Parameter-free Robust Optimization for the Maximum-Sharpe Portfolio Problem

Deepayan Chakrabarti

Information, Risk, and Operations Management, McCombs School of Business, Austin, TX
deepay@utexas.edu

Abstract

How can we optimize for the Sharpe ratio if we only have limited training data? Estimates of mean asset returns are noisy, and this noise hurts the out-of-sample Sharpe ratio of current methods. The minimum-variance portfolio, which ignores mean returns, often has a better Sharpe ratio. We develop a parameter-free and scalable method called AlphaRob for this problem. AlphaRob's portfolio is a convex combination of two prespecified portfolios. To select the best combination, AlphaRob fuses robust optimization with a new notion of a portfolio's regret that accounts for the training data's size. Our analysis only needs mild assumptions on the distribution of asset returns. AlphaRob significantly outperforms competing methods on several simulated and real-world datasets, even after adjusting for transaction costs. AlphaRob is 7.5% better on average than the nearest competitor, and 28% better than the next-best combination portfolio method. Using our regret of regret, we are also able to explain the performance of the minimum-variance portfolio.

Keywords: Finance, Robust optimization, Sharpe ratio, Portfolio optimization

1. Introduction

The Sharpe ratio and its derivatives are popular measures of portfolio performance (Hanke & Penev, 2018; Guerreiro & Fonseca, 2020). Given the mean and covariance of asset returns, the portfolio with the maximum Sharpe ratio has a known form. But often, we must use noisy parameter estimates computed from limited training data. The optimization step amplifies estimation errors, resulting in portfolios with poor out-of-sample performance (Jobson & Korkie, 1980; Frost & Savarino, 1986; Jorion, 1986; Michaud, 1989; Zhao et al., 2019). This problem has attracted much attention recently (Kolm et al., 2014).

One approach is to build better estimators. Regularized covariances have smaller mean-squared error than the sample covariance matrix (Ledoit & Wolf, 2004; Bickel & Levina, 2008; Ledoit & Wolf, 2012). Such covariances also arise from constrained portfolios and from regularized versions of the minimum-variance portfolio problem (Frost & Savarino, 1988; Jagannathan & Ma, 2003; Brodie et al., 2009; Fan et al., 2012; DeMiguel et al., 2009a). To estimate the mean vector, we can "shrink" the sample mean towards a fixed target vector (Jorion, 1986; Frost & Savarino, 1986; DeMiguel et al., 2013).

Prior work using such estimators points to a surprising result: the minimum-variance portfolio using a regularized covariance matrix often achieves the best out-of-sample Sharpe ratio (Jagannathan & Ma, 2003; Garlappi et al., 2007; DeMiguel et al., 2009b, 2013; Ledoit & Wolf, 2017; Zhao et al., 2019). Note that the minimum-variance portfolio does not optimize for the Sharpe ratio and completely ignores mean returns. Chopra & Ziemba (1993) showed that errors in mean estimates hurt portfolio performance more than covariance estimation errors. But this only suggests a need for better algorithms to adjust for the mean estimation errors. These previous studies motivate our two main questions:

How can we use noisy estimates of mean asset returns to improve Sharpe ratios?
Why do minimum-variance portfolios often achieve high Sharpe ratios?

We consider these questions in the context of a popular class of portfolios called combination portfolios. Here, the portfolio is a linear combination of two or more prespecified portfolios, such as the equal-weight, minimum-variance, or mean-variance portfolios. Each prespecified portfolio has some advantages (e.g., robustness to errors, or low standard deviation). A carefully chosen combination portfolio can inherit the best mix of these properties to achieve a better Sharpe ratio than any of the prespecified portfolios. Combination portfolios are optimal for a robust version of the mean-variance problem (Garlappi et al., 2007). They also arise from shrinkage estimation and robust optimization (Jorion, 1986; Frost & Savarino, 1986; DeMiguel et al., 2013; Scherer, 2007). Different combination weights can yield portfolios with very different performance. So, choosing the right combination weight is the main problem.

Existing methods to choose combination weights are inadequate for the Sharpe ratio problem. There are several reasons for this. First, most prior work focuses on the mean-variance objective (Garlappi et al., 2007; Kan & Zhou, 2007; Tu & Zhou, 2011; Kirby & Ostdiek, 2012). The mean-variance objective has a user-specified risk aversion parameter, and this affects the combination weight. So we must compare the Sharpe ratios for a range of combinations, one for each risk aversion setting. But estimation errors prevent us from making reliable comparisons. DeMiguel et al. (2013) propose several other objectives (or "calibrations") for choosing the combination weight. But the calibrated combinations still depend on the risk aversion parameter. Further, there is no theoretical justification for preferring one calibration over the others.

Another problem is that several methods make strong distributional assumptions. Kan & Zhou (2007) and Deng et al. (2013) assume Gaussian returns. The uncertainty sets for Garlappi et al. (2007) are also inspired by Gaussian returns. DeMiguel et al. (2013) consider calibrations based on Gaussian returns, but also provide a non-parametric approach.

Finally, the minimum-variance portfolio often achieves a higher out-of-sample Sharpe ratio than existing methods to choose the combination weight (Garlappi et al., 2007; DeMiguel et al., 2013).

We will choose the combination weight via a robust optimization to account for estimation errors. In robust optimization, we first construct an "uncertainty set" of plausible values for the unknown parameters such as the mean and covariance. Then we select the portfolio with the best worst-case performance over the uncertainty set. A variety of uncertainty sets have been proposed for the mean and covariance (Goldfarb & Iyengar, 2003; Tütüncü & Koenig, 2004; Garlappi et al., 2007), for only the covariance (Qiu et al., 2015; Ceria

& Stubbs, 2006; Zhao et al., 2019), for the distribution of returns (Ji & Lejeune, 2020), and for the Sharpe ratio under normality assumptions (Deng et al., 2013). Others have studied robustness for different risk measures (DeMiguel & Nogales, 2009; Huang et al., 2010; Kakouris & Rustem, 2014), or over multiple periods (Gülpınar & Rustem, 2007; Shen & Zhang, 2008). See Xidonas et al. (2020) for a review and Scutellà & Recchia (2013) for connections to robust statistics. However, existing robust approaches are sensitive to the uncertainty set's size and shape, and often need substantial computation, which does not scale (Scherer, 2007).

1.1. Our Approach

We consider portfolios that are convex combinations of the minimum-variance and maximum-Sharpe portfolios. These portfolios are created using mean and covariance estimates computed from n training samples of asset returns. We derive the optimal combination and show how it depends on n. For large n, there is no estimation error, and the optimal combination converges to the maximum-Sharpe portfolio. For small n, there is more estimation error. Since the minimum-variance portfolio ignores the mean, it avoids this source of error. So the optimal portfolio is close to the minimum-variance portfolio here.

We cannot find the optimal combination in practice because it depends on unknown parameters (the mean and covariance). For these parameters, we can only say what values are likely given the training data. So we build a robust portfolio that is never far from optimal for any of these likely parameter values.

We formalize this as a problem of minimizing regret under uncertainty. The regret of a portfolio is the Sharpe ratio of the optimal combination relative to the given portfolio. Thus our notion of regret adapts to the training data size n. We score each portfolio based on its worst-case regret, that is, the regret if the actual parameters happen to be the worst possible instance among all reasonable parameter values. Then, we select the portfolio with the smallest worst-case regret. This portfolio's Sharpe ratio is then guaranteed to be relatively close to the optimal combination portfolio for the actual parameter values.

Our algorithm to find the robust regret-minimizing portfolio is called AlphaRob. AlphaRob is based on our analysis of the regret, for which we only need the mild assumption that the distribution of asset returns has finite moments. In particular, we do not assume a parametric form, such as Gaussian returns. Further, we prove that estimation error affects our regret minimization problem only through a single random variable. So AlphaRob only needs an uncertainty set for this random variable, irrespective of the number of assets.

Apart from theoretical guarantees, AlphaRob also has several practical advantages. The simple one-dimensional uncertainty set makes AlphaRob fast and scalable. AlphaRob also needs no parameter-fitting; we use the same settings for all datasets. Our results are robust to small changes in these settings. Finally, we can adapt AlphaRob to situations where shorting is not allowed.

We compare AlphaRob against competing methods on both simulated data and 12 real-world datasets. We run tests with four choices of training sample sizes for each dataset, from $n = 15$ to $n = 120$. For the $n = 120$ setting, the Sharpe ratio of AlphaRob is, on average, 7.5% better than the next-best method. It can be $15\% - 21\%$ better for specific datasets, and the differences are statistically significant even after adjusting for transaction costs. AlphaRob outperforms the next-best combination portfolio by 28%, on average.

We can also explain the performance of the minimum-variance portfolio via its regret. The percentage difference in Sharpe ratios between AlphaRob and the minimum-variance portfolio has a Spearman correlation of around 0.95 with the latter's regret. We show that the regret increases with training size and the average level of risk-adjusted excess return across assets. For significantly correlated assets, like individual stocks, the regret is small. The minimum-variance portfolio is difficult to beat when its regret is less than 1.05, that is, when it is within 5% of the optimal. For less correlated assets, such as factor-based assets, the regret is much higher (1.24 for some datasets). Here, AlphaRob is up to 23% better than the minimum-variance portfolio.

We note that AlphaRob is not designed to counter outliers, and it assumes that training and test data have the same distribution. If these assumptions do not hold, a different analysis is required.

The rest of the paper is organized as follows. We formulate our problem in Section 2 and present our analysis in Section 3. Then, we propose our robust optimization method and the AlphaRob algorithm in Section 4. We validate AlphaRob on simulated data in Section 5 and on real-world datasets in Section 6. We conclude with a discussion in Section 7. The e-companion contains all proofs and supplementary material.

2. Formal setup

We want a policy that takes training data as input and outputs a portfolio. Suppose we are given a dataset $\mathcal{D}_n \in \mathbb{R}^{n\times p}$ of n independent and identically distributed sample returns for p assets. Then policy π constructs a portfolio $\pi(\mathcal{D}_n) =: \boldsymbol{w} \in \mathbb{R}^p$. Each component of $\boldsymbol{w}$ represents the amount allocated to one asset. The total portfolio must sum to one: $\boldsymbol{w}^T\mathbf{1} = 1$, where $\mathbf{1}$ is the all-ones vector. If we hold this portfolio for one timestep, and the asset returns in that timestep are given by $\boldsymbol{r} \in \mathbb{R}^p$, then the return of the portfolio is $\boldsymbol{w}^T\boldsymbol{r}$.

