Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)

Portfolio Blending via Thompson Sampling

Weiwei Shen[†, ‡] and Jun Wang[†]
[†]School of Computer Science and Software Engineering
East China Normal University, Shanghai, China
[‡]GE Global Research Center, Niskayuna, NY, USA,
realsww@gmail.com, wongjun@gmail.com

Abstract

As a definitive investment guideline for institutions and individuals, Markowitz's modern portfolio theory is ubiquitous in financial industry. However, its noticeably poor out-of-sample performance due to the inaccurate estimation of parameters evokes unremitting efforts of investigating effective remedies. One common retrofit that blends portfolios from disparate investment perspectives has received growing attention. While even a naive portfolio blending strategy can be empirically successful, how to effectually and robustly blend portfolios to generate stable performance improvement remains less explored. In this paper, we present a novel online algorithm that leverages Thompson sampling into the sequential decision-making process for portfolio blending. By modeling blending coefficients as probabilities of choosing basis portfolios and utilizing Bayes decision rules to update the corresponding distribution functions, our algorithm sequentially determines the optimal coefficients to blend multiple portfolios that embody different criteria of investment and market views. Compared with competitive trading strategies across various benchmarks, our method shows superiority through standard evaluation metrics.

1 Introduction

The modern portfolio theory framework pioneered by [Markowitz, 1952] has been instrumental in developing and understanding financial markets and investment decision making. Thus far its mean-variance paradigm remains the pervasive formulation of portfolio choice problems in both academia and industry [Brandt, 2010; Kolm *et al.*, 2014]. Its increasing popularity among pension funds, mutual funds and 401(k) plans has called for thorough understanding and careful implementing. Generally, the mean-variance framework formalizes the concept of return-risk tradeoff that investors should consider return and risk together to determine the allocation of funds among investment alternatives. In particular, it suggests that among available portfolios that achieve a particular return objective, investors should invest the portfolio with the smallest variance. All other portfolios are "inefficient" in terms of having a higher variance representing a higher risk. However, due to the hurdle of accurately estimating involved parameters, the mean-variance portfolio often performs poorly in out-of-sample settings [Broadie, 1993].

On the other hand, the concept of blending portfolios arising from different investment perspectives to construct a new portfolio can be traced back to the ingenious two-fund separation theorem by [Tobin, 1958]. In the mean-variance framework, the two-fund separation theorem states that the efficient portfolio can be considered as a linear combination of two portfolios. Given the unsatisfactory out-of-sample performance of the mean-variance portfolio, the two-fund separation theorem naturally brings us the opportunity of blending portfolios to achieve better performance than the mean-variance portfolio and other heuristic strategies. However, as the pivotal drivers of performance, blending coefficients that characterize the combination of portfolios demands a systematic and comprehensive way to determine.

Meanwhile, the massive amounts of data in the financial industry spark the use of advanced data analysis tools to implement online portfolio strategies. As machine learning algorithms have shown extreme efficiency in the automated process of large datasets, over years researchers have made significant efforts of designing real time data stream based portfolio strategies [Blum and Kalai, 1999; Cover and Ordentlich, 1996; Borodin *et al.*, 2004; Agarwal *et al.*, 2006; Li and Hoi, 2012; Shen *et al.*, 2014; Shen and Wang, 2015]. Illustration over a wide range of online portfolio strategies may be found in the survey by [Li and Hoi, 2014], and the references therein.

In this paper, we address the conundrum of appropriately determining blending coefficients of portfolios in an online setting by a machine learning algorithm. We believe that it is a step in the development of exploiting machine learning algorithms for portfolio choice problems. In particular, we first construct three basis portfolios in finance prepared for blending and formulate the portfolio blending problem into a Thompson sampling problem. Then we model blending coefficients as probabilities of choosing basis portfolios and rest on Bayes decision rules to update the distribution characterizing those probabilities. With two sets of different basis portfolios, we design two blended portfolios accordingly. To justify their performance from various angles, we employ a suite of standard finance metrics consisting of *Sharpe ratios*, *volatility*, and *maximum drawdowns*. Our extensive empirical studies and comparisons of the two blended portfolios with seven competing strategies over five real-world market datasets conspicuously illustrate the superiority of the proposed Thompson sampling based blending algorithm.

2 Background and Related Work

In this section, we briefly discuss two topics, i.e., *Thompson sampling* and *portfolio blending*. The former covers a short history, the current advance, and the formulation in a bandit setting of Thompson sampling; the latter comprises of the discussion about the two-fund theorem with shrinkage rules and representative work.

2.1 Thompson Sampling

As a heuristic solution to the well-known exploration-exploitation problem, Thompson sampling was first induced by [Thompson, 1933] in the early 1930's. Surprisingly, unlike other probability matching methods, such as Bayes decision rules, Thompson sampling remained unpopular for an extremely long time in the research community. Recently, Thompson sampling has been revisited by many researchers and successfully applied to various machine learning problems, such as reinforcement learning [Granmo, 2010], online advertising [Graepel *et al.*, 2010] and Markov decision processes [Strens, 2000]. In particular, for multi-armed bandit learning problems, a recent empirical study shows that Thompson sampling is a highly promising strategy of addressing the exploration-exploitation tradeoff [Chapelle and Li, 2011]. Despite of its simplicity, Thompson sampling achieves comparable performance with competing methods such as upper confidence bound (UCB) and ϵ-greedy methods. In addition, although in contrast with UCB [Auer *et al.*, 2002] Thompson sampling lacks strong theoretical guarantees on the regret, recent studies have shown that it converges asymptoticly in the bandit learning context [Granmo, 2010; Agrawal and Goyal, 2012; Gopalan *et al.*, 2014]. Also, the role of risk in bandit learning has started to be acknowledged and studied [Sani *et al.*, 2012; Shen *et al.*, 2015]. We briefly describe the Thompson sampling algorithm below.

