In our experiments, we intentionally choose diversified and large datasets to fairly evaluate our new strategy.

**Fama and French datasets:** In the finance community, the Fama and French datasets have been widely recognized as high-quality and standard evaluation protocols (Fama and French 1992). Based on various types of financial segments of the U.S. stock market, the datasets contain carefully constructed portfolios from historical data. In general, they have an extensive coverage of assets classes and span a long period. In our experiments, the FF100 dataset includes monthly returns of 100 assets over forty years.

Table 1: Summary of the testing datasets

#	Dataset	Frequency	Time Period	$m$	$n$
1.	FF100	Monthly	07/01/1963 - 12/31/2004	498	100
2.	ETF139	Weekly	01/01/2008 - 10/30/2012	252	139
3.	EQ181	Weekly	01/01/2008 - 10/30/2012	252	181
4.	SP434	Daily	09/05/2001 - 08/09/2013	2999	434

**Real-world market datasets:** Three real-world datasets are used in our experiments: ETF139, EQ181 and SP434. Specifically, ETF139 contains 139 exchange-traded funds (ETF). Due to the advantages of having clear tax and fee structures, high liquidity and diversity, ETFs have become popularized among investors. EQ181 is constructed from the individual equities from the large-cap segment of the pool of the Russell Top 200 Index. We exclude those stocks with missing historical data from the start of our testing periods and finally receive a total of 181 assets. Likewise, for the SP434 dataset, we filter the daily return data of the 500 firms listed in the S&P 500 Index and retain 434 assets.

Table 1 summarizes these two types of benchmarks. They implicitly underline different perspectives in performance assessment. On the one hand, FF100 and SP434 highlight the long-term performance due to their long period of time spans and limited selection bias. On the other hand, ETF139 and EQ181 reflect the vicissitude market environment after the recent financial crisis starting from 2007. The four datasets have diverse trading frequencies: monthly, weekly and daily. Thus, through empirical evaluations on those datasets, we can thoroughly understand the performance of each strategy.

### Competing Portfolio Strategies

To comprehensively assess the proposed strategy, we consider ten state-of-the-art competing portfolios: (a) **Equally-weighted portfolio (EW)**: EW is a naive yet robust strategy. It has been shown outperforming 14 sophisticated models across seven real-world datasets as well as one simulated dataset at monthly frequency of 2000 years (DeMiguel, Garlappi, and Uppal 2009). Thus, EW is suggested to serve as the first touchstone in portfolio research. (b) **Value-weighted portfolio (VW)**: VW forms a market mimicking passive portfolio. Most active mutual fund managers have the difficulty in outperforming passive benchmarks such as market indices even before netting out fees (Fama and French 2010). (c) **Minimum-variance portfolio (MV)**: MV based on sample moments has been shown outperforming the classical mean-variance portfolio in different markets and time spans (Scherer 2011). (d) **Resampled efficient portfolio by (Michaud 1989) (RES)**: Since its inception, RES remains the most quoted portfolio strategy based on resampling. (e) **Two-fund portfolio by (Tu and Zhou 2011) (TZT)**: TZT infuses the classical mean-variance and the EW portfolios to achieve both estimation error mitigation and wealth growth. (f) **Three-fund portfolio by (Kan and Zhou 2007) (KZT)**: KZT encompasses the risk-free asset, the mean-variance and MV portfolios as to cancel out estimation errors in the mean-variance portfolio by incorporating the MV portfolio. (g) **Four-fund portfolio by (Tu and Zhou 2011) (TZF)**: TZF