To measure the quality of π, we assume that training and test samples $\mathcal{D}_n$ and $\boldsymbol{r}$ are repeatedly generated from a distribution $f(.)$. The expected Sharpe ratio of π under $f(.)$ is given by

$$\mathcal{S}(\pi|f) = \frac{E_{\mathcal{D}_n,\boldsymbol{r}}\left[\pi(\mathcal{D}_n)^T\boldsymbol{r}\right]}{\sqrt{Var_{\mathcal{D}_n,\boldsymbol{r}}\left[\pi(\mathcal{D}_n)^T\boldsymbol{r}\right]}} \qquad \text{where } \mathcal{D}_n \sim f^{\otimes n},\ \boldsymbol{r} \sim f. \tag{1}$$

Here, $f^{\otimes n}(.)$ is the product distribution of n independent samples drawn from the distribution $f(.)$. The policy $\pi^\star$ with the highest expected Sharpe ratio is well-known. Let $\boldsymbol{\mu}$ and Σ be the mean and covariance matrix of the distribution $f(.)$. If $\mathbf{1}^T\Sigma^{-1}\boldsymbol{\mu} > 0$, then[1]

$$\pi^\star(\mathcal{D}_n) := \boldsymbol{w}_{\text{MS}} = \frac{\Sigma^{-1}\boldsymbol{\mu}}{\mathbf{1}^T\Sigma^{-1}\boldsymbol{\mu}}. \tag{2}$$

This is a fixed policy that does not depend on $\mathcal{D}_n$. So the denominator of Eq. 1 depends only on the variability of the test sample $\boldsymbol{r}$. But, in practice, we do not know $\boldsymbol{\mu}$ and Σ, and we can only construct policies that depend on

$\mathcal{D}_n$. The variability due to $\mathcal{D}_n$ can dominate that due to $\boldsymbol{r}$, especially for small sample sizes. Thus, finding a policy with a high expected Sharpe ratio under an unknown $f(.)$ requires a different analysis than for a known $f(.)$.

When $f(.)$ is unknown, we must rely on sample estimates. The focus of this paper is on countering the errors in estimating $\boldsymbol{\mu}$. Hence, for our analysis, we consider portfolios constructed using the sample mean $\hat{\boldsymbol{\mu}}$ and the true covariance Σ. The assumption of a known Σ allows us to isolate and analyze the effect of estimation error in $\hat{\boldsymbol{\mu}}$. In practice, we will use a robust estimator of Σ. We will show in Section 5 that this does not make a qualitative difference to our results.

We consider portfolios that are convex combinations of the minimum-variance portfolio and the sample-based maximum-Sharpe portfolio:

$$\pi_\beta(\mathcal{D}_n) = (1-\beta)\cdot \boldsymbol{w}_{\text{MV}} + \beta \cdot \hat{\boldsymbol{w}}_{\text{MS}}, \quad \text{where } \boldsymbol{w}_{\text{MV}} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}} \text{ and } \hat{\boldsymbol{w}}_{\text{MS}} = \frac{\Sigma^{-1}\hat{\boldsymbol{\mu}}}{\mathbf{1}^T\Sigma^{-1}\hat{\boldsymbol{\mu}}}. \tag{3}$$

Note that the expected Sharpe ratio $\mathcal{S}(\pi_\beta(\mathcal{D}_n)|f)$ need not be monotonic in β (Eq. 1). So the β that achieves the highest Sharpe ratio could be anywhere in the interval $[0, 1]$. The best β depends on the sample size n. For small n, $\|\hat{\boldsymbol{w}}_{\text{MS}} - \boldsymbol{w}_{\text{MS}}\|$ can be large, and the expected Sharpe ratio of $\hat{\boldsymbol{w}}_{\text{MS}}$ may be much lower than $\boldsymbol{w}_{\text{MS}}$. In this case, we can match the minimum-variance portfolio $\boldsymbol{w}_{\text{MV}}$ by setting $\beta = 0$. As noted earlier, $\boldsymbol{w}_{\text{MV}}$ performs very well empirically. As n increases, the sample mean $\hat{\boldsymbol{\mu}}$ converges almost surely to the true mean $\boldsymbol{\mu}$, so $\hat{\boldsymbol{w}}_{\text{MS}} \to \boldsymbol{w}_{\text{MS}}$. When $n \to \infty$, $\pi(\mathcal{D}_n)$ traces out the efficient frontier as β is varied from zero to one. Here, $\beta = 1$ gives the optimal maximum-Sharpe portfolio.

We note that alternative combinations are also possible. However, the combination of $\boldsymbol{w}_{\text{MV}}$ and $\hat{\boldsymbol{w}}_{\text{MS}}$ has several advantages. As Kan & Zhou (2007) note, $\boldsymbol{w}_{\text{MV}}$ depends only on the covariance matrix Σ and not on $\hat{\boldsymbol{\mu}}$, so it can be estimated more accurately than $\hat{\boldsymbol{w}}_{\text{MS}}$. Prior work also shows that $\boldsymbol{w}_{\text{MV}}$ by itself often performs better than alternatives, such as the equal-weighted portfolio (Ledoit & Wolf, 2017; Zhao et al., 2019). Finally, since $\hat{\boldsymbol{w}}_{\text{MS}} \to \boldsymbol{w}_{\text{MS}}$ as $n \to \infty$, $\hat{\boldsymbol{w}}_{\text{MS}}$ must be one endpoint of the combination to guarantee asymptotic optimality.

Notation. For ease of analysis, we define the following:

$$\hat{\boldsymbol{z}} := \Sigma^{-1/2}\hat{\boldsymbol{\mu}}, \quad \boldsymbol{z} := \Sigma^{-1/2}\boldsymbol{\mu}, \quad \boldsymbol{m} := \Sigma^{-1/2}\boldsymbol{1}, \quad s := \boldsymbol{z}^T\boldsymbol{m}, \quad \theta := \frac{s}{\|\boldsymbol{m}\|\|\boldsymbol{z}\|} = \frac{\boldsymbol{m}^T\boldsymbol{z}}{\|\boldsymbol{m}\|\|\boldsymbol{z}\|}.$$

We will occasionally refer to a portfolio $\boldsymbol{w}$ with the vector $\boldsymbol{x}$ where $\boldsymbol{w} = \Sigma^{-1/2}\boldsymbol{x}/(\boldsymbol{1}^T\Sigma^{-1/2}\boldsymbol{x})$. For instance, $\boldsymbol{x} = \boldsymbol{m} \Rightarrow \boldsymbol{w} = \boldsymbol{w}_{\text{MV}}$, and $\boldsymbol{x} = \hat{\boldsymbol{z}} \Rightarrow \boldsymbol{w} = \hat{\boldsymbol{w}}_{\text{MS}}$. We will use $\mathcal{S}_w(\boldsymbol{w})$ and $\mathcal{S}_x(\boldsymbol{x})$ to refer to the expected Sharpe ratio of a policy that constructs the portfolio $\boldsymbol{w}$ corresponding to $\boldsymbol{x}$. We can show that

$$\|\boldsymbol{z}\| = \mathcal{S}_w(\boldsymbol{w}_{\text{MS}}), \qquad \|\boldsymbol{m}\| = \frac{1}{\sqrt{Var(\boldsymbol{w}_{\text{MV}})}}, \qquad -1 \leq \theta = \frac{\mathcal{S}_w(\boldsymbol{w}_{\text{MV}})}{\mathcal{S}_w(\boldsymbol{w}_{\text{MS}})} \leq 1. \tag{4}$$

Thus, $\boldsymbol{m}$ and $\boldsymbol{z}$ are vectors whose lengths and angles correspond to the variances and Sharpe ratios of $\boldsymbol{w}_{\text{MV}}$ and $\boldsymbol{w}_{\text{MS}}$. Since $\boldsymbol{w}_{\text{MV}}$ often has a high Sharpe ratio on real-world datasets, we expect θ to be a positive fraction not close to zero.

3. Analysis of the Expected Sharpe Ratio

We start our analysis with the problem of finding the optimal combination portfolio. Observe that

$$\mathcal{S}_x(\boldsymbol{m} + \alpha\hat{\boldsymbol{z}}) = \mathcal{S}_w\left(\frac{\Sigma^{-1/2}(\boldsymbol{m} + \alpha\hat{\boldsymbol{z}})}{\boldsymbol{1}^T\Sigma^{-1/2}(\boldsymbol{m} + \alpha\hat{\boldsymbol{z}})}\right) = \mathcal{S}_w\left(\frac{\boldsymbol{w}_{\text{MV}} \cdot (\boldsymbol{1}^T\Sigma^{-1/2}\boldsymbol{m}) + \alpha \cdot \hat{\boldsymbol{w}}_{\text{MS}} \cdot (\boldsymbol{1}^T\Sigma^{-1/2}\hat{\boldsymbol{z}})}{\boldsymbol{1}^T\Sigma^{-1/2}(\boldsymbol{m} + \alpha\hat{\boldsymbol{z}})}\right)$$

$$= \mathcal{S}_w\left((1-\beta) \cdot \boldsymbol{w}_{\text{MV}} + \beta \cdot \hat{\boldsymbol{w}}_{\text{MS}}\right),$$

$$\text{where } 1 - \beta = \left(1 + \alpha \cdot \frac{\boldsymbol{1}^T\Sigma^{-1/2}\hat{\boldsymbol{z}}}{\boldsymbol{1}^T\Sigma^{-1/2}\boldsymbol{m}}\right)^{-1} = \left(1 + \alpha \cdot \left(\boldsymbol{w}_{\text{MV}}^T\hat{\boldsymbol{\mu}}\right)\right)^{-1}. \tag{5}$$

So we will solve the following problem[2]:

$$\text{find the optimal } \alpha^\star = \arg\max_{\alpha \geq 0} \mathcal{S}_x(\boldsymbol{m} + \alpha \cdot \hat{\boldsymbol{z}}). \tag{6}$$

This can be translated into the corresponding $\beta^\star$ via Eq. 5. The next theorem analyzes the expected Sharpe ratio $\mathcal{S}_x(\boldsymbol{x})$ for any $\boldsymbol{x}$.

Theorem 1. *Let $\boldsymbol{x}$ be a function of $\mathcal{D}_n$ such that $\boldsymbol{m}^T\boldsymbol{x} \neq 0$ almost surely. Then,*

$$\mathcal{S}_x(\boldsymbol{x}) = \frac{E_{\mathcal{D}_n}\left[\frac{\boldsymbol{z}^T\boldsymbol{x}}{\boldsymbol{m}^T\boldsymbol{x}}\right]}{\sqrt{E_{\mathcal{D}_n}\left[\frac{\boldsymbol{x}^T\boldsymbol{x}}{(\boldsymbol{m}^T\boldsymbol{x})^2}\right] + Var_{\mathcal{D}_n}\left(\frac{\boldsymbol{z}^T\boldsymbol{x}}{\boldsymbol{m}^T\boldsymbol{x}}\right)}} \tag{7}$$

Recall that the expected Sharpe ratio depends on both the test sample $\boldsymbol{r}$ and the training samples $\mathcal{D}_n$ (Eq. 1). The two terms in the denominator in Eq. 7 account for variance due to these two sources of randomness. For example, consider the minimum-variance portfolio $\boldsymbol{w}_{\text{MV}}$. This corresponds to $\boldsymbol{x} = \boldsymbol{m} = \Sigma^{-1/2}\mathbf{1}$, which does not depend on $\mathcal{D}_n$ (we take Σ as given). So the second term in the denominator vanishes. The small denominator improves the expected Sharpe ratio of $\boldsymbol{w}_{\text{MV}}$. Next, we consider the case of $\boldsymbol{x} = \boldsymbol{m} + \alpha \cdot \hat{\boldsymbol{z}}$.

