Why Naïve $1/N$ Diversification Is Not So Naïve, and How to Beat It?*

Ming Yuan

Columbia University

Email: ming.yuan@columbia.edu

Guofu Zhou

Washington University in St. Louis

Email: zhou@wustl.edu

Current version: January 2022

*We are grateful to Victor DeMiguel, John Dooley, Lorenzo Garlappi, Raymond Kan, Nathan Lassance, Ľuboš Pástor, Jun Tu, Raman Uppal, Xiaolu Wang, Alex Weissensteiner, MIchael Wolf, Paolo Zaffaroni, and seminar and conference participants at Washington University in St. Louis and 15th International Conference on Computational and Financial Econometrics (2021) for insightful and helpful comments.

Send correspondence to Guofu Zhou, Olin School of Business, Washington University in St. Louis, St. Louis, MO 63130; e-mail: zhou@wustl.edu; phone: 314-935-6384.

Why Naïve $1/N$ Diversification Is Not So Naïve,

and How to Beat It?

Abstract

In this paper, we study the portfolio choice problem under estimation risk and show why the $1/N$ rule is very difficult to beat in applications and studies. First, as long as the dimensionality is high relative to sample size, we show that the usual estimated investment strategies are biased even asymptotically. Second, we show that the $1/N$ rule is optimal in a one-factor model with diversifiable risks as dimensionality increases, irrespectively of the sample size, making investment theory-based rules inadequate as they suffer from estimation errors. Third, we provide strategies that can outperform the $1/N$ under suitable conditions.

JEL Classification: C53, G11, G12, G17

Keywords: portfolio choice; estimation risk; mean-variance optimization; naïve diversification; $1/N$.

1. Introduction

Portfolio choice is obvious truly one of the most important aspects of investment theory, and the mean-variance framework pioneered by Markowitz (1952) seminal work has been the major model used in practice in asset allocation and active portfolio management.[1] However, to implement the mean-variance optimal portfolio, both asset expected return and covariance matrix must be estimated, introducing the well-known estimation risk problem. Brown (1976) and Bawa, Brown, and Klein (1979) and Jorion (1986), Jagannathan and Ma (2003), MacKinlay and Ľuboš Pástor (2000) and Ledoit and Wolf (2004) and Kan and Zhou (2007) are examples of earlier studies that provide various strategies to overcome the estimation risk, but the performances of these strategies are sample size dependent. In their highly influential study, DeMiguel, Garlappi and Uppal (2009) compare these strategies with the naïve $1/N$ investment strategy that invests equally among N risky assets.[2] They find astonishingly that

> "*the estimation window needed for the sample-based mean-variance strategy and its extensions to outperform the* $1/N$ *benchmark is around 3000 months for a portfolio with 25 assets and about 6000 months for a portfolio with 50 assets.*"

Their finding raises a serious issue on the value of investment theory as the sample size is too large for the theory to be realistically applied in practice. To overcome the estimation risk, DeMiguel, Garlappi, Nogales and Uppal (2009), Duchin and Levy (2009), Tu and Zhou (2011), Kirby and Ostdiek (2012), DeMiguel, Martin-Utrera and Nogales (2015), Ledoit and Wolf (2017), Lassance, Martin-Utrera and Simaan (2021), Kan, Wang and Zhou (2021), among others, propose additional portfolio strategies that have improved performances. However, it is unknown why those strategies can beat the $1/N$ in some data sets and still cannot in others.

[1] See, e.g., Grinold and Kahn (1999), Litterman (2003), Meucci (2005), Qian, Hua, and Sorensen (2007) and references therein.

[2] Their paper has over 3100 Google citations. The $1/N$ strategy was known as the Talmud rule more than 1500 years ago (Duchin and Levy (2009)).

In this paper, we aim at providing a deep theoretical understanding of both of the $1/N$ and some of the major related estimated portfolio strategies, shedding light on why the estimated rules often cannot beat the $1/N$, as well as providing conditions under which our proposed rules can beat the $1/N$. Specifically, we make three major contributions. First, we obtain both the exact and asymptotic distributions of the Sharpe ratio of the combination strategy that combines the plug-in with the estimated global minimum portfolio (GMV). The exact distribution is useful to understand what drives the performance of the rule. Kan, Wang and Zheng (2020) provide the first exact distribution for the Sharpe of the plug-in rule, ours extend their result to the general three-fund case. In particular, we obtain also the exact distribution of the Sharpe ratio of the GMV, which is widely used in practice to avoid estimating the mean (see, e.g., Kempf and Memmel (2006), Basak, Jagannathan and Ma (2009) and Bodnar, Parolya and Schmid (2018) for its properties).