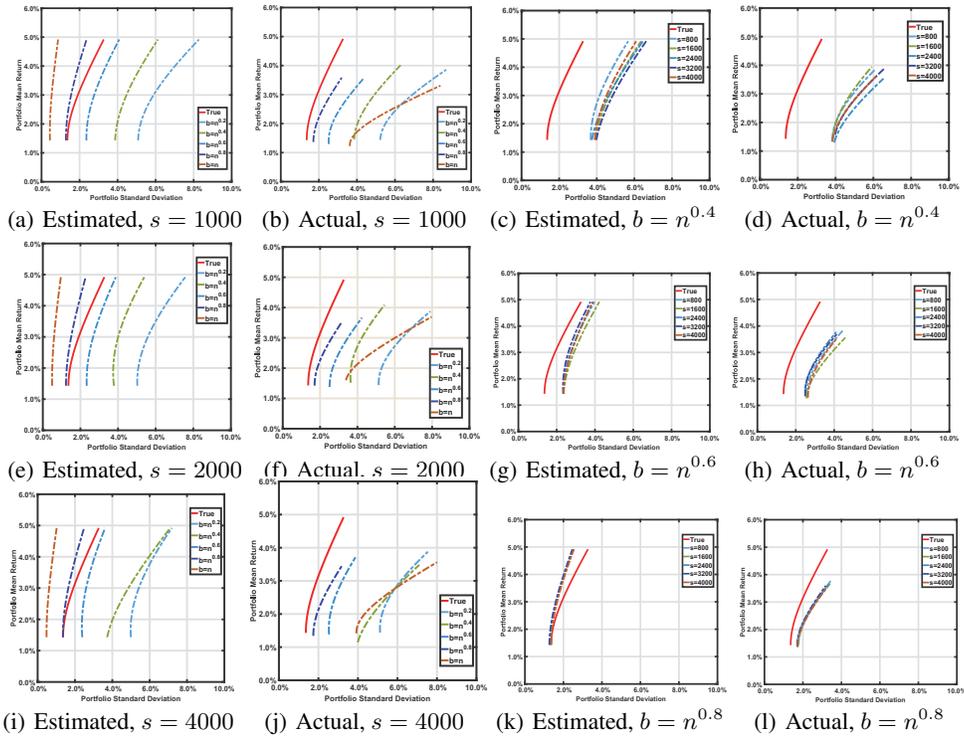


Figure 1: SSR efficient frontiers based on different hyperparameters in a simulation study. The whole dataset SP434 with  $n = 434$  assets is used to estimate the means and the covariance matrix as inputs for simulation. See Table 1 for more information about SP434. By assuming the asset return follows a corresponding multivariate normal distribution,  $\tau = 500$  return data for training are then synthesized by simulation.

is formed by mixing the KZT and the EW portfolios. Their study shows that it performs comparably with EW in some special cases and better in general. (h) **Two-fund portfolio by (Kan, Wang, and Zhou 2016) (KWZ)**: KWZ is an updated version of TZT particularly for portfolios only with risky assets. It is targeting at outperforming EW. (i) **Covariance shrinkage estimator based portfolio by (Ledoit and Wolf 2004) (SKC)**: SKC hinges on a popular and robust shrinkage estimator for the covariance matrix estimation that works well even when the training sample size is critical. (j) **On-line passive aggressive mean reversion portfolio by (Li et al. 2012) (PAMR)**: PAMR from machine learning researchers has been shown robustly outperforming 12 portfolio strategies on six datasets.

## Performance Metrics

Following (DeMiguel, Garlappi, and Uppal 2009), we employ the “rolling-horizon” setting for the sequential out-of-sample performance evaluation. From time  $t_\tau$  to  $t_{m-1}$ , at each rebalancing time  $t_h$  for  $h = \tau, \dots, m-1$ , we first use the return data  $\{\mathbf{R}_k\}_{k=h-\tau+1}^h$  to determine the portfolio weights  $\hat{\omega}_h$ . Second, we compute the realized out-of-sample net return  $\hat{r}_{h+1}$  for the subsequent trading period. Then, we evaluate the out-of-sample characteristics of portfolios by four standard metrics in finance (Brandt 2010): (i) **Sharpe Ratio (SR)**: SR is a common risk-adjusted return measure for a portfolio strat-

egy considering both return and risk. Simply, SR is calculated as the portfolio return normalized by its standard deviation:  $SR = \tilde{r}/\tilde{\sigma}$  with  $\tilde{r}$  as the mean of portfolio net returns and  $\tilde{\sigma}$  as the standard deviation:

$$\tilde{r} = \frac{1}{m - \tau} \sum_{k=\tau+1}^m \hat{r}_k, \quad \tilde{\sigma} = \sqrt{\frac{1}{m - \tau} \sum_{k=\tau+1}^m (\tilde{r} - \hat{r}_k)^2}. \quad (2)$$

To compare strategies based on different rebalancing frequencies, we report the annualized Sharpe ratio as  $\sqrt{H}SR$ , where the scaling factor  $H$  represents the number of rebalancing times per year. In calculation, we use  $H = 12, 52$  and  $252$  for monthly, weekly and daily rebalances, respectively. (ii) **Volatility (VO)**: VO is a broadly computed quantitative risk measure in finance. Similar to SR, we report the annualized volatility using the scaling factor  $H$  and the standard deviations of returns  $\tilde{\sigma}$  in equation (2). (iii) **Turnover Rate (TO)**: TO is a crucial metric to quantify the volume of rebalancing. Due to the existence of market frictions such as transaction costs and taxes, a high TO leads to extra trading costs and could drastically degrade after-cost return performance. TO is computed by

$$TO = \frac{1}{m - \tau - 1} \sum_{k=\tau}^{m-2} \|\hat{\omega}_{k+} - \hat{\omega}_{k+1}\|_1, \quad (3)$$

where  $\hat{\omega}_{k+}$  is the portfolio weight vector before rebalancing at  $t_{k+1}$  and  $\|\cdot\|_1$  denotes  $l_1$ -norm. Briefly, the above equation