Theorem 2. *Assume that the distribution $f(.)$ of asset returns has finite moments. Let $\boldsymbol{\chi} = \boldsymbol{m} + \alpha \cdot \boldsymbol{z}$. Then,*

$$\mathcal{S}_x\left(\boldsymbol{m} + \alpha \cdot \hat{\boldsymbol{z}}\right) = \frac{(\boldsymbol{z}^T\boldsymbol{\chi}) + \dfrac{\alpha^3}{n} \cdot \dfrac{\left(\|\boldsymbol{m}\|^2\|\boldsymbol{z}\|^2 - s^2\right)}{\left(\boldsymbol{m}^T\boldsymbol{\chi}\right)^2}}{\sqrt{\|\boldsymbol{\chi}\|^2 + \dfrac{\alpha^2}{n}\left[(p-1) + \dfrac{\left(\|\boldsymbol{m}\|^2\|\boldsymbol{z}\|^2 - s^2\right)\cdot\left(3\alpha^2 + \|\boldsymbol{\chi}\|^2\right)}{\left(\boldsymbol{m}^T\boldsymbol{\chi}\right)^2}\right]}} + o\left(\frac{1}{n}\right), \tag{8}$$

Theorem 3. *Suppose $\theta > 0$ and $\alpha \geq 0$. If $n \gg \frac{1/\theta^2 - 1}{\|\boldsymbol{z}\|^2}$ and $p \gg \frac{(1-\theta^2)(\|\boldsymbol{z}\|^2+3)}{\theta^2}$, then Eq. 8 can be simplified to*

$$\mathcal{S}_x\left(\boldsymbol{m} + \alpha \cdot \hat{\boldsymbol{z}}\right) \approx \frac{\boldsymbol{z}^T\boldsymbol{\chi}}{\sqrt{\|\boldsymbol{\chi}\|^2 + \dfrac{\alpha^2 \cdot p}{n}}} = \frac{\|\boldsymbol{z}\|}{\sqrt{1 + q\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right)}} \tag{9}$$

$$\text{where } q\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right) = \frac{\|\boldsymbol{m}\|^2\left(1-\theta^2\right) + \alpha^2(p/n)}{\left(\|\boldsymbol{m}\|\theta + \alpha\|\boldsymbol{z}\|\right)^2}. \tag{10}$$

We can use this simpler form as long as n and p are large enough. The conditions listed in Theorem 3 are reasonable for our datasets. For example, for the Fama-French equally-weighted dataset with 100 assets, these conditions only require $n \gg 2.5$ and $p \gg 4.5$. We will use this simpler formula henceforth. With this simplification, the optimal $\alpha^\star$ that achieves the highest expected Sharpe ratio (Eq. 6) has a simple form.

Theorem 4. *Suppose $0 < \theta < 1$. Define $\Delta = \mathcal{S}_w\left(\boldsymbol{w}_{MS}\right)^2 - \mathcal{S}_w\left(\boldsymbol{w}_{MV}\right)^2$. For $\alpha \geq 0$, the expected Sharpe ratio formula of Eq. 9 is unimodal, and achieves its* *maximum at*

$$\alpha^\star = \frac{n\|\boldsymbol{m}\|\|\boldsymbol{z}\|}{p}\left(\frac{1}{\theta} - \theta\right), \tag{11}$$

$$\textit{or equivalently, } \beta^\star = \frac{n}{p}\cdot\Delta\cdot\left(\frac{\boldsymbol{w}_{MV}^T\hat{\boldsymbol{\mu}}}{\boldsymbol{w}_{MV}^T\boldsymbol{\mu}}\right).$$

The corresponding expected Sharpe ratio is

$$\mathcal{S}_w\left((1-\beta^\star)\cdot\boldsymbol{w}_{MV} + \beta^\star\cdot\hat{\boldsymbol{w}}_{MS}\right) = \frac{\mathcal{S}_w\left(\boldsymbol{w}_{MS}\right)}{\sqrt{1 + \frac{1}{\frac{\mathcal{S}_w(\boldsymbol{w}_{MV})^2}{\Delta} + \frac{n}{p}\cdot\mathcal{S}_w(\boldsymbol{w}_{MS})^2}}}.$$

Recall that $\alpha^\star$ represents the optimal tilt away from $\boldsymbol{w}_{\text{MV}}$ towards $\hat{\boldsymbol{w}}_{\text{MS}}$. When the sample size n increases, we have more confidence in the estimated returns $\hat{\boldsymbol{\mu}}$. So $\alpha^\star$ also increases. In the limit of $n\to\infty$, we recover the optimal solution $\boldsymbol{w}_{\text{MS}}$. Also, $\alpha^\star$ increases with $\|\boldsymbol{z}\|$ because $\|\boldsymbol{z}\| = \mathcal{S}_w\left(\boldsymbol{w}_{\text{MS}}\right)$ (Eq. 4). So, a higher value of $\|\boldsymbol{z}\|$ implies greater benefits in moving from $\boldsymbol{w}_{\text{MV}}$ towards $\hat{\boldsymbol{w}}_{\text{MS}}$. Similarly, when $\theta = \mathcal{S}_w\left(\boldsymbol{w}_{\text{MV}}\right)/\mathcal{S}_w\left(\boldsymbol{w}_{\text{MS}}\right)$ is large, $\boldsymbol{w}_{\text{MV}}$ is almost as good as $\boldsymbol{w}_{\text{MS}}$. So there is less need to tilt towards $\hat{\boldsymbol{w}}_{\text{MS}}$. Hence, $\alpha^\star$ decreases with increasing θ. Finally, $\alpha^\star$ is inversely proportional to the number of assets p since each new asset adds to the variance in the returns, and hence increases the variability of $\hat{\boldsymbol{w}}_{\text{MS}}$.

4. Robust Optimization

To calculate the optimal $\alpha^\star$, we need to know $\|\boldsymbol{m}\|$, $\|\boldsymbol{z}\|$, and θ. Since we assume that Σ is known, we know $\|\boldsymbol{m}\| = \sqrt{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}$. But we only have the estimated quantities $\hat{\boldsymbol{z}}$ and $\hat{\theta}$. We cannot plug-in $\|\hat{\boldsymbol{z}}\|$ and $\hat{\theta}$ in place of $\|\boldsymbol{z}\|$ and θ because the former are *biased* estimates of the latter.

Theorem 5. *Assume that the returns distribution has finite moments. Then,*

$$E\|\hat{\boldsymbol{z}}\| = \|\boldsymbol{z}\| + \frac{p-1}{2n\|\boldsymbol{z}\|} + o\left(\frac{1}{n}\right), \qquad E\hat{\theta} = \theta\left(1 - \frac{p-1}{2n\|\boldsymbol{z}\|^2}\right) + o\left(\frac{1}{n}\right).$$

So, we expect $\|\hat{\boldsymbol{z}}\| > \|\boldsymbol{z}\|$ and $\hat{\theta} < \theta$. This implies that we are likely to overestimate $\alpha^\star$ if we plug in $\|\hat{\boldsymbol{z}}\|$ and $\hat{\theta}$ in Eq. 11. Even if we constructed unbiased estimators of $\|\boldsymbol{z}\|$ and θ, these would still have estimation errors. Plugging estimates into Eq. 11 does not account for these errors.

We propose a robust optimization to solve this problem. First, we construct confidence intervals $\mathcal{I}_{\|\boldsymbol{z}\|} = [\|\boldsymbol{z}\|_-, \|\boldsymbol{z}\|_+]$ and $\mathcal{I}_\theta = [\theta_-, \theta_+]$, with $0 \leq \|\boldsymbol{z}\|_- \leq \|\boldsymbol{z}\|_+$ and $0 < \theta_- \leq \theta_+ \leq 1$. We can do this via standard techniques such as the jackknife or the bootstrap. Then, we find a robust choice for α:

$$\alpha_{rob} = \arg\max_{\alpha \geq 0} \min_{\substack{\|\boldsymbol{z}\| \in \mathcal{I}_{\|\boldsymbol{z}\|} \\ \theta \in \mathcal{I}_\theta}} \frac{\mathcal{S}_x\left(\boldsymbol{m} + \alpha \cdot \hat{\boldsymbol{z}}\right)}{\mathcal{S}_x\left(\boldsymbol{m} + \alpha^\star \cdot \hat{\boldsymbol{z}}\right)}, \tag{12}$$

where $\alpha^\star$ is a function of $\|\boldsymbol{z}\|$ and θ (Eq. 11). The ratio in Eq. 12 is the "normalized" expected Sharpe ratio. The inverse of this ratio represents our regret if we use α instead of $\alpha^\star$. Thus, for a given $\|\boldsymbol{z}\|$ and θ, the optimal $\alpha^\star$ achieves a regret of one, and all other choices of α incur a higher regret. By setting $\alpha = \alpha_{rob}$, we incur the least regret even if the actual values of $\|\boldsymbol{z}\|$ and θ are the worst possible values in $\mathcal{I}_{\|\boldsymbol{z}\|}$ and $\mathcal{I}_\theta$.