Consider a set of actions $\mathcal{A}$ and a reward r. In each round, a player chooses an action $\alpha \in \mathcal{A}$ and then receives the corresponding reward $r \in \mathbb{R}$ following a probability distribution that depends on the issued action. The player attempts to determine a policy that can generate an action set $\{\alpha_1, \ldots, \alpha_k, \ldots, \alpha_m\}$ that creates the maximum cumulative reward after playing m rounds. In a Bayesian setting, the set of past observations $\mathcal{D}$ that consists of $\{(\alpha_1, r_1), \ldots, (\alpha_k, r_k)\}$ is modeled as a parametric likelihood function $\mathcal{P}(r|\alpha, \boldsymbol{\theta})$ with a set of parameters $\boldsymbol{\theta}$. By assuming a prior distribution $\mathcal{P}(\boldsymbol{\theta})$ on those parameters, the posterior distribution is given by $\mathcal{P}(\boldsymbol{\theta}|\mathcal{D}) \propto \prod_k \mathcal{P}(r_k|\alpha_k, \boldsymbol{\theta})\mathcal{P}(\boldsymbol{\theta})$. Denoting by $\boldsymbol{\theta}^*$ the set of unknown true parameters, the optimal action at time t_k is determined by maximizing the expected reward, i.e., $\alpha_k^* = \arg\max_{\alpha_k} \mathbb{E}(r_k|\alpha_k, \boldsymbol{\theta}^*)$. However, since $\boldsymbol{\theta}^*$ is unknown, by randomly selecting an action according to its probability of being optimal, the action is chosen with probability:

$$\int \mathbb{I}\big[\mathbb{E}(r_k|\alpha_k, \boldsymbol{\theta}) = \max_{\alpha'_k} \mathbb{E}(r_k|\alpha'_k, \boldsymbol{\theta})\big]\mathcal{P}(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}, \qquad (1)$$

where $\mathbb{I}$ is the indicator function. The implementation of Thompson sampling strategy can be realized by samplings, which is straightforward in many applications including multi-armed bandit problems. Briefly, in each round, the set of parameters $\boldsymbol{\theta}$ is sampled from the posterior $\mathcal{P}(\boldsymbol{\theta}|\mathcal{D})$ and the action α_k are chosen to maximize $\mathbb{E}(r_k|\alpha_k, \boldsymbol{\theta})$. A detailed description of Thompson sampling research may be found in [Russo and Van Roy, 2014].

2.2 Portfolio Blending

Although blending portfolios to construct a better performing portfolio sounds naive, the observed empirical results have demonstrated its superiority [DeMiguel *et al.*, 2009]. Theoretically, the portfolio structure induced by Tobin's two-fund theorem implies that the two-fund theorem falls under the rubric of applying shrinkage directly to the portfolio weights. Since shrinkage estimators mitigate estimation error by introducing bias, the approach of blending portfolios provides a pathway to improving the mean-variance portfolio.

While the effectiveness of blending disparate portfolios varies by virtue of the specified shrinking target, they can often outperform the mean-variance portfolio and other heuristic portfolios [Meucci, 2009]. In particular, [Kan and Zhou, 2007] propose a three-fund blending portfolio to further improve the models based on Bayes-Stein shrinkage estimators [Jorion, 1986]. They include the third fund as to diminish the adverse impact of estimation error in terms of hedging the estimation risk embedded in the first two funds. [Tu and Zhou, 2011] consider optimally blending the equally-weighted portfolio with the mean-variance portfolio or with their early proposed three-fund blending portfolio. They calibrate the blending coefficients under the assumption of independent and identically distributed (i.i.d.) normal returns by maximizing investors' expected utility. Their results show that their four-fund blending portfolio outperforms the mean-variance portfolio but not always performs as well as the equally-weighted portfolio. Recently, [DeMiguel *et al.*, 2013] attack the similar problem as [Tu and Zhou, 2011] by testing more economic criteria for coefficient calibration. Their results show the variance minimization criterion is most robust. Furthermore, among numerous approaches to improving the performance of the mean-variance portfolio, many of them essentially share the concept of portfolio blending in different forms [Jorion, 1986; Ledoit and Wolf, 2008]. A more comprehensive review of those variants of the mean-variance portfolio may be referred to [Kolm *et al.*, 2014].

3 Methodology

In this section, we first introduce the notations and finance terms used in this paper. Then we discuss three basis portfolios for blending, formulate the problem of portfolio blending into a Bernoulli bandit problem, and calibrate the blending coefficients by Thompson sampling. Finally, we summarize the proposed algorithm.

3.1 Notations

In a 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 Δt represents 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 denoted as $\mathbf{R}_k = (R_{k,1}, \dots, R_{k,i}, \dots, R_{k,n})^\top$. 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.