Table 2: Portfolio performance of strategies

Dataset	Metrics	SSR	EW	VW	MV	RES	TZT	KZT	TZF	KWZ	SKC	PAMR
FF100	SR	<b>1.51</b>	0.93	1.03	0.62	0.62	0.83	0.11	1.04	0.68	1.36	0.43
	<i>p</i> -value	0.00	1.00	0.01	0.18	0.16	0.36	0.00	0.30	0.29	0.02	0.00
	VO (%)	<b>12.64</b>	18.29	17.98	29.90	30.03	21.17	344.16	18.81	40.11	13.51	27.71
	TO (%)	38.95	2.37	<b>0.00</b>	781.11	802.13	544.12	25245.53	232.76	126.60	55.75	133.81
ETF139	MDD (%)	<b>33.71</b>	37.28	37.03	81.09	81.15	58.62	100	44.06	88.25	34.98	54.16
	SR	<b>1.61</b>	0.51	0.50	0.73	0.65	0.42	0.97	0.32	-0.27	1.13	0.86
	<i>p</i> -value	0.10	1.00	0.55	0.83	0.83	0.82	0.14	0.51	0.19	0.49	0.21
	VO (%)	<b>2.09</b>	21.20	21.06	3.82	3.89	21.78	56.46	18.60	22.57	2.67	36.85
EQ181	TO (%)	25.54	1.08	<b>0.00</b>	801.99	825.08	403.91	10520.42	338.78	5762.82	20.21	153.38
	MDD (%)	<b>1.27</b>	22.35	22.51	4.42	4.59	24.83	19.35	20.44	28.11	1.40	34.33
	SR	1.76	0.97	0.97	0.80	0.80	1.34	-0.24	1.00	<b>1.96</b>	1.48	0.30
	<i>p</i> -value	0.57	1.00	0.84	0.90	0.86	0.41	0.41	0.87	0.43	0.76	0.21
SP434	VO (%)	<b>7.79</b>	15.43	15.29	23.72	23.87	17.45	45.50	15.07	45.57	8.87	22.86
	TO (%)	15.70	1.85	<b>0.00</b>	552.75	568.96	140.72	1332.92	55.33	114.74	27.34	140.48
	MDD (%)	<b>3.07</b>	9.19	9.17	9.97	10.43	10.21	22.18	8.45	16.81	6.62	18.7
	SR	<b>1.92</b>	0.78	0.72	1.28	1.27	-0.10	0.40	0.82	0.46	1.27	1.14
SP434	<i>p</i> -value	0.00	1.00	0.52	0.23	0.23	0.08	0.41	0.70	0.46	0.23	0.26
	VO (%)	<b>6.98</b>	20.16	18.26	13.18	13.18	61.12	564.66	17.14	138.27	7.78	61.12
	TO (%)	6.56	1.50	<b>0.00</b>	131.73	144.76	3816.62	5911.32	57.59	17.21	24.9	130.39
	MDD (%)	18.14	46.34	44.71	<b>15.29</b>	16.01	100	100	39.49	99.99	26.63	65.30

Note: We use short sliding widows  $\tau = 120, 150, 200$  and  $500$  for the four datasets, respectively. SSR, MV, RES and SKC implement the minimum-variance portfolio without return targets. For SSR,  $b = n^{0.7}$  and  $s = 15k$ . The *p*-values under the SR results quantify the statistical significance of the difference in SR between two comparing portfolios. EW is the benchmark portfolio. As returns are hardly i.i.d., the studentized circular block bootstrapping methodology in (Ledoit and Wolf 2008) is used to compute the *p*-values with 5000 bootstrap resamples and a block with the size of 5.

generates an average absolute value of the rebalancing trades across all the assets and over all the trading periods. (iv) **Maximum Drawdown (MDD)**: MDD is one of the topmost risk measures for money management professionals, as large drawdowns often lead to fund redemptions (Magdon-Ismail and Atiya 2004). MDD is computed as the maximum drop of the cumulative wealth over a tested time period:

$$\text{MDD} = \max_{k \in [\tau, m]} (M_k - W_k), \quad (4)$$

with the running maximum of the cumulative wealth  $M_k$  and the cumulative wealth  $W_j$  at time  $t_j$  obtained by

$$M_k = \max_{j \in [\tau, k]} W_j \text{ with } W_j = \prod_{l=\tau+1}^j (1 + \hat{r}_l), \quad (5)$$

where we assume investors start with one dollar.