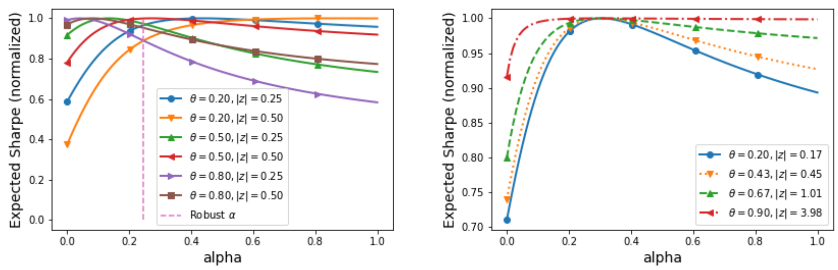
Remark 1. *Setting $\alpha = 0$ yields the minimum-variance portfolio. Its regret is*

$$\frac{1}{\left(Regret\left(\boldsymbol{w}_{MV}\right)\right)^2} = \frac{1}{1+\nu} \cdot 1 + \frac{\nu}{1+\nu} \cdot \theta^2; \qquad \nu = n \cdot \left(\frac{1}{\theta^2} - 1\right) \cdot \frac{\|\boldsymbol{z}\|^2}{p}$$

Thus, Regret($\boldsymbol{w}_{MV}$) increases with ν. When $n = 0$ (so $\nu = 0$), $\boldsymbol{w}_{MV}$ is optimal (regret=1). As $n \to \infty$, the regret gradually increases to $1/\theta = \mathcal{S}_w\left(\boldsymbol{w}_{MS}\right)/\mathcal{S}_w\left(\boldsymbol{w}_{MV}\right)$. We can further expand the parameter ν as

$$\nu = n \times \underbrace{\left(\frac{\mathcal{S}_w\left(\boldsymbol{w}_{MS}\right)^2}{\mathcal{S}_w\left(\boldsymbol{w}_{MV}\right)^2} - 1\right)}_{\substack{\text{relative difference} \\ \text{when } n \to \infty}} \times \underbrace{\left(\frac{\sum_{i=1}^p \gamma_i/\sigma_{\epsilon_i}^2 \cdot \mu_i}{\sum_{i=1}^p \gamma_i/\sigma_{\epsilon_i}^2}\right)}_{\substack{\text{weighted average of} \\ \text{mean asset returns}}} \times \underbrace{\left(\frac{\sum_{i=1}^p \gamma_i/\sigma_{\epsilon_i}^2}{p}\right)}_{\substack{\text{average of excess return} \\ \text{over excess variance}}},$$

where γ_i and $\sigma_{\epsilon_i}^2$ are the excess expected returns and excess variance for asset i left over after hedging its returns using all other assets (Stevens, 1998). The last term measures how uncorrelated the assets are. It can be high if the assets represent different factors, since factors are designed to capture distinct information.



(a) Different modes

(b) Same mode

Figure 1: *Example curves of the normalized expected Sharpe ratio for different values of $\|\boldsymbol{z}\|$ and θ:* (a) The robust solution α_{rob} is such that for any $\alpha \neq \alpha_{rob}$, there is some $(\|\boldsymbol{z}\|, \theta)$ for which the normalized expected Sharpe ratio is smaller for α than for α_{rob}. (b) When multiple curves have the same mode, the curve with the smallest θ is the lowest curve.

So for factor-based assets, Regret($\boldsymbol{w}_{MV}$) can be high, and our robust optimization can be much better than the minimum-variance portfolio. But when the assets are individual stocks, the minimum-variance portfolio may be close to optimal.

Figure 1(a) plots the normalized expected Sharpe ratio against α for several choices of $(\|\boldsymbol{z}\|, \theta)$. The curves are unimodal, by Theorem 4. Each curve attains a maximum value of one due to the normalization. Intuitively, α_{rob} is at the intersection of two curves such that all other curves lie above the intersection point. We find α_{rob} in two steps. First, in Section 4.1, we simplify the robust optimization. We prove that many $(\|\boldsymbol{z}\|, \theta)$ are redundant for the robust optimization (see Figure 1(b)). By pruning these from $\mathcal{I}_{\|\boldsymbol{z}\|} \times \mathcal{I}_\theta$, we get a one-dimensional uncertainty interval. Then, in Section 4.2, we present our AlphaRob algorithm to solve the robust optimization of Eq. 12. AlphaRob generates a sequence of α values. We prove that this sequence converges to α_{rob}.

4.1. Simplification of the robust optimization

We will first apply a monotonic transform to the normalized expected Sharpe ratio, to simplify analysis. Define

$$\begin{aligned} h\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right) &= 2 \cdot \log\left(\frac{\mathcal{S}_x\left(\boldsymbol{m} + \alpha \cdot \hat{\boldsymbol{z}}\right)}{\mathcal{S}_x\left(\boldsymbol{m} + \alpha^\star \cdot \hat{\boldsymbol{z}}\right)}\right) \\ &= \log(1 + q\left(\alpha^\star \mid \|\boldsymbol{z}\|, \theta\right)) - \log(1 + q\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right)), \end{aligned} \tag{13}$$

with $q\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right)$ as defined in Eq. 10. Note that $\alpha^\star$ is a function of $(\|\boldsymbol{z}\|, \theta)$, and $h\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right) \leq 0$ everywhere. The, Eq. 12 is equivalent to

$$\alpha_{rob} = \arg\max_{\alpha \geq 0} \min_{\substack{\|\boldsymbol{z}\| \in \mathcal{I}_{\|\boldsymbol{z}\|} \\ \theta \in \mathcal{I}_\theta}} h\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right). \tag{14}$$

Now, consider the set of $(\|\boldsymbol{z}\|, \theta)$ pairs for which the curves $h\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right)$ have their mode at a given location $\alpha^\star = \gamma$:

$$\Gamma_\gamma := \left\{(\|\boldsymbol{z}\|, \theta) \mid \|\boldsymbol{z}\| \in \mathcal{I}_{\|\boldsymbol{z}\|}, \theta \in \mathcal{I}_\theta, \alpha^\star(\|\boldsymbol{z}\|, \theta) = \gamma\right\}, \tag{15}$$

where the notation emphasizes that $\alpha^\star$ is a function of $\|\boldsymbol{z}\|$ and θ. We now prove that among all such curves, the one with the smallest θ is the lowest curve.

Theorem 6. *Let* $\theta_\gamma = \min\left\{\theta \in \mathcal{I}_\theta \mid \exists \|\boldsymbol{z}\| \in \mathcal{I}_{\|\boldsymbol{z}\|} \text{ such that } (\|\boldsymbol{z}\|, \theta) \in \Gamma_\gamma\right\}$. *Then,*

- *The set* $\{\|\boldsymbol{z}\| \mid (\|\boldsymbol{z}\|, \theta_\gamma) \in \Gamma_\gamma\}$ *has only one element; call it* $\|\boldsymbol{z}\|_\gamma$.
- *For any* $\alpha \geq 0$, $h\left(\alpha \mid \|\boldsymbol{z}\|_\gamma, \theta_\gamma\right) \leq h\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right)$ *for all* $(\|\boldsymbol{z}\|, \theta) \in \Gamma_\gamma$.

Now, define $h\left(\alpha \mid \gamma\right) = h\left(\alpha \mid \|\boldsymbol{z}\|_\gamma, \theta_\gamma\right)$, where $\|\boldsymbol{z}\|_\gamma$ and θ_γ are uniquely defined by Theorem 6. So $h\left(\alpha \mid \gamma\right)$ represents the lowest curve among all curves whose mode is at $\alpha^\star = \gamma$. Recall that $\alpha^\star$ is monotonically increasing in $\|\boldsymbol{z}\|$ and decreasing in θ (Eq. 11). So γ takes values in the interval $\mathcal{I}_\gamma = \left[\alpha^\star\left(\|\boldsymbol{z}\|_-, \theta_+\right), \alpha^\star\left(\|\boldsymbol{z}\|_+, \theta_-\right)\right]$. By Theorem 6, for any $\alpha \geq 0$,

$$\min_{\substack{\|\boldsymbol{z}\| \in \mathcal{I}_{\|\boldsymbol{z}\|} \\ \theta \in \mathcal{I}_\theta}} h\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right) = \min_{\gamma \in \mathcal{I}_\gamma} h\left(\alpha \mid \gamma\right).$$

So, the robust optimization problem of Eq. 14 simplifies to

$$\alpha_{rob} = \arg\max_{\alpha \geq 0} \min_{\gamma \in \mathcal{I}_\gamma} h\left(\alpha \mid \gamma\right). \tag{16}$$

4.2. Algorithm

We now present our algorithm, called AlphaRob, to solve Eq. 16 and find α_{rob}. Intuitively, α_{rob} is the value of α at which two curves intersect, and all other curves lie above that intersection point. AlphaRob starts with an initial guess about these two curves and then iteratively refines them. Now the curves $h(\alpha \mid \gamma)$ are indexed by their mode γ. We initialize AlphaRob with two curves whose modes are at the ends of the interval $\mathcal{I}_\gamma$. Then we iterate over two steps (see Figure 2). First, we find the intersection $\alpha^{(t)}$ of the curves. Then we find two curves with modes $\gamma_{lo}^{(t+1)} \leq \alpha^{(t)}$ and $\gamma_{hi}^{(t+1)} \geq \alpha^{(t)}$, that incur the largest regret at $\alpha^{(t)}$. These become our guesses for the next iteration. Both steps only need a grid search on the interval $\mathcal{I}_\gamma$. We repeat these steps until the sequence $\alpha^{(t)}$ converges. Algorithm 1 shows these steps. The next theorem shows that AlphaRob converges to the solution α_{rob} of Eq. 16.

Theorem 7. *Suppose $\theta_- > 0$ and $\|\boldsymbol{z}\|_+ < \infty$. Let $\kappa_{rob} = \min_{\gamma \in \mathcal{I}_\gamma} h(\alpha_{rob} \mid \gamma)$ and $\kappa^{(t)} = h\left(\alpha^{(t)} \mid \gamma_{lo}^{(t)}\right)$. Then, for all $t' \geq 1$, $\kappa_{rob} \leq \kappa^{(t')} \leq \kappa^{(1)} < 0$ unless the interval $\mathcal{I}_\gamma$ is degenerate. Further, the iterates $\kappa^{(t)}$ converge towards κ_{rob} as*

$$\kappa^{(t)} - \kappa_{rob} \leq \left(\kappa^{(1)} - \kappa_{rob}\right) \cdot \left(1 + C \cdot |\kappa^{(1)}|\right)^{-(t-1)},$$

for some constant $C > 0$ that depends on $\mathcal{I}_{\|\boldsymbol{z}\|}$ and $\mathcal{I}_\theta$.

AlphaRob needs the interval $\mathcal{I}_\gamma$ as input. $\mathcal{I}_\gamma$ is a function of $\mathcal{I}_{\|\boldsymbol{z}\|}$ and $\mathcal{I}_\theta$. We use 3σ confidence intervals for both $\mathcal{I}_{\|\boldsymbol{z}\|}$ and $\mathcal{I}_\theta$. We estimate the standard deviation σ via the jackknife, using only the training samples. We center the interval $\mathcal{I}_{\|\boldsymbol{z}\|}$ at $\sqrt{\|\hat{\boldsymbol{z}}\|^2 - p/n}$ since $E[\|\hat{\boldsymbol{z}}\|^2] = \|\boldsymbol{z}\| + p/n$. For $\mathcal{I}_\theta$, we center it at an unbiased estimate of θ, again computed by the jackknife. Note that confidence intervals can also be built using other methods, and can incorporate any prior beliefs. The choice of confidence intervals is orthogonal to our algorithm.