Denote by $\boldsymbol{\omega}_k = (\omega_{k,1}, \dots, \omega_{k,i}, \dots, \omega_{k,n})^\top$ 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. We assume the sum of all the portfolio weights equals one, i.e., $\boldsymbol{\omega}_k^\top \mathbf{1} = \sum_{i=1}^n \omega_{k,i} = 1$, where $\mathbf{1}$ is a column vector with ones as its entities. If $\omega_{k,i} > 0$, it indicates 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 to invest other assets. If the price of the borrowed asset rebounds, investors will suffer from a loss. The maximum loss for a long position will be the total amount of invested wealth and the maximum loss of a short sale position could be infinity theoretically. Given gross returns and portfolio weights, we can compute the realized portfolio before-cost *net return* μ_k from time t_{k-1} to t_k as $\mu_k = \mathbf{R}_k^\top \boldsymbol{\omega}_k - 1$.

3.2 Basis Portfolios

In our study, we focus on three basis portfolios for blending, i.e., the equally-weighted, the value-weighted and the minimum-variance portfolios. Those portfolios are standard in finance and easy to compute from data.

Equally-weighted portfolio (EW): EW simply ignores all data information and distributes the investment equally among all the assets:

$$\boldsymbol{\omega}_k^{\text{EW}} = \frac{1}{n}\mathbf{1}. \tag{2}$$

Value-weighted portfolio (VW): As a passive market mimicking strategy, VW is calculated by:

$$\boldsymbol{\omega}_k^{\text{VW}} = \frac{\boldsymbol{\omega}_{k-1} \circ \mathbf{R}_{k-1}}{\boldsymbol{\omega}_{k-1}^\top \mathbf{R}_{k-1}}, \tag{3}$$

where $\circ$ denotes the Hadamard product of two vectors. VW assigns a weight to each asset equal to its market capitalization divided by the total market capitalization of all the assets at each rebalancing time.

Minimum-variance portfolio (MV): Denote by $\boldsymbol{\Sigma}_k$ the covariance matrix of the n asset returns $\mathbf{R}_k$ at time t_k. MV as a variant of the mean-variance portfolio is computed by:

$$\boldsymbol{\omega}_k^{\text{MV}} = \underset{\boldsymbol{\omega}_k^\top \mathbf{1} = 1}{\arg\min}\, \boldsymbol{\omega}_k^\top \boldsymbol{\Sigma}_k \boldsymbol{\omega}_k = \frac{\boldsymbol{\Sigma}_k^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}_k^{-1}\mathbf{1}}. \tag{4}$$

3.3 Portfolio Blending with Thompson Sampling

After obtaining the weights of basis portfolios, we take a linear combination to construct the blending portfolios. In particular, we blend the equally-weighted and the minimum-variance portfolios as:

$$\boldsymbol{\omega}_k^{\text{EM}} = \delta_k \boldsymbol{\omega}_k^{\text{MV}} + (1-\delta_k)\boldsymbol{\omega}_k^{\text{EW}}, \tag{5}$$

and blend the value-weighted and the minimum-variance portfolios as:

$$\boldsymbol{\omega}_k^{\text{VM}} = \delta_k \boldsymbol{\omega}_k^{\text{MV}} + (1-\delta_k)\boldsymbol{\omega}_k^{\text{VW}}, \tag{6}$$

where $0 \le \delta_k \le 1$ is the blending coefficient acting as the main driver of the performance after determining the basis portfolios. Intuitively, given a dynamic trading environment, an optimal blending should perform at least as well as any individual strategy. In this paper, we make the sequential decision on the blending coefficient δ_k by applying Thompson sampling to a Bernoulli bandit problem, as discussed below.

First, we consider the blending coefficient δ_k as the probability of choosing the minimum-variance portfolio $\boldsymbol{\omega}_k^{\text{MV}}$. Intuitively, the blending portfolio can be read as the expectation of different portfolios if the blending coefficients are the corresponding probabilities of choosing those portfolios. If basis portfolios are constructed according to different projections of future market conditions, the blending coefficient δ_k acting as a probability captures the market view of investors. For example, if investors lack information to create sophisticated strategies, they may rely more on EW, i.e., put more weight on $\boldsymbol{\omega}_k^{\text{EW}}$. Next, we assume the probability of choosing MV follows a Beta distribution with parameters a and b, i.e., $\delta_k \sim \text{Beta}(a, b)$. The Beta distribution with the support $(0, 1)$ has the probability density function as $f(x; a, b) \propto x^{a-1}(1-x)^{b-1}$ with parameters $a > 0$ and $b > 0$ and the mean $a/(a+b)$. Given the probability density function $f(x; a, b)$, the higher the a and b the tighter is the concentration around its mean. The Beta distribution is advantageous for Bernoulli rewards because if the prior is a $\text{Beta}(a, b)$ distribution, after observing a Bernoulli test, the posterior distribution is $\text{Beta}(a+1, b)$ or $\text{Beta}(a, b+1)$, depending upon whether the test offers a success or a failure. In the limit of $a \to \infty$, investors will be certain to select this portfolio; in contrast, if $b \to \infty$ investors will surely not invest in this portfolio.