The asymptotic distribution concerns the case when N is large, but $N < T$ and N/T approaches a constant as sample size T increases to infinity. This large dimension case is quite relevant in practice and in existing studies. For example, the sample size T is 120 whereas N varies from 3 to 24 (see, e.g., Table 3 of DeMiguel, Garlappi and Uppal (2009)). Complimenting Ao, Li and Zhen (2019) who obtain the asymptotic distribution of the Sharpe ratio for the plug-in rule, we obtain it for the general three-fund strategy. The asymptotic distribution is valuable for understanding easier the performance of the various rules. In particular, while it is well known that the general three-fund, a special case of which is the plug-in, is optimally for large T but fixed N, we show that it is asymptotically biased in the large N case.

The second contribution of our paper is to show that the $1/N$ is optimal when N is large enough, under the condition that the asset returns are governed by a one-factor model with diversifiable risks. This result is very intuitive once known, and is yet very powerful. If there is one factor that prices all the assets with diversifiable risks, then the $1/N$ portfolio must be a portfolio of the factor with the average betas as the weight plus idiosyncratic risks. Since the risks are diversifiable, they will not alter the Sharpe ratio of the $1/N$ portfolio in the limit. Hence, the Sharpe ratio of the $1/N$

portfolio must approach that of the factor and be optimal.

The third contribution is to propose new rules that can beat the $1/N$ under certain conditions. We consider two cases which require two fundamentally different solutions. In the first case when $N < T$, we consider the combination rule of combining the plug-in with the $1/N$. Tu and Zhou (2011) is the first to study such a rule. In contrast to their study, we focus on maximizing the Sharpe ratio, which is a popular criterion in practice, instead of the expected mean-variance utility. We obtain both the exact distribution of the Sharpe ratio of the combination rule, and the asymptotic distribution in the high dimensional case. We also solve explicitly the combination coefficient that maximizes the asymptotic Sharpe ratio.

In the case of $N > T$, we consider a long-short portfolio that long assets with extreme positive alphas relative to the $1/N$, and short those with extreme negative ones. We show that adding this portfolio to the $1/N$ will improve it strictly unless all the true alphas are zeros.

It will be useful to discuss the difference and limitation of our paper. We focus here on Sharpe ratios, while many studies focus on the expected mean-variance utility, which Lassance, Martin-Utrera and Simaan (2021) extend into a more robust framework by considering the uncertainty associated with utility maximization. Without estimation errors, all are equivalent. With estimation errors, the objectives are mathematically different. Lassance (2021) provides an interesting comparison between the two setups in the fixed N case. In contrast, we focus on Sharpe ratio only and on comparing existing strategies with the $1/N$ in the high dimension case. Our paper shows the optimality of the $1/N$ rule only in a restricted one-factor model. In a general multi-factor APT, Raponi, Uppal and Zaffaroni (2021) provide a new strategy with superior out-of-sample performance.

The rest of the paper is organized as follows. Section 2 discuss properties of the common estimated rules. Section 3 provides conditions under which the $1/N$ is optimal. Section 4 explores ways to outperform the $1/N$. Section 5 concludes.

2. Bias of estimated rules

Consider the standard mean-variance portfolio choice problem in which an investor chooses his optimal portfolio among N risky assets and a riskfree asset. Denote the returns of the N risky assets at time t by r_t, and r_f the return on the riskfree asset. Let $R_t = r_t - r_f$ be an N-vector of the excess returns, with mean μ and covariance matrix Σ.

In the absence of estimation errors, the optimal portfolio weights are well known,

$$w = \frac{1}{\gamma}\Sigma^{-1}\mu, \tag{1}$$

where γ is a risk aversion parameter, and the optimized Sharpe ratio is

$$SR = \sqrt{\mu^\top \Sigma^{-1}\mu}. \tag{2}$$

However, the parameters μ and Σ are unknown, and are usually estimated from data.

Suppose there are T periods of observed excess returns data, the common sample estimates are

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t, \tag{3}$$

$$\hat{\Sigma} = \frac{1}{T-1}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})^\top. \tag{4}$$

The data are usually assumed identically and independently distributed (iid). For obtaining exact distributional results, they are further assumed to be normal, that is, $R_t \sim N(\mu, \Sigma)$.

The plug-in rule or the estimated optimal portfolio weights are those replacing the unknown parameters by their estimates,

$$\hat{w} = \frac{1}{\gamma}\hat{\Sigma}^{-1}\hat{\mu}. \tag{5}$$

Since it is the estimates rather than the true parameters that are used, this introduces the estimation errors in the portfolio weights, making the resulting portfolio will not necessarily achieve the optimal Sharpe ratio.