Remark 2. *Algorithm 1 is designed for the simplified expected Sharpe ratio formula (Eq. 9). In rare cases, this can lead to $\alpha^{(t)}$ values that are too large. So*

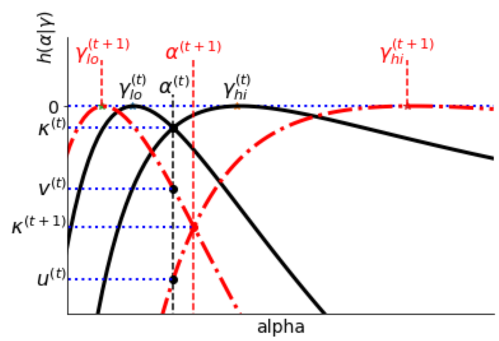


Figure 2: *Description of* ALPHAROB*:* Iteration t starts with two curves whose modes are at $\gamma_{lo}^{(t)}$ and $\gamma_{hi}^{(t)}$. They intersect at $\alpha^{(t)}$, where both curves have a value $\kappa^{(t)}$. Then, we find new curves with modes $\gamma_{lo}^{(t+1)} \le \alpha^{(t)}$ and $\gamma_{hi}^{(t+1)} \ge \alpha^{(t)}$ that have the lowest values at $\alpha^{(t)}$. This is shown by $u^{(t)}$ and $v^{(t)}$. In the next iteration, we find the intersection of these curves at $\alpha^{(t+1)}$, with value $\kappa^{(t+1)}$. We repeat until convergence.

Algorithm 1 Solution to the Robust Optimization

function ALPHAROB(n, p, $\|\boldsymbol{m}\|$, $\mathcal{I}_\gamma$)‘
    $t \leftarrow 1$
    $\gamma_{lo}^{(1)} \leftarrow \min(\mathcal{I}_\gamma)$
    $\gamma_{hi}^{(1)} \leftarrow \max(\mathcal{I}_\gamma)$
    repeat
        $\alpha^{(t)} \leftarrow \alpha \in \left[\gamma_{lo}^{(t)}, \gamma_{hi}^{(t)}\right]$ s.t. $h\left(\alpha \mid \gamma_{lo}^{(t)}\right) = h\left(\alpha \mid \gamma_{hi}^{(t)}\right)$ ▷ Find intersection of curves
        $\gamma_{lo}^{(t+1)} \leftarrow \arg\min_{\gamma \le \alpha^{(t)}} h\left(\alpha^{(t)} \mid \gamma\right)$ ▷ Update curves
        $\gamma_{hi}^{(t+1)} \leftarrow \arg\min_{\gamma \ge \alpha^{(t)}} h\left(\alpha^{(t)} \mid \gamma\right)$
        $t \leftarrow t + 1$
    until the sequence $\alpha^{(t)}$ converges
    return $\alpha^{(t)}$
end function

we do a post-processing step where we consider the full formula for the variance of the combination portfolio (Eq. 8). We use this to recalculate the intersections $\alpha^{(t)}$ *and also to choose the best* α_{rob} *among the* $\alpha^{(t)}$.

Remark 3. *As n increases, we expect $\mathcal{I}_{\|\boldsymbol{z}\|}$ and $\mathcal{I}_\theta$ to shrink as $O(\sqrt{n})$ by the Central Limit Theorem. Surprisingly, $\mathcal{I}_\gamma$ may become wider with n. This is because the formula for $\alpha^\star$ grows linearly with n (Eq. 11). So the interval $\mathcal{I}_\gamma$ is of the form $n \cdot O(1 \pm 1/\sqrt{n})$. Note that as $n \to \infty$, both ends of the interval $\mathcal{I}_\gamma$ grow with n. So, $\alpha_{rob} \to \infty$ and we recover the optimal maximum-Sharpe portfolio in the limit.*

5. Experiments on Simulated Datasets

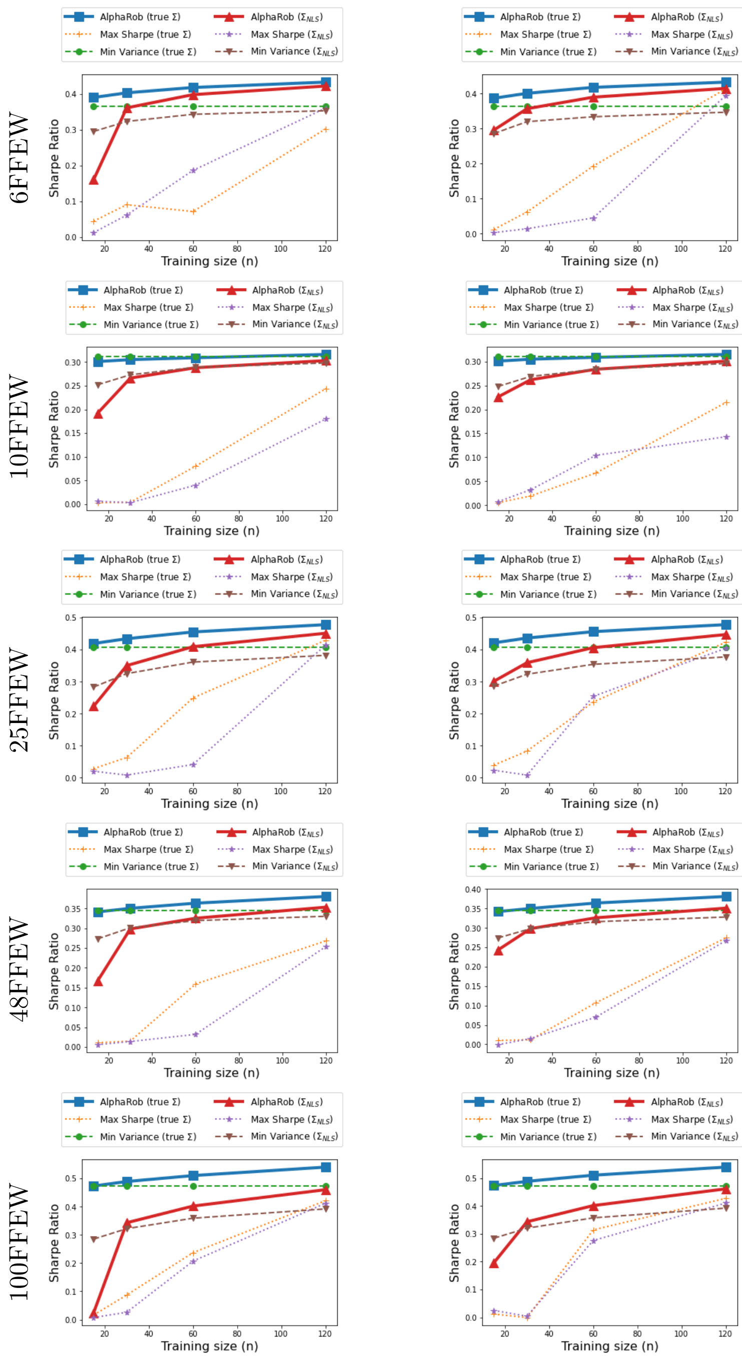
We will now test AlphaRob on simulated data where the optimal $\alpha^\star$ is known. We will vary the training length n, the number of assets p, and the distribution of returns. We will also connect the performance of the minimum-variance portfolio to its regret, and explore no-shorting variants of AlphaRob.

Experiment setup: We generate n i.i.d. sample returns from either a Gaussian distribution or a heavy-tailed t-distribution with a given mean $\boldsymbol{\mu}$ and covariance Σ. We run AlphaRob over these n samples to calculate α_{rob}. We repeat this 1000 times. This gives the empirical distribution for α_{rob}. We report the Sharpe ratio for combination portfolios with α drawn from this empirical distribution.

To set $(\boldsymbol{\mu}, \Sigma)$ to realistic values, we compute them from Fama-French datasets. These have monthly asset returns from July 1963 until July 2015. Each asset is a portfolio of firms either weighted equally or by value. For example, the dataset `10FFEW` has ten assets built from equally-weighted portfolios of firms grouped by industry. Similarly, each asset in `10FFVW` is a value-weighted portfolio of firms. The list of all datasets is in the supplementary material. For each dataset, we find the mean $\boldsymbol{\mu}$ and covariance Σ of the monthly returns, and use this $(\boldsymbol{\mu}, \Sigma)$ to simulate sample returns for our experiment.

AlphaRob also needs a covariance matrix as input. We run tests with both the true Σ and a robust covariance estimate Σ_{NLS} (Ledoit & Wolf, 2017).

Sharpe Ratio Results: Figure 3 plots the Sharpe ratios of AlphaRob, the minimum-variance portfolio, and the estimated maximum-Sharpe portfolio for five datasets. Results on another five datasets are similar, and are presented in

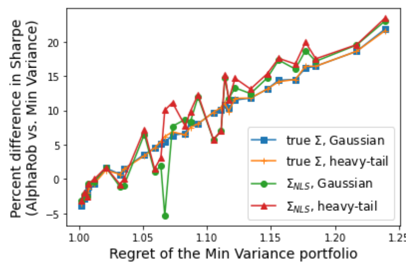


(a) Gaussian distribution (b) t distribution

Figure 3: Sharpe ratios after simulating returns for different number of assets. AlphaRob is comparable or better than other methods given $n \geq 30$ training samples.

	n=15	n=30	n=60	n=120
Σ_{NLS}, Gaussian	0.83	3.70	7.60	12.44
Σ_{NLS}, t dist.	0.82	3.70	7.72	12.38
true Σ, Gaussian	-26.51	2.99	8.36	12.67
true Σ, t dist.	-10.03	5.09	9.02	12.94

(a) Percentage difference in Sharpe (AlphaRob versus minimum-variance)



(b) Sharpe difference versus regret ($n \geq 30$)

Figure 4: *Sharpe Ratio of* AlphaRob *versus minimum-variance on simulated data:* (a) The percent difference, averaged over 10 datasets, grows with the training size n. AlphaRob is better for $n \geq 30$. (b) The percent difference is positively correlated with the regret of the minimum-variance portfolio. Each point corresponds to one dataset and one training size.

the supplementary material. The Sharpe ratio for AlphaRob is always better than the estimated maximum-Sharpe portfolio. Apart from the smallest training size ($n = 15$), AlphaRob is comparable or better than the minimum-variance portfolio. When $n = 120$, AlphaRob is 12% better; even for $n = 30$ training samples, AlphaRob is $3\% - 5\%$ better (Table 4(a)).

The predictive power of regret: Figure 4(b) shows a strong positive correlation (Spearman rank correlation > 0.93, p-value $< 10^{-13}$) between the outperformance of AlphaRob and the regret of the minimum-variance portfolio (see Remark 1). Thus, for any dataset and training size, a single number — the regret of the minimum-variance portfolio — tells us how well the latter performs, and how far AlphaRob can improve upon it.