Further, to design our Bernoulli test, we set up a benchmark blending portfolio with its blending coefficient equal to the mean of the $\text{Beta}(a, b)$ distribution i.e., $\bar{\delta}_k = a/(a+b)$. Therefore, the corresponding benchmark portfolios are:

$$\bar{\boldsymbol{\omega}}_k^{\text{EM (VM)}} = \bar{\delta}_k \boldsymbol{\omega}_k^{\text{MV}} + (1-\bar{\delta}_k)\boldsymbol{\omega}_k^{\text{EW (VW)}}, \tag{7}$$

where we use $\boldsymbol{\omega}_k^{\text{EW (VW)}}$ for short to represent the portfolio weight vector $\boldsymbol{\omega}_k^{\text{EW}}$ or $\boldsymbol{\omega}_k^{\text{VW}}$. We then sample one $\tilde{\delta}_k$ from the $\text{Beta}(a, b)$ distribution and construct the testing blending portfolios as:

$$\tilde{\boldsymbol{\omega}}_k^{\text{EM (VM)}} = \tilde{\delta}_k \boldsymbol{\omega}_k^{\text{MV}} + (1-\tilde{\delta}_k)\boldsymbol{\omega}_k^{\text{EW (VW)}}. \tag{8}$$

After observing the gross return yield by the rebalancing, we call it a success or a failure based on:

$$\begin{cases} \text{Success} & \mathbf{R}_k^\top \tilde{\boldsymbol{\omega}}_k^{\text{EM (VM)}} > \mathbf{R}_k^\top \bar{\boldsymbol{\omega}}_k^{\text{EM (VM)}} \text{ and } \tilde{\delta}_k > \bar{\delta}_k \\ \text{Success} & \mathbf{R}_k^\top \tilde{\boldsymbol{\omega}}_k^{\text{EM (VM)}} < \mathbf{R}_k^\top \bar{\boldsymbol{\omega}}_k^{\text{EM (VM)}} \text{ and } \tilde{\delta}_k < \bar{\delta}_k \\ \text{Failure} & \mathbf{R}_k^\top \tilde{\boldsymbol{\omega}}_k^{\text{EM (VM)}} > \mathbf{R}_k^\top \bar{\boldsymbol{\omega}}_k^{\text{EM (VM)}} \text{ and } \tilde{\delta}_k < \bar{\delta}_k \\ \text{Failure} & \mathbf{R}_k^\top \tilde{\boldsymbol{\omega}}_k^{\text{EM (VM)}} < \mathbf{R}_k^\top \bar{\boldsymbol{\omega}}_k^{\text{EM (VM)}} \text{ and } \tilde{\delta}_k > \bar{\delta}_k \end{cases}. \tag{9}$$

Algorithm 1 Portfolio Blending via Thompson Sampling

Inputs: $m, n, \mathbf{R}_{-\tau+1}, \ldots, \mathbf{R}_m, \tau$
for $k = 1 \to m$ **do**
  Compute the equally-weighted portfolio $\boldsymbol{\omega}_k^{\text{EW}}$;
  Compute the value-weighted portfolio $\boldsymbol{\omega}_k^{\text{VW}}$ by (3);
  Estimate the covariance matrix of asset returns $\boldsymbol{\Sigma}_k$ by $\{\mathbf{R}_{k-\tau}, \ldots, \mathbf{R}_{k-1}\}$ and compute the minimum-variance portfolio $\boldsymbol{\omega}_k^{\text{MV}}$ by (4);
  Initialize the Beta distribution by $a = 1$ and $b = 1$;
  for $j = 1 \to \tau$ **do**
    Compute the benchmark blending coefficient $\bar{\delta}_j$;
    Construct the benchmark portfolio $\bar{\boldsymbol{\omega}}_j^{\text{EM (VM)}}$ by (7);
    Sample one $\tilde{\delta}_j$ from the Beta(a, b) distribution;
    Construct the testing portfolio $\tilde{\boldsymbol{\omega}}_j^{\text{EM (VM)}}$ by (8);
    Compare the testing and benchmark portfolios according to (9) by using $\mathbf{R}_{j-\tau}$;
    Update a and b:
    if Success **then**
      $a = a + 1$;
    else
      $b = b + 1$;
  Compute the optimal blending coefficient δ_k^* by (10);
  Construct the proposed TS-EM portfolio and the TS-VM portfolio $\boldsymbol{\omega}_k^{\text{TS-EM (TS-VM)}}$ by (11);
Output:
The series of portfolios $\boldsymbol{\omega}_k^{\text{TS-EM (TS-VM)}}$ and the portfolio before-coast net returns $\mu_k^{\text{TS-EM (TS-VM)}}$ for $k = 1, \ldots, m$.

Specifically, if $\mathbf{R}_k^\top \tilde{\boldsymbol{\omega}}_k^{\text{EM (VM)}} > \mathbf{R}_k^\top \bar{\boldsymbol{\omega}}_k^{\text{EM (VM)}}$ and $\tilde{\delta}_k > \bar{\delta}_k$ or $\mathbf{R}_k^\top \tilde{\boldsymbol{\omega}}_k^{\text{EM (VM)}} < \mathbf{R}_k^\top \bar{\boldsymbol{\omega}}_k^{\text{EM (VM)}}$ and $\tilde{\delta}_k < \bar{\delta}_k$, we call it a success because investors have made a wise decision about the overweight or the underweight on MV. Otherwise, we call it a failure because investors have made an inadvisable bet on the weight. A success suggests updating the parameters such that in the next round of rebalance investors should have a higher probability of choosing MV, and vice versa.[1]

Furthermore, similar to the steps in [Agrawal and Goyal, 2012], we apply Thompson sampling to implementing the distribution updating step. We start with the initial prior as Beta$(1, 1)$ and τ periods of historical data. Given no information about the performance of portfolios, Beta$(1, 1)$, i.e., a standard uniform distribution, is reasonable to investors. At each rebalance time, investors construct the aforementioned Bernoulli test, observe a success or a failure thereafter, and correspondingly update the posterior distribution. After the training period with τ rebalances, the algorithm ends up with the updated distribution as Beta$(1 + a_\tau, 1 + b_\tau)$, by assuming investors have encountered a_τ successes and b_τ failures.