## Results

Table 2 presents the overall performance of the compared 11 portfolios across the four tested benchmarks. SSR apparently outperforms the classical Markowitz's type portfolio MV and Michaud's bagging based portfolio RES. Consistent with the observations in (Becker, Gürtler, and Hibbeln 2015), MV and RES have almost identical performance, and the bootstrap resampling steps in Michaud's portfolio seem futile. In most of the cases, SSR has higher SRs, lower VO and lower MDDs than the challenging baselines EW and VW, the up-to-date portfolio blending strategies TZT, KZT and TZF, the market timing portfolio PAMR, as well as the shrinkage estimator based portfolio SKC. In sum, SSR behaves as a robust strategy with high return and low risk.

Notably, VW and EW are intrinsically designed as passive strategies for low turnover rates. For example, VW mimics the market trend in a passive way and requires no rebalancing, resulting in zero TOs constantly. After excluding those two

passive strategies, as a moderately active trading strategy, SSR achieves the lowest TOs among all the remaining nine portfolio strategies. It has been recognized that high turnover rates stem from the sensitivity of portfolio optimization solutions to the estimation errors in input parameters (Best and Grauer 1991). Hence, portfolios with high estimation risks are sensitive to changes in the market and could have inflated before-cost returns. For instance, although KWZ has a little higher SR than SSR on EQ181, its TO is ten times larger than that of SSR, which inevitably leads to huge trading costs. This observation in turn indicates that in a multi-period trading setup the sacrificed diversification benefits in the subset resampling step are compensated not only by diminished estimation errors but also by low turnover rates. More experimental results with analysis are included in the long version of this work.

## 5 Conclusions and Discussions

In this paper, we develop a new portfolio selection strategy via subset resampling to promote the applicability of Markowitz's portfolio optimization to a large pool of assets. By sacrificing some diversification benefits, the new strategy receives the compensation of largely ameliorated estimation errors. We provide analysis on the hyperparameter selection and offer extensive comparison studies with ten representative portfolio strategies on four diversified datasets. As the work extends the depth of the portfolio research via ensemble methods, we believe that it represents an effort in leveraging successfully applied machine learning algorithms into long-lasting finance problems. Our future work includes explicitly incorporating the consideration of market frictions such as transaction costs and taxes into more sophisticated ensemble learning algorithms (Rokach 2010; Shen and Wang 2015).