When the regret is small (close to 1), the minimum-variance portfolio is nearly optimal. So even small errors can make AlphaRob's Sharpe ratio worse than the minimum-variance portfolio. We see that AlphaRob is comparable or slightly worse (by up to 3%) compared to the minimum-variance portfolio when the latter's regret is less than 1.05. In other words, the minimum-variance portfolio is hard to beat if it is within 5% of the optimal combination portfolio for a given training size. But as its regret grows (up to 1.24 in some settings), AlphaRob outperforms by a wider margin (up to 23% better).

Note that to calculate the regret, we need the true mean and covariance. So we cannot assume the regret is known while building our portfolio. But AlphaRob automatically captures the benefits (when regret is high) while avoiding the pitfalls (when regret is low).

Robustness of results: Figure 3 also shows that results are similar whether sample returns follow a Gaussian or t-distribution. This is because our analysis only made modest moment assumptions, which hold for many distributions. Also, the Sharpe ratios using the estimated covariance Σ_{NLS} are usually worse than using the true covariance Σ, as expected. But when AlphaRob outperforms the minimum-variance portfolio under Σ, it often does so under Σ_{NLS}.

Imposing a no-shorting constraint: We first find the minimum-variance and maximum-Sharpe portfolios under the no-shorting constraint. Any convex combination of the two constrained portfolios also satisfies the constraint. So we extend AlphaRob to find the best convex combination.

We can do this by a simple pre-processing step. Jagannathan & Ma (2003) showed that the minimum-variance portfolio under the no-shorting constraint corresponds to an unconstrained portfolio using a modified covariance matrix. Similarly, constrained mean-variance portfolios correspond to a modification of the vector of mean asset returns. We now run AlphaRob using this modified covariance matrix and mean vector.

With these modifications, AlphaRob has a 4% higher Sharpe ratio than the minimum-variance portfolio for 6FFEW. But Sharpe ratios are nearly identical for all other datasets. The reason is that the modifications reduce the regret of the minimum-variance portfolio. Jagannathan & Ma (2003) show that the modifications tend to reduce any large positive correlations between assets and increase mean returns that are too low. So there are fewer opportunities to exploit differences between stocks while hedging risk. We find a regret of 1.06 for 6FFEW and less than 1.03 on all the other Fama-French datasets. As mentioned earlier, when the regret is below 1.05, AlphaRob becomes comparable to the minimum-variance portfolio. But when shorting is allowed, the regret can be much larger, and AlphaRob outperforms much more (Figure 4(b)).

6. Experiments on Real-World Datasets

We will first compare the Sharpe ratio of AlphaRob against competing methods on 12 real-world datasets for four training sizes. Then we compare other performance measures, and show a sensitivity analysis.

Experiment setup: In addition to the Fama-French datasets used in Section 5, we construct two monthly-return datasets with $p = 200$ and $p = 500$ stocks. We use the method of Zhao et al. (2019) to select the top stocks by market value over eleven years. This lets us train over $n = 120$ months and test over the next 12 months. We repeat this process until our datasets have as many periods as the Fama-French datasets.

Now, for any dataset of monthly asset returns, we start with the first n months as the training set. From this training set, each competing method constructs a portfolio. The portfolio's returns are calculated on month $n + 1$. Then, we shift the training "window" by one month, i.e., from months 2 to $n+1$. We repeat this process until we cover the entire time series.

Note that we use the monthly return data directly, instead of calculating an overall $(\boldsymbol{\mu}, \Sigma)$ and then simulating returns as in Section 5. Hence, the results here are not comparable to those in Section 5. Also, we can only use robust covariance estimates (such as Σ_{NLS}) since the actual covariance is unknown and may vary over time.

Evaluation metrics: We calculate Sharpe ratios in two ways. The *overall Sharpe* calculates the mean and standard deviation of portfolio returns over all windows and takes their ratio. The *average one-year Sharpe* splits the portfolio returns into one-year periods, calculate the Sharpe ratio for each period, and then take the average of the Sharpe ratios over all periods. The overall Sharpe is intuitive for stationary time series. The average one-year Sharpe may be more useful if financial conditions change significantly over the course of the dataset.

Experimental Details: We run AlphaRob with 3σ confidence intervals, and set the minimum θ to $\theta_- = 0.2$. These settings were chosen from the sensitivity analysis presented later. We compare against the minimum-variance (Min

Var) and the estimated maximum-Sharpe (Est. Max Sharpe) portfolios built using the NLS and L2 covariance estimators (Ledoit & Wolf, 2004, 2012). We also compare against three combination portfolio methods (AA, EQL MV-min, and TZ), a robust portfolio (CS), and a method based on conjugate descent (PARR) (Garlappi et al., 2007; DeMiguel et al., 2013; Tu & Zhou, 2011; Ceria & Stubbs, 2006; DeMiguel et al., 2009a)[3]. For some methods, their authors do not present any parameter tuning method to optimize for Sharpe ratio. So we report best-case results using parameters picked from out-of-sample data. Note that AlphaRob needs no parameter tuning.

Comparison of Sharpe Ratios: We compare methods based on the relative difference between their Sharpe ratios. Since Sharpe ratios measure returns per unit of risk, the relative difference between two Sharpe ratios corresponds to the relative difference in returns for the same amount of risk[4]. Table 1(a) and 1(b) show the overall Sharpe ratio and the average one-year Sharpe ratio when the training set is $n = 120$ months. We separate the CS and PARR methods since they are optimized with out-of-sample data.

Compared to the remaining methods, AlphaRob is the best on eleven out of twelve datasets for overall Sharpe, and on ten datasets for average one-year Sharpe. The next-best method is Min Var (NLS), which is 7.5% worse on average in terms of overall Sharpe ratio, and 5.5% worse for the average one-year Sharpe ratio. For some datasets, AlphaRob is 21% better than Min Var (NLS). Further, the difference between AlphaRob and Min Var (NLS) is statistically significant for five datasets.

Against competing combination portfolios, AlphaRob is 28% better on average than the next-best algorithm (AA). Even when compared to CS and PARR under their optimal settings, AlphaRob fares well. AlphaRob is, on average, 3% better than PARR in terms of the overall Sharpe ratio, and roughly comparable in terms of the average one-year Sharpe ratio. The CS method works well when the number of assets is small ($p = 6$ or 10). But for large p, the Sharpe ratio of CS is usually worse.

	48FFVW	200Stocks	10FFVW	500Stocks	10FFEW	48FFEW	6FFVW	100FFVW	25FFVW	100FFEW	6FFEW	25FFEW
AlphaRob	**0.24**	**0.26**	0.27**	**0.28**	**0.33**	**0.33**	**0.35**	**0.38**	**0.40**	**0.42**	**0.45**	**0.48**
Min Var (NLS)	**0.24**	**0.26**	0.29	**0.28**	0.27*	0.27	0.32**	0.36	0.36**	0.40	0.39***	0.43*
Min Var (L2)	0.23***	0.22***	**0.30**	0.27	0.29	0.28	0.29***	0.34**	0.34**	0.38*	0.33***	0.40**
AA (γ=1)	0.18***	0.21	0.29	0.27	0.28*	0.26**	0.34	0.17***	0.38	0.21***	0.41**	0.46
AA (γ=3)	0.17***	0.22	0.29	**0.28**	0.27*	0.24**	0.33*	0.19***	0.37*	0.21***	0.40***	0.45*
EQL MV-min (γ=1)	0.18***	0.08**	0.15***	0.06***	0.25**	0.26**	0.30*	0.17***	0.36	0.20***	0.39**	0.44*
EQL MV-min (γ=3)	0.14**	0.04***	0.08***	−0.01***	0.21***	0.24**	0.27**	0.17***	0.27***	0.20***	0.37**	0.35**
TZ (γ=1)	0.10**	x	0.12***	x	0.20*	0.07**	0.29**	0.02***	0.22***	0.10***	0.37**	0.37**
TZ (γ=3)	0.08***	x	0.10**	x	0.17	0.03***	0.29**	0.07***	0.20***	0.06**	0.37**	0.36***
Est. Max Sharpe (NLS)	0.11**	0.01***	0.10**	−0.02***	0.18	0.19*	0.29**	0.21***	0.25**	0.24***	0.37**	0.37**
Est. Max Sharpe (L2)	0.09***	0.02***	0.11**	−0.02***	0.17	0.12	0.26***	−0.00**	0.26**	0.18***	0.35***	0.39**
Equal Weight	0.22	0.22	0.24*	0.23	0.23*	0.23	0.23***	0.24***	0.23***	0.24***	0.23***	0.23***
AlphaRob	0.24	0.26	0.27	**0.28**	**0.33**	**0.33**	**0.35***	**0.38**	**0.40**	**0.42**	0.45	**0.48**
CS (best)	0.18***	x	0.29	x	0.31	0.27**	**0.35**	0.18***	0.39	0.21***	**0.46**	0.47
PARR (best)	**0.27**	**0.29**	**0.30**	**0.28**	0.30	0.30	0.33	0.36	0.36*	0.40	0.41**	0.45

(a) Overall Sharpe ratios. Note that CS and PARR are optimized with out-of-sample data.

	48FFVW	200Stocks	10FFVW	500Stocks	10FFEW	48FFEW	6FFVW	100FFVW	25FFVW	100FFEW	6FFEW	25FFEW
AlphaRob	0.31	**0.32**	0.34**	**0.38**	**0.38**	**0.42**	**0.42**	**0.49**	**0.47**	**0.50**	**0.55**	**0.57**
Min Var (NLS)	0.31	**0.32**	0.35*	**0.38**	0.33*	0.35**	0.41	0.45*	0.44*	0.48	0.51	0.54*
Min Var (L2)	0.29	0.28	**0.36**	0.35	0.37	0.37*	0.38*	0.44*	0.44	0.48	0.45***	0.52**
AA (γ=1)	0.23***	0.26	0.35	0.36	0.34*	0.32***	**0.42**	0.18***	**0.47**	0.26***	0.53	**0.57**
AA (γ=3)	0.23***	0.26*	0.35	0.37	0.33**	0.29***	0.41	0.20***	0.45	0.25***	0.53	0.56
EQL MV-min (γ=1)	0.22***	0.15***	0.22***	0.11***	0.29***	0.32***	0.34**	0.18***	0.42**	0.23***	0.46**	0.52**
EQL MV-min (γ=3)	0.16***	0.11***	0.17***	0.04***	0.25***	0.33**	0.31**	0.18***	0.31***	0.23***	0.44**	0.42***
TZ (γ=1)	0.13***	x	0.20***	x	0.27***	0.20***	0.34**	0.06***	0.28***	0.13***	0.45**	0.44***
TZ (γ=3)	0.12***	x	0.19***	x	0.26***	0.14***	0.34**	0.12***	0.26***	0.11***	0.45**	0.43***
Est. Max Sharpe (NLS)	0.16***	0.05***	0.20***	0.01***	0.27***	0.29***	0.34**	0.31***	0.32***	0.34***	0.45**	0.44***
Est. Max Sharpe (L2)	0.15***	0.04***	0.22***	0.01***	0.27***	0.29***	0.33**	0.21***	0.33***	0.28***	0.46*	0.47**
Equal Weight	**0.32**	0.30	0.33	0.32**	0.31	0.30*	0.34***	0.34**	0.34**	0.34**	0.33***	0.34***
AlphaRob	0.31	0.32	0.34	**0.38**	**0.38**	**0.42**	**0.42**	**0.49**	**0.47**	**0.50**	0.55	**0.57**
CS (best)	0.23***	x	0.35	x	0.36	0.33***	**0.42**	0.19***	**0.47**	0.25***	**0.56**	**0.57**
PARR (best)	**0.34**	**0.36**	**0.37**	**0.38**	**0.38**	0.40	0.41	0.45*	0.44	0.49	0.53	0.56

(b) Average one-year Sharpe ratios. CS and PARR are optimized with out-of-sample data.