Finally, we determine the blending coefficient as the mean of the most updated distribution as:

$$\delta_k^* = \frac{(1 + a_\tau)}{(1 + a_\tau + 1 + b_\tau)}. \tag{10}$$

Namely, the proposed Thompson sampling based equally-weighted and minimum-variance blending portfolio (TS-EM) and the value-weighted and minimum-variance blending portfolio (TS-VM) read:

$$\boldsymbol{\omega}_k^{\text{TS-EM (TS-VM)}} = \delta_k^* \boldsymbol{\omega}_k^{\text{MV}} + (1 - \delta_k^*) \boldsymbol{\omega}_k^{\text{EW (VW)}}. \tag{11}$$

Accordingly, the realized portfolio before-cost net return μ_k from time t_{k-1} to t_k will be

$$\mu_k^{\text{TS-EM (TS-VM)}} = \mathbf{R}_k^\top \boldsymbol{\omega}_k^{\text{TS-EM (TS-VM)}} - 1. \tag{12}$$

On the one hand, while surpassing either EW or MV has been shown arduous, the proposed TS-EM portfolio aims to perform at least as well as EW and MV via the new blending algorithm. On the other hand, by incorporating market trend information in VW and risk control mechanism in MV, the proposed TS-VM portfolio attempts to exploit the interplay of VW and MV, thereby constructing a superior blending portfolio. In addition, we estimate the covariance matrix $\boldsymbol{\Sigma}_k$ by a factor model [Fan *et al.*, 2008] based on the historical data in sliding windows with the size of τ training data. Algorithm 1 succinctly summarizes the detailed procedure of constructing these two blending portfolios.

4 Experiments

In this section, we perform empirical studies to evaluate the proposed portfolio blending algorithm. We first describe the experimental settings, including a brief introduction of the testing benchmarks and the evaluation metrics. Then we will report the results and compare with seven state-of-the-art competing portfolio strategies.

4.1 Data

To fairly appraise the new method, following [DeMiguel *et al.*, 2009; Shen *et al.*, 2014] in our experiments we choose five datasets from two distinct classes of benchmarks that represent both academic standards and real-world market datasets.

Fama and French datasets (FF) [Fama and French, 1992]: As standard evaluation protocols and oft-adopted testbeds in the finance community, the FF datasets are constructed portfolios of broad financial segments of the U.S. stock market. The datasets at the monthly frequency spanning a period of forty years have an extensive coverage to asset classes. **Real-world market datasets** [Shen *et al.*, 2015]: The real-world datasets including ETF139 and EQ181 are crawled from *Yahoo! Finance* on a weekly basis from 2008 to 2012. The ETF139 dataset consists of 139 exchange-traded funds that are traded like stocks in the U.S. market. Not only do they offer investors more flexibility and channels to the market, but also they have the advantages on taxes and interests of the investment over mutual funds. The EQ181 dataset contains individual equities from the large-cap segment of the Russell 200 index that covers 63% of total market capitalization. After removing those stocks with missing historical data from the start of our testing periods, we finally collect a total of 181 U.S. stocks to form the EQ181 dataset.

We summarize those two groups of benchmarks in Table 1. They essentially embody different perspectives for performance assessment. On the one hand, the FF25, FF48 and FF100 datasets underline the long-term performance since the forty-year spanning would introduce limited selection

[1]If $\mathbf{R}_k^\top \tilde{\boldsymbol{\omega}}_k^{\text{EM (VM)}} = \mathbf{R}_k^\top \bar{\boldsymbol{\omega}}_k^{\text{EM (VM)}}$, we do not update the parameters; if $\tilde{\delta}_k = \bar{\delta}_k$, we simply re-sample from the Beta distribution.

Table 1: Summary of the testing datasets

#	Dataset	Frequency	Time Period	m	n	Description
1.	FF25	Monthly	07/01/1963 - 12/31/2004	498	25	Twenty-five portfolios of firms sorted by size and book-to-market
2.	FF48	Monthly	07/01/1963 - 12/31/2004	498	48	Forty-eight industry portfolios representing the U.S. stock market
3.	FF100	Monthly	07/01/1963 - 12/31/2004	498	100	One hundred portfolios of firms sorted by size and book-to-market
4.	ETF139	Weekly	01/01/2008 - 10/30/2012	252	139	One hundred and thirty-nine exchange-traded funds
5.	EQ181	Weekly	01/01/2008 - 10/30/2012	252	181	One hundred and eighty-one U.S. large-cap equities

bias and performance manipulation. On the other hand, the ETF139 and EQ181 datasets emphasize the robustness with respect to the higher trading frequency and the vicissitude market environment after the recent financial crisis in 2007.