## References

- Agarwal, A.; Hazan, E.; Kale, S.; and Schapire, R. E. 2006. Algorithms for portfolio management based on the Newton method. In *Proceedings of the 23rd International Conference on Machine Learning*, 9–16.
- Becker, F.; Gürtler, M.; and Hibbeln, M. 2015. Markowitz versus Michaud: Portfolio optimization strategies reconsidered. *The European Journal of Finance* 21(4):269–291.
- Best, M. J., and Grauer, R. R. 1991. On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies* 4(2):315–342.
- Blum, A., and Kalai, A. 1999. Universal portfolios with and without transaction costs. *Machine Learning* 35(3):193–205.
- Borodin, A.; El-Yaniv, R.; and Gogan, V. 2004. Can we learn to beat the best stock? *Journal of Artificial Intelligence* 21:579–594.
- Brandt, M. W. 2010. Portfolio choice problems. In Ait-Sahalia, Y., and Hansen, L. P., eds., *Handbooks of Financial Econometrics*. Elsevier. 269–336.
- Breiman, L. 1996. Bagging predictors. *Machine learning* 24(2):123–140.
- Broadie, M. 1993. Computing efficient frontiers using estimated parameters. *Annals of Operations Research* 45(1):21–58.
- Cover, T. M. 1991. Universal portfolios. *Mathematical Finance* 1(1):1–29.
- DeMiguel, V.; Garlappi, L.; and Uppal, R. 2009. Optimal versus naive diversification: How inefficient is the  $1/N$  portfolio strategy? *The Review of Financial Study* 22:1915–1953.
- Derbeko, P.; El-Yaniv, R.; and Meir, R. 2002. Variance optimized bagging. In *European Conference on Machine Learning*, 60–72. Springer.
- Dietterich, T. G. 2000. Ensemble methods in machine learning. In *International Workshop on Multiple Classifier Systems*, 1–15.
- Fama, E. F., and French, K. R. 1992. The cross-section of expected stock returns. *Journal of Finance* 47(2):427–465.
- Fama, E. F., and French, K. R. 2010. Luck versus skill in the cross-section of mutual fund returns. *The Journal of Finance* 65(5):1915–1947.
- Fan, J.; Zhang, J.; and Yu, K. 2012. Asset allocation and risk assessment with gross exposure constraints for vast portfolios. *Journal of American Statistical Association* 412–428.
- Harvey, C. R.; Liechty, J. C.; Liechty, M. W.; and Müller, P. 2010. Portfolio selection with higher moments. *Quantitative Finance* 10(5):469–485.
- Ho, T. K. 1998. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(8):832–844.
- Jorin, P. 1986. Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative analysis* 21:279–292.
- Kan, R., and Zhou, G. 2007. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis* 42(03):621–656.
- Kan, R.; Wang, X.; and Zhou, G. 2016. Optimal portfolio selection with and without risk-free asset. Available at SSRN 2729429.
- Karoui, N. E., and Purdom, E. 2016. The bootstrap, covariance matrices and PCA in moderate and high-dimensions. *arXiv preprint arXiv:1608.00948*.
- Kleiner, A.; Talwalkar, A.; Sarkar, P.; and Jordan, M. I. 2014. A scalable bootstrap for massive data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76(4):795–816.
- Kolm, P. N.; Tütüncü, R.; and Fabozzi, F. J. 2014. 60 years of portfolio optimization: Practical challenges and current trends. *European Journal of Operational Research* 234(2):356–371.
- Ledoit, O., and Wolf, M. 2004. Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management* 30(4):110–119.
- Ledoit, O., and Wolf, M. 2008. Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance* 15:850–859.
- Li, B., and Hoi, S. C. 2014. Online portfolio selection: A survey. *ACM Computing Survey* 46(3):35.
- Li, B.; Hoi, P.; Hoi, S. C.; and Gopalkrishnan, V. 2012. PAMR: Passive aggressive mean reversion strategy for portfolio selection. *Machine Learning* 87:221–258.
- Magdon-Ismail, M., and Atiya, A. F. 2004. Maximum drawdown. *Risk Magazine* 17(10):99–102.
- Markowitz, H., and Usmen, N. 2003. Resampled frontiers versus diffuse Bayes: An experiment. *Journal Of Investment Management* 1(4):9–25.
- Markowitz, H. 1952. Portfolio selection. *Journal of Finance* 7:77–91.
- Merton, R. C. 1980. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8:323–361.
- Michaud, R. O., and Michaud, R. O. 2008. *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Oxford University Press.
- Michaud, R. O. 1989. The Markowitz optimization enigma: is ‘optimized’ optimal? *Financial Analysts Journal* 45(1):31–42.
- Polikar, R. 2006. Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine* 6(3):21–45.
- Rokach, L. 2010. Ensemble-based classifiers. *Artificial Intelligence Review* 33(1-2):1–39.
- Scherer, B. 2002. Portfolio resampling: Review and critique. *Financial Analysts Journal* 58(6):98–109.
- Scherer, B. 2011. A note on the returns from minimum variance investing. *Journal of Empirical Finance* 18(4):652–660.
- Shen, W., and Wang, J. 2015. Transaction costs-aware portfolio optimization via fast Löwner-John ellipsoid approximation. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, 1854–1860.
- Shen, W., and Wang, J. 2016. Portfolio blending via Thompson sampling. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, 1983–1989.
- Shen, W.; Wang, J.; Jiang, Y.-G.; and Zha, H. 2015. Portfolio choices with orthogonal bandit learning. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, 974–980.
- Shen, W.; Wang, J.; and Ma, S. 2014. Doubly regularized portfolio with risk minimization. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 1286–1292.
- Skurichina, M., and Duin, R. P. W. 2002. Bagging, boosting and the random subspace method for linear classifiers. *Pattern Analysis & Applications* 5(2):121–135.
- Tu, J., and Zhou, G. 2011. Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics* 99:204–215.
- Wolf, M. 2013. Resampling vs. shrinkage for benchmarked managers. *Wilmott Magazine* 76–81.
- Zhou, Z.-H. 2012. *Ensemble Methods: Foundations and Algorithms*. CRC press.