	48FFVW	200Stocks	10FFVW	500Stocks	10FFEW	48FFEW	6FFVW	100FFVW	25FFVW	100FFEW	6FFEW	25FFEW
AlphaRob	0.18	0.17	0.25**	0.19	**0.29**	**0.26**	**0.32**	**0.29**	**0.32**	**0.33**	**0.41**	**0.40**
Min Var (NLS)	0.20	0.18	0.28	0.19	0.24*	0.21	0.30*	0.28	0.29**	0.32	0.37***	0.37**
Min Var (L2)	0.18	0.09	**0.29**	0.14	0.28	0.24	0.28***	0.23**	0.30**	0.28*	0.32***	0.37**
AA (γ=1)	0.09	−0.07	0.27	0.12	0.25*	0.16**	0.31	−0.25***	0.29	−0.21***	0.39**	0.37
AA (γ=3)	0.09	−0.06	0.28	0.13	0.24*	0.14**	0.31	−0.23***	0.28	−0.21***	0.38**	0.36
EQL MV-min (γ=1)	0.08	−0.16*	0.09***	−0.11***	0.17**	0.15**	0.18*	−0.28***	0.18	−0.26***	0.31**	0.28
EQL MV-min (γ=3)	−0.10	−0.19**	0.02***	−0.14***	0.09**	0.03**	0.12**	−0.34***	−0.17***	−0.36***	0.27**	0.00**
TZ (γ=1)	−0.03*	x	0.08***	x	0.14	−0.02**	0.24**	−0.47***	0.07**	−0.88***	0.32**	0.22**
TZ (γ=3)	−0.05**	x	0.06**	x	0.11	−0.12***	0.24**	−0.82***	0.04**	−0.96**	0.31**	0.21**
Est. Max Sharpe (NLS)	0.03*	−0.09***	0.06**	−0.11***	0.13	0.11*	0.24**	−0.06***	0.12**	0.07***	0.32**	0.24**
Est. Max Sharpe (L2)	0.01**	−0.12***	0.08**	−0.16***	0.14	0.07	0.22***	−0.20**	0.17**	−0.23***	0.33**	0.30**
Equal Weight	**0.22**	**0.22**	0.24*	**0.23**	0.23*	0.23	0.23***	0.24***	0.23***	0.24***	0.23***	0.23***
AlphaRob	0.18	0.17	0.25	0.19	**0.29**	0.26	**0.32***	**0.29**	**0.32**	**0.33**	0.41	**0.40**
CS (best)	0.09**	x	0.27	x	0.27	0.16	**0.32**	−0.24***	0.30	−0.22***	**0.42**	0.38
PARR (best)	**0.26**	**0.26**	**0.29**	**0.26**	**0.29**	**0.30**	0.31	0.25	0.27*	0.30	0.39**	0.38

(c) Overall Sharpe Ratio adjusted for transaction costs (50 basis points).

Table 1: Comparison of Sharpe ratios on real-world datasets for training size $n = 120$ months. A cross denotes that the method does not apply or it did not finish with a feasible portfolio. Significance levels at 0.1, 0.05, and 0.01 are indicated by one, two, and three stars respectively. Significance is checked following Ledoit & Wolf (2008) for the overall Sharpe ratio, and by paired t-tests for the average one-year Sharpe ratio.

	48FFVW	200Stocks	10FFVW	500Stocks	10FFEW	48FFEW	6FFVW	100FFVW	25FFVW	100FFEW	6FFEW	25FFEW
AlphaRob	3.85***	3.34	3.73***	**3.04**	3.82**	4.13***	4.28***	3.85**	4.02***	3.85	4.81**	3.99***
Min Var (NLS)	**3.65**	**3.32**	3.58**	**3.04**	3.54	3.69**	**4.05**	**3.65**	3.72	**3.72**	**4.46**	3.68
Min Var (L2)	3.69	3.49***	**3.53**	3.08	**3.49**	**3.63**	**4.05**	3.86***	**3.67**	3.85***	4.50	**3.67**
AA (γ=1)	4.27***	4.62***	3.62***	3.53***	3.53	4.29***	4.08	10.01***	3.91***	10.15***	4.48	3.91**
AA (γ=3)	4.20***	4.34***	3.61***	3.49***	3.54	4.20***	4.07	7.76***	3.88***	7.71***	4.47	3.86**
EQL MV-min (γ=1)	4.35***	9.33	5.68*	6.85***	6.87***	4.34***	6.21***	7.51***	5.15***	7.49***	6.61***	5.09***
EQL MV-min (γ=3)	7.33***	14.39	8.42	14.85	9.67***	7.17***	7.07***	8.61***	11.33***	9.52***	7.41***	10.70***
TZ (γ=1)	10.23***	x	6.60	x	9.76***	42.51**	5.98***	159.91	8.81***	72.61***	6.68*	6.70***
TZ (γ=3)	12.26***	x	7.29	x	11.25*	599.86*	6.08***	63.52***	9.89**	103.27***	6.79*	6.99***
Est. Max Sharpe (NLS)	9.51***	9.82***	7.50	10.20**	10.35**	13.04**	6.00***	11.22*	8.35***	7.34***	6.72**	6.25***
Est. Max Sharpe (L2)	10.05***	9.91***	7.01	8.57***	10.15	18.27	6.12***	66.45**	7.24**	11.02***	6.11*	5.45***
Equal Weight	4.89***	4.57***	4.31***	4.79***	5.73***	5.69***	4.92***	5.20***	5.11***	5.41***	5.42***	5.35***
CS (best)	4.34***	x	3.62***	x	3.88***	4.38***	4.32***	7.50***	3.96***	8.02***	4.98***	3.98**
PARR (best)	3.87**	3.56***	3.60	3.51***	4.10**	4.38***	4.09	3.87***	3.87***	3.87***	4.47	3.71

(a) Standard deviation ($n = 120$).

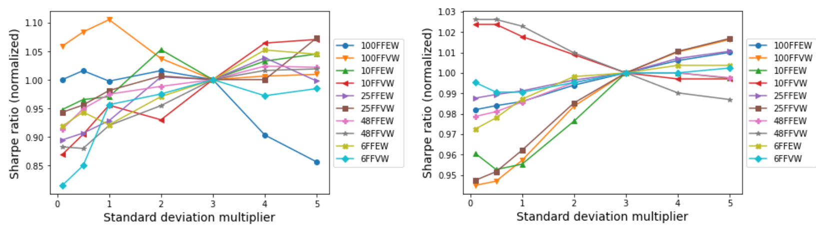
	48FFVW	200Stocks	10FFVW	500Stocks	10FFEW	48FFEW	6FFVW	100FFVW	25FFVW	100FFEW	6FFEW	25FFEW
AlphaRob	3.30	2.33	1.24	2.46	2.60	2.79	1.69	2.25	1.97	2.02	3.56	1.44
Min Var (NLS)	3.56	2.44	1.10	2.46	2.79	3.11	1.64	2.32	1.52	2.53	2.40	2.18
Min Var (L2)	3.60	1.95	1.19	2.19	3.42	3.28	2.02	3.19	1.98	2.88	2.95	2.48
AA (γ=1)	2.88	5.00	1.16	3.64	2.81	2.47	1.72	2.61	1.84	1.02	2.41	2.12
AA (γ=3)	2.76	3.71	1.11	3.23	2.81	1.95	1.72	0.79	1.73	0.49	2.36	2.18
EQL MV-min (γ=1)	3.22	64.89	19.71	24.85	7.46	2.97	2.49	0.83	2.17	0.57	8.64	2.41
EQL MV-min (γ=3)	4.67	131.79	68.15	144.97	12.30	5.04	3.53	1.36	3.34	8.00	11.23	3.35
TZ (γ=1)	6.24	x	49.36	x	53.90	196.07	4.61	271.94	20.03	54.39	15.88	4.48
TZ (γ=3)	12.08	x	67.02	x	84.48	396.92	4.68	44.27	26.22	62.42	16.15	4.68
Est. Max Sharpe (NLS)	11.66	51.38	78.29	81.16	67.75	53.16	4.69	69.09	21.57	12.73	15.35	4.25
Est. Max Sharpe (L2)	13.95	35.19	68.51	24.86	90.52	170.78	7.60	351.71	22.99	23.02	17.00	6.43
Equal Weight	3.12	2.09	2.28	2.79	3.35	3.89	2.90	3.07	2.96	3.35	3.28	3.30
CS (best)	2.91	x	1.13	x	2.30	2.79	1.59	0.82	1.94	0.47	2.58	2.06
PARR (best)	3.27	2.63	1.25	2.82	3.79	4.29	1.72	2.34	1.62	2.24	2.52	2.00

(b) Average kurtosis of portfolios ($n = 120$).

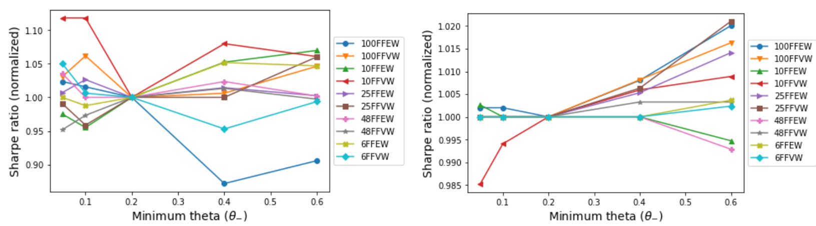
Table 2: Alternate performance measures.