4.2 Competing Portfolios

To comprehensively evaluate the performance of the two proposed portfolios, we consider seven state-of-the-art competing portfolios: (a) **Equally-weighted portfolio** (EW): EW in equation (2) has been shown to outperform 14 sophisticated models across seven empirical datasets as well as one simulated dataset at monthly frequency of 2000 years [DeMiguel *et al.*, 2009]. Thus, EW is commonly suggested to serve as the first obvious but challenging benchmark in portfolio research. (b) **Value-weighted portfolio** (VW): While VW in equation (3) forms a passive portfolio, most active mutual fund managers have the difficulty of outperforming passive benchmarks such as the market even before netting out fees [Fama and French, 2010]. (c) **Minimum-variance portfolio** (MV): MV in equation (4) has consistently shown robust performance in different market conditions [Jagannathan and Ma, 2003]. (d) **Two-fund portfolio by [Tu and Zhou, 2011]** (TZT): TZT blends the traditional mean-variance and the EW portfolios to achieve both estimation error reduction and wealth growth. (e) **Three-fund portfolio by [Kan and Zhou, 2007]** (KZT): KZT encompasses the risk-free, the mean-variance and MV portfolios to diminish the inherent estimation error in the mean-variance portfolio by blending its alike variant. (f) **Four-fund portfolio by [Tu and Zhou, 2011]** (TZF): TZF is formed by mixing the KZT and the EW portfolios. Their study shows it performs comparably with EW in some special cases and better in general. (g) **Online moving average reversion based portfolio by [Li and Hoi, 2012]** (MAR): MAR developed by machine learning researchers has been shown to outperform 12 portfolio strategies across five datasets.

In sum, the first three strategies, i.e., EW, VW and MV, have been the common baselines for portfolio research in finance. They have been broadly adopted as the touchstones of portfolio performance. They also represent the special cases of blending with fixed blending coefficients. The next three portfolios, i.e., TZT, KZT and TZF, are well recognized as important portfolio blending strategies so far. They reflect the up-to-date efforts of researchers on portfolio blending.

4.3 Performance Metrics

We employ the "rolling-horizon" settings suggested in [DeMiguel *et al.*, 2009]. Specifically, the sliding windows with the size of $\tau = 120$ months or $\tau = 200$ weeks of training data are used to construct portfolios for the subsequent month or week.[2] We compute the out-of-sample performance of the portfolios by the following standard criteria in finance [Brandt, 2010]: (i) *Sharpe ratios*; (ii) *volatility*, and (iii) *maximum drawdowns*. In addition, we incorporate the information of the turnover of each strategy through deducting the return by a proportional transaction cost [Broadie and Shen, 2016]. We set a cost factor c equal to 50 basis points per transaction to obviate inflated return from large turnovers, as suggested in [DeMiguel *et al.*, 2009].

First, the *Sharpe ratio* (SR), which measures the reward-to-risk ratio of a portfolio strategy, is computed as the portfolio return normalized by its standard deviation:

$$\text{SR} = \frac{\hat{\mu}}{\hat{\sigma}}, \tag{13}$$

where the mean of portfolio after-cost net return $\hat{\mu}$ and the corresponding standard deviation $\hat{\sigma}$ are computed as

$$\hat{\mu} = \frac{1}{m}\sum_{k=1}^{m}\tilde{\mu}_k \text{ and } \hat{\sigma} = \sqrt{\frac{1}{m}\sum_{k=1}^{m}(\tilde{\mu}_k - \hat{\mu})^2}, \tag{14}$$

where $\tilde{\mu}_k = \mu_k(1 - c\|\boldsymbol{\omega}_{k^+} - \boldsymbol{\omega}_k\|_1)$ denotes the after-cost net return from time t_{k-1} to t_k, $\boldsymbol{\omega}_{k^+}$ represents the portfolio weight vector before rebalancing at t_{k+1} and $\|\cdot\|_1$ denotes l_1-norm. SR heightens the significance of gauging portfolio performance with the dual consideration of risk and return.

Second, the *volatility* is a quantitative risk measure of investment. The calculation of the portfolio volatility relates to the standard deviation of returns $\hat{\sigma}$ by (14). To compare strategies based on different rebalancing frequencies, we compute the annualized volatility by $\sqrt{H}\hat{\sigma}$ with H the total number of rebalancing times each year. In our experiments, we set $H = 12$ and $H = 52$ for monthly and weekly rebalances, respectively.

Third, we report the *maximum drawdown* (MDD) for each strategy [Magdon-Ismail and Atiya, 2004]. The maximum drawdown is defined as the maximum drop of the cumulative wealth from its running maximum over a period of time:

$$\text{MDD} = \max_{k\in[0,m]}(M_k - W_k), \tag{15}$$

where the drawdown $M_k - W_k$ is defined as the drop of the wealth from its running maximum M_k:

$$M_k = \max_{j\in[0,k]} W_j, \tag{16}$$

[2]The study in [DeMiguel *et al.*, 2009] shows portfolio performance generally does not vary considerably by using longer than five years of monthly data.