AlphaRob is best on eight datasets even after adjusting for transaction costs (Table 1(c)). Following DeMiguel et al. (2009b), we impose a 50 basis point penalty on transactions. AlphaRob is statistically significantly better than Min Var (NLS) on five datasets. These are the same five datasets where AlphaRob outperforms Min Var (NLS) significantly even without transaction costs (Table 1(a)). So, the inclusion of transaction costs does not change the comparative advantage of AlphaRob over Min Var (NLS). The Equal Weight portfolios improves relative to others because it has zero transaction costs. Several competing methods have a negative Sharpe ratio after transaction costs.

Overall, AlphaRob is comparable or better than Min Var (NLS) on most Fama-French datasets. For `200Stocks` and `500Stocks`, the regret of Min Var is only 1.03 due to high correlations[5]. So AlphaRob is comparable to Min Var (NLS) here. While the above results are for a training size of $n = 120$, the supplementary material shows similar results for other training sizes too.



(a) Sharpe ratio vs. Std. dev. multiplier for $n = 15$ (left) and $n = 120$ (right).



(b) Sharpe ratio vs. minimum θ for $n = 15$ (left) and $n = 120$ (right).

Figure 5: Sensitivity analysis for the parameters of AlphaRob. We set the standard deviation multiplier to 3 and the minimum θ to 0.2 by default.

Other Performance Measures: Table 2 shows the standard deviation and kurtosis of all methods. The standard deviation of AlphaRob is lower than other competing methods, except Min Var (NLS). Between AlphaRob and Min Var (NLS), the greatest percentage differences are for `48FFEW`, `25FFVW`, and `25FFEW`. For all these datasets, the Sharpe ratio of AlphaRob is significantly better than Min Var (NLS). Thus, AlphaRob is assuming extra variability only when it helps the Sharpe ratio. The kurtosis of AlphaRob is comparable to the minimum-variance portfolios and better than the other methods.

Sensitivity Analysis: By default, AlphaRob uses 3σ confidence intervals for both $\|\boldsymbol{z}\|$ and θ. In other words, we set the standard deviation multiplier to a default of 3. Plot 5(a) shows that the Sharpe ratio can fall if the standard deviation multiplier is smaller than 2. A small multiplier means smaller confidence intervals $\mathcal{I}_{\|\boldsymbol{z}\|}$ and $\mathcal{I}_{\theta}$. These are less likely to contain the actual parameters $\|\boldsymbol{z}\|$ and θ. Hence, AlphaRob may choose a poor α_{rob}. The Sharpe ratio can also fall if the multiplier is greater than 4 when the training sample size n is small. The reason is that for small n, the estimates of $\|\boldsymbol{z}\|$ and θ have a substantial error. So the standard deviations are often significant. A large multiplier makes the robust solution too conservative in such cases. Any multiplier between two and four works well.

AlphaRob also sets the minimum value of θ to $\theta_- = 0.2$. This implies a constraint on the best achievable Sharpe ratio: $\mathcal{S}_w(\boldsymbol{w}_{\text{MS}}) \leq \mathcal{S}_w(\boldsymbol{w}_{\text{MV}})/\theta_-$ by Eq. 4. Plot 5(b) shows the results for varying θ_-. If θ_- is too large, we downplay the benefits of the maximum-Sharpe portfolio. So AlphaRob chooses an α_{rob} that is too small. But if θ_- is too small, we may be too optimistic and choose large values for α_{rob}. This is especially true for small n, where the estimate $\hat{\theta}$ may be far from the true θ. We see that $\theta_- \in [0.1, 0.2]$ works well in all settings.

7. Discussion

We showed how noisy estimates of mean returns can be fruitfully used in the maximum-Sharpe portfolio problem. To account for the estimation errors, our algorithm (AlphaRob) uses a robust optimization, with two important details. First, we only consider a limited set of portfolios, called combination portfolios. This restriction allows us to capture uncertainty about mean returns with an interval. We need not create an uncertainty set for the entire mean returns vector. The simple uncertainty set lets AlphaRob scale to problems with many assets. Second, our robust objective is to minimize worst-case regret. We determine a portfolio's regret by comparing its Sharpe ratio to the optimal combination portfolio *adjusted for the training sample size.* AlphaRob is fast and needs no parameter tuning. It performs as well or better than competing methods on both simulated and real-world datasets.

We also show that the performance of the minimum-variance portfolio is negatively correlated with its regret. The regret increases with training size and the average level of risk-adjusted excess return across assets. The latter term can be high for uncorrelated assets, such as those based on factors. Then, the regret of the minimum-variance portfolio is high, and AlphaRob achieves a significantly better Sharpe ratio. However, when the assets are individual stocks, their high correlations reduce the regret of the minimum-variance portfolio. If this regret is within 5% of the optimal, the minimum-variance portfolio is hard to beat. AlphaRob automatically captures the benefits (when regret is high) while avoiding the pitfalls (when regret is low). So, investors and portfolio managers can use AlphaRob for all regret levels.

Several extensions of our work are possible. We could extend the analysis from Sharpe ratios to general utility functions. The supplementary material presents some results for the mean-variance utility function. But, for each utility function, we need a bespoke algorithm to find a robust portfolio. A second extension would be to impose uncertainty sets on the covariance matrix as well. Our analysis shows that the covariance affects the expected Sharpe ratio only via a few functions of the return distribution. A careful analysis would be needed to characterize the behavior of these functions under the uncertainty set. Finally, the analysis of the expected Sharpe ratio can be extended to include the variance of the robust portfolio. The robust portfolio depends on a confidence interval estimated from data. So we need to consider the variability of confidence intervals constructed by the jackknife.

Endnotes

[1]This condition holds for all our datasets.

[2]An alternative is to seek the best *constant* β with the highest expected Sharpe ratio. We can show that the optimal expected Sharpe ratio is the same as for our formulation. However, Thm. 6 no longer holds, and we need it to have a simple regret-minimization algorithm.

[3]The details are presented in the supplementary material. We do not compare against the no-shorting portfolio (Jagannathan & Ma, 2003), the L_1-regularized minimum-variance portfolio (Brodie et al., 2009), Bayesian methods, and the method of Kan & Zhou (2007), since Min Var (NLS) and the Equal Weight portfolio perform as well or better (DeMiguel et al., 2009b; Zhao et al., 2019).

[4]A portfolio manager may invest in a portfolio of assets until the standard deviation hits a maximum threshold. So if two portfolios are available, then the relative difference in returns is given by the relative difference in Sharpe ratios. The portfolio manager may strongly prefer one portfolio to another if relative differences in Sharpe ratio are high, even if absolute differences are modest.

[5] We calculate the regret from estimates of the mean and covariance from the full dataset, assuming stationarity.

References

Bickel, P. J., & Levina, E. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics*, *36*, 199–227.

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

Ceria, S., & Stubbs, R. A. (2006). Incorporating estimation errors into portfolio selection: Robust portfolio construction. *Journal of Asset Management; London*, *7*, 109–127.

Chopra, V. K., & Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, *19*, 249–257.

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

DeMiguel, V., Garlappi, L., & Uppal, R. (2009b). Optimal versus naïve diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, *22*, 1915–1953.

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

DeMiguel, V., & Nogales, F. J. (2009). Portfolio selection with robust estimation. *Operations Research*, *57*, 560–577.

Deng, G., Dulaney, T., Mccann, C., & Wang, O. (2013). Robust portfolio optimization with Value-at-Risk-adjusted Sharpe ratios. *Journal of Asset Management; London*, *14*, 293–305.

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

Frost, P. A., & Savarino, J. E. (1986). An empirical Bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, *21*, 293–305.

Frost, P. A., & Savarino, J. E. (1988). For better performance: Constrain portfolio weights. *The Journal of Portfolio Management*, *15*, 29–34.

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

Goldfarb, D., & Iyengar, G. (2003). Robust Portfolio Selection Problems. *Mathematics of Operations Research*, *28*, 1–38.

Guerreiro, A. P., & Fonseca, C. M. (2020). An analysis of the Hypervolume Sharpe-Ratio Indicator. *European Journal of Operational Research*, *283*, 614–629.

Gülpınar, N., & Rustem, B. (2007). Worst-case robust decisions for multi-period mean–variance portfolio optimization. *European Journal of Operational Research*, *183*, 981–1000.

Hanke, M., & Penev, S. (2018). Comparing large-sample maximum Sharpe ratios and incremental variable testing. *European Journal of Operational Research*, *265*, 571–579.

Huang, D., Zhu, S., Fabozzi, F. J., & Fukushima, M. (2010). Portfolio selection under distributional uncertainty: A relative robust CVaR approach. *European Journal of Operational Research*, *203*, 185–194.

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

Ji, R., & Lejeune, M. (2020). Data-Driven Optimization of Reward-Risk Ratio Measures. *INFORMS Journal on Computing*, *(in press)*.

Jobson, J. D., & Korkie, B. (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, *75*, 544–554.

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

Kakouris, I., & Rustem, B. (2014). Robust portfolio optimization with copulas. *European Journal of Operational Research*, *235*, 28–37.

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

Kirby, C., & Ostdiek, B. (2012). It's All in the Timing: Simple Active Portfolio Strategies that Outperform Naïve Diversification. *The Journal of Financial and Quantitative Analysis*, *47*, 437–467.

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

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*, 365–411.

Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, *15*, 850–859.

Ledoit, O., & Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, *40*, 1024–1060.

Ledoit, O., & Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *The Review of Financial Studies*, *30*, 4349–4388.

Michaud, R. O. (1989). The Markowitz optimization enigma: is 'optimized' optimal? *Financial Analysts Journal*, *45*, 31–42.

Qiu, H., Han, F., Liu, H., & Caffo, B. (2015). Robust Portfolio Optimization. In *Advances in Neural Information Processing Systems 28* (pp. 46–54).

Scherer, B. (2007). Can robust portfolio optimisation help to build better portfolios? *Journal of Asset Management; London*, *7*, 374–387.

Scutellà, M. G., & Recchia, R. (2013). Robust portfolio asset allocation and risk measures. *Annals of Operations Research*, *204*, 145–169.

Shen, R., & Zhang, S. (2008). Robust portfolio selection based on a multi-stage scenario tree. *European Journal of Operational Research*, *191*, 864–887.

Stevens, G. V. (1998). On the inverse of the covariance matrix in portfolio analysis. *The Journal of Finance*, *53*, 1821–1827.

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

Tütüncü, R., & Koenig, M. (2004). Robust Asset Allocation. *Annals of Operations Research*, *132*, 157–187.

Xidonas, P., Steuer, R., & Hassapis, C. (2020). Robust portfolio optimization: a categorized bibliographic review. *Annals of Operations Research*, *292*, 533–552.

Zhao, L., Chakrabarti, D., & Muthuraman, K. (2019). Portfolio construction by mitigating error amplification: The bounded-noise portfolio. *Operations Research*, *67*, 965–983.