Table 2: Portfolio performance of strategies

Dataset	Metrics	TS-EM	TS-VM	EW	VW	MV	TZT	KZT	TZF	MAR
FF25	SR (%)	33.84	**34.85**	26.40	28.53	30.95	28.48	23.86	33.93	17.76
	p-value	0.00	0.00	1.00	0.02	0.03	0.00	0.30	0.00	0.01
	Vol (%)	13.75	13.75	17.60	17.49	**13.37**	18.15	19.43	16.42	16.97
	MDD (%)	35.95	**35.83**	38.03	37.80	39.51	45.52	57.30	42.37	40.10
FF48	SR (%)	**27.75**	27.39	23.98	23.37	24.14	16.75	15.16	25.81	21.55
	p-value	0.00	0.00	1.00	0.27	0.84	0.33	0.86	0.00	0.74
	Vol (%)	**13.51**	13.58	16.87	16.76	13.96	23.14	18.43	15.62	15.48
	MDD (%)	36.08	**35.84**	42.07	41.53	38.45	51.65	56.13	39.94	41.04
FF100	SR (%)	36.70	**37.90**	26.75	29.76	37.60	10.10	37.48	26.80	20.49
	p-value	0.00	0.00	1.00	0.01	0.00	0.00	0.00	0.00	0.02
	Vol (%)	**13.72**	13.75	18.29	18.10	13.73	26.81	13.34	18.22	17.70
	MDD (%)	36.13	**36.04**	37.38	37.03	37.51	59.88	36.78	37.34	38.33
ETF139	SR (%)	**10.79**	10.58	10.32	10.51	3.52	-30.58	-4.64	10.36	8.36
	p-value	0.05	0.05	1.00	0.44	0.48	0.05	0.80	0.00	0.76
	Vol (%)	10.53	10.38	18.17	18.03	**3.24**	53.72	4.18	17.60	16.59
	MDD (%)	6.51	6.53	11.45	11.33	**2.67**	74.52	4.39	11.03	10.09
EQ181	SR (%)	**15.64**	15.39	13.09	13.44	10.80	-16.30	8.34	13.11	11.75
	p-value	0.00	0.01	1.00	0.42	0.63	0.64	0.45	0.00	0.72
	Vol (%)	9.81	9.81	15.43	15.29	**8.80**	83.65	9.01	15.43	14.49
	MDD (%)	4.80	**4.80**	9.24	9.17	7.35	82.85	8.78	9.22	8.99

where the after-cost cumulative wealth W_k is computed by $W_k = \prod_{j=1}^{k} \boldsymbol{\omega}_j^\top \tilde{\mu}_j$. Since large drawdowns inevitably lead to fund redemptions, MDD has been the top-one risk measure for money management professionals.

To further quantify the statistical significance of the difference in SR between two comparing portfolios, we also report the p-values under the corresponding SR results. To compute the p-values for the case of non-i.i.d. returns, we adopt the studentized circular block bootstrapping methodology in [Ledoit and Wolf, 2008]. In particular, we set the EW portfolio as the benchmark with 1000 bootstrap resamples, 95% significance level, and a block with the size of 5.

4.4 Results

Table 2 presents the overall performance of the compared nine portfolios across the tested five benchmarks. In particular, we report the *Sharpe ratios*, the *volatility* and the *maximum drawdowns* for all portfolios to comprehensively evaluate performance with the emphasis on the tradeoff between return and risk. In most testing cases, the two proposed blending portfolios clearly outperform both the challenging baselines circulated in financial research, i.e., EW and VW, and representative blending strategies, i.e, TZT, KZT and TZF. We observe that the new portfolios consistently produce the highest risk-adjust return across all the benchmarks with statistical significance. In addition, they often yield lower investment risks than the other three compared blending strategies, reflected by the smaller *volatility* and *maximum drawdowns*. Even in some cases, our blending portfolios generate lower risk than the MV strategy whose sole objective is investment risk minimization. Those observations echo with the intrinsic design of our algorithm in calibrating blending coefficients for portfolios with moderate risk according to portfolio gross return. Further, the proposed strategies demonstrate statistically significant better performance with a noticeable effect size than their basis portfolios in risk and return evaluation metrics. As the performance of the blending portfolios stems from the tradeoff between the gains from the blending coefficient and the losses from the estimation errors in estimating that additional parameter, we interpret those positive findings in performance as the evidence supporting the new algorithm.

In summary, our blending strategies formed by two sets of basis portfolios have embedded careful risk control mechanism and market dynamics. Therefore, in most of testing cases our methods can generate superior performance, i.e., higher risk-adjusted returns, lower volatility and drawdown risks, and outperform individual basis portfolios as well as other representative blending portfolios.

5 Conclusions and Discussions

In this paper, we develop a machine learning algorithm of viably blending portfolios from different investment principles to generate robust and high-quality portfolio strategies. Through casting the question of determining blending coefficients into a Bernoulli bandit problem, we implement Thompson sampling to obtain optimal blending portfolios. Two blended portfolios with different basis portfolios consistently outperform seven highly competitive strategies across five datasets. Our results not only address the "$1/n$" portfolio challenge [DeMiguel *et al.*, 2009] but also demonstrate the insights of adapting portfolio strategies to accommodate parameter estimation errors. In our future work, we will extend the current blending algorithm for multiple portfolios by Dirichlet distribution [Silverthorn and Miikkulainen, 2010].

References

[Agarwal *et al.*, 2006] A. Agarwal, E. Hazan, S. Kale, and R. E. Schapire. Algorithms for portfolio management based on the Newton method. In *Proceedings of the 23th International Conference on Machine Learning*, pages 9–16, 2006.

[Agrawal and Goyal, 2012] S. Agrawal and N. Goyal. Analysis of Thompson sampling for the multi-armed bandit problem. In *Proceedings of the 25th Annual Conference on Learning Theory*, pages 39.1–39.26, 2012.

[Auer *et al.*, 2002] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

[Blum and Kalai, 1999] A. Blum and A. Kalai. Universal portfolios with and without transaction costs. *Machine Learning*, 35(3):193–205, 1999.

[Borodin *et al.*, 2004] A. Borodin, R. El-Yaniv, and V. Gogan. Can we learn to beat the best stock? *Journal of Artificial Intelligence*, 21:579–594, 2004.

[Brandt, 2010] M. W. Brandt. Portfolio choice problems. In Y. Ait-Sahalia and L. P. Hansen, editors, *Handbooks of Financial Econometrics*, pages 269–336. Elsevier, 2010.

[Broadie and Shen, 2016] M. Broadie and W. Shen. High-dimensional portfolio optimization with transaction costs. *International Journal of Theoretical and Applied Finance*, 2016.

[Broadie, 1993] M. Broadie. Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, 45(1):21–58, 1993.

[Chapelle and Li, 2011] O. Chapelle and L. Li. An empirical evaluation of Thompson sampling. In *Advances in Neural Information Processing Systems*, pages 2249–2257, 2011.

[Cover and Ordentlich, 1996] T. M. Cover and E. Ordentlich. Universal portfolios with side information. *IEEE Transactions on Information Theory*, 42(2):348–363, 1996.

[DeMiguel *et al.*, 2009] V. DeMiguel, L. Garlappi, and R. Uppal. Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy? *The Review of Financial Study*, 22:1915–1953, 2009.

[DeMiguel *et al.*, 2013] V. DeMiguel, A. Martin-Utrera, and F. J. Nogales. Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking & Finance*, 37(8):3018–3034, 2013.

[Fama and French, 1992] E. F. Fama and K. R. French. The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–465, 1992.

[Fama and French, 2010] E. F. Fama and K. R. French. Luck versus skill in the cross-section of mutual fund returns. *The Journal of Finance*, 65(5):1915–1947, 2010.

[Fan *et al.*, 2008] J. Fan, Y. Fan, and J. Lv. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147:186–197, 2008.

[Gopalan *et al.*, 2014] A. Gopalan, S. Mannor, and Y. Mansour. Thompson sampling for complex online problems. In *Proceedings of the 31st International Conference on Machine Learning*, pages 100–108, 2014.

[Graepel *et al.*, 2010] T. Graepel, J. Q. Candela, T. Borchert, and R. Herbrich. Web-scale Bayesian click-through rate prediction for sponsored search advertising in Microsoft's Bing search engine. In *Proceedings of the 27th International Conference on Machine Learning*, pages 13–20, 2010.

[Granmo, 2010] O. C. Granmo. Solving two-armed bernoulli bandit problems using a Bayesian learning automaton. *International Journal of Intelligent Computing and Cybernetics*, 3(2):207–234, 2010.

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

[Jorion, 1986] P. Jorion. Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(03):279–292, 1986.

[Kan and Zhou, 2007] R. Kan and G. Zhou. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(03):621–656, 2007.

[Kolm *et al.*, 2014] P. N. Kolm, R. Tütüncü, and F. J. Fabozzi. 60 years of portfolio optimization: Practical challenges and current trends. *European Journal of Operational Research*, 234(2):356–371, 2014.

[Ledoit and Wolf, 2008] O. Ledoit and M. Wolf. Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, 15:850–859, 2008.

[Li and Hoi, 2012] B. Li and S. C. Hoi. On-line portfolio selection with moving average reversion. In *Proceedings of the 29th International Conference on Machine Learning*, 2012.

[Li and Hoi, 2014] B. Li and S. C. Hoi. Online portfolio selection: A survey. *ACM Computing Survey*, 46(3):35, 2014.

[Magdon-Ismail and Atiya, 2004] M. Magdon-Ismail and A. F. Atiya. Maximum drawdown. *Risk Magazine*, 17(10):99–102, 2004.

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

[Meucci, 2009] A. Meucci. *Risk and Asset Allocation*. Springer Science & Business Media, 2009.

[Russo and Van Roy, 2014] D. Russo and B. Van Roy. Learning to optimize via posterior sampling. *Mathematics of Operations Research*, 39(4):1221–1243, 2014.

[Sani *et al.*, 2012] A. Sani, A. Lazaric, and R. Munos. Risk-aversion in multi-armed bandits. In *Advances in Neural Information Processing Systems*, pages 3275–3283, 2012.

[Shen and Wang, 2015] W. Shen and J. Wang. Transaction costs-aware portfolio optimization via fast Löwner-John ellipsoid approximation. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, 2015.

[Shen *et al.*, 2014] W. Shen, J. Wang, and S. Ma. Doubly regularized portfolio with risk minimization. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 2014.

[Shen *et al.*, 2015] W. Shen, J. Wang, Y.-G. Jiang, and H. Zha. Portfolio choices with orthogonal bandit learning. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, 2015.

[Silverthorn and Miikkulainen, 2010] B. Silverthorn and R. Miikkulainen. Latent class models for algorithm portfolio methods. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, 2010.

[Strens, 2000] M. Strens. A Bayesian framework for reinforcement learning. In *Proceedings of the 17th International Conference on Machine Learning*, pages 943–950, 2000.

[Thompson, 1933] W. R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, pages 285–294, 1933.

[Tobin, 1958] J. Tobin. Liquidity preference as behavior towards risk. *The Review of Economic Studies*, pages 65–86, 1958.

[Tu and Zhou, 2011] J. Tu and G. Zhou. Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, 99:204–215, 2011.