The estimated global minimum portfolio (GMV) is also very popular in practice. Without estimation errors, the GMV portfolio weights are

$$w_g = \frac{\Sigma^{-1} 1_N}{1_N^\top \Sigma^{-1} 1_N}, \tag{6}$$

and so the sample version are

$$\hat{w}_g = \frac{\hat{\Sigma}^{-1} 1_N}{1_N^\top \hat{\Sigma}^{-1} 1_N}. \tag{7}$$

It may be noted that the estimation errors occur only in $\hat{\Sigma}^{-1}$.

Kan and Zhou (2007), to improve the plug-in rule, propose the following three-fund rule,

$$\hat{w}_\lambda = (1-\lambda)\frac{1}{\gamma}\hat{\Sigma}^{-1}\hat{\mu} + \lambda \hat{\Sigma}^{-1} 1_N, \qquad \lambda \in [0,1], \tag{8}$$

which is a combination of the plug-in, estimated GMV, and the risk free asset.[3] Lassance, Martin-Utrera and Simaan (2021) examine how such a rule performs in their proposed robustness framework.

Let SR_λ be the Sharpe ratio of the three-fund rule $\hat{w}_\lambda$. We are here interested in the exact distribution of SR_λ. Let μ_g and σ_g^2 be the expected return and variance of the GMV portfolio,

$$\mu_g = \frac{\mu^\top \Sigma^{-1} 1_N}{1_N^\top \Sigma^{-1} 1_N}, \qquad \sigma_g^2 = \frac{1}{1_N^\top \Sigma^{-1} 1_N}.$$

[3] Kan and Zhou (2007) two different combination coefficients, which can be scaled to allow their sum be one without affecting the Sharpe ratio.

Its Sharpe ratio is given by

$$SR_g = \frac{\mu_g}{\sigma_g} = \frac{\mu^\top \Sigma^{-1} 1_N}{(1_N^\top \Sigma^{-1} 1_N)^{1/2}}.$$

Then the distribution is given by (all proofs are provided in the appendix)

Proposition 1: Assume that $T > N+2$. Let $W \sim Wishart(I_N/(T-1), T-1)$ and $X \sim N(0, I_N/T)$ be independent of each other. Then the exact distribution of SR_λ is

$$SR_\lambda =_d \frac{A}{B^{1/2}}, \tag{9}$$

where

$$\begin{aligned} A \;=\; & ((1-\lambda)SR^2 + \lambda\gamma\sigma_g^{-1}SR_g) \cdot (e_1^\top W^{-1} e_1) + \lambda\gamma\sigma_g^{-1}\left(SR^2 - SR_g^2\right)^{1/2} \cdot (e_2^\top W^{-1} e_1) \\ & + \frac{(1-\lambda)\left((1-\lambda)SR^2 + \lambda\gamma\sigma_g^{-1}SR_g\right)}{((1-\lambda)^2 SR^2 + \lambda^2\gamma^2\sigma_g^{-2} + 2(1-\lambda)\lambda\gamma\sigma_g^{-1}SR_g)^{1/2}} \cdot (e_1^\top W^{-1} X) \\ & + (1-\lambda)\lambda\gamma\sigma_g^{-1}\left(\frac{SR^2 - SR_g^2}{(1-\lambda)^2 SR^2 + \lambda^2\gamma^2\sigma_g^{-2} + 2(1-\lambda)\lambda\gamma\sigma_g^{-1}SR_g}\right)^{1/2} \cdot (e_2^\top W^{-1} X). \end{aligned}$$

and

$$\begin{aligned} B \;=\; & ((1-\lambda)^2 SR^2 + \lambda^2\gamma^2\sigma_g^{-2} + 2(1-\lambda)\lambda\gamma\sigma_g^{-1}SR_g)(e_1 W^{-2} e_1) + (1-\lambda)^2 (X^\top W^{-2} X) \\ & + 2(1-\lambda)((1-\lambda)^2 SR^2 + \lambda^2\gamma^2\sigma_g^{-2} + 2(1-\lambda)\lambda\gamma\sigma_g^{-1}SR_g)^{1/2}(e_1^\top W^{-2} X) \end{aligned}$$

with e_1 and e_2 being the first two canonical basis of $\mathbb{R}^N$.

Proposition 1 shows that the exact distribution of the three-fund rule is the same as the righthand side, which is related to the three usual mean-variance frontier parameters. The role played by N and T are not explicit, but implicit in W and Z which are independent of any unknown parameters μ and Σ. Although the distribution is of unknown form, it can be computed with arbitrary accuracy via Monte Carlo integration for any given key mean-variance frontier parameters.

Let us examine some special cases. When $\lambda = 0$, we have

Corollary 1: For any $T > N+2$, the exact distribution of the Sharpe ratio of the plug-in rule is

$$SR_{\hat{w}} =_d \frac{SR^2(e_1^\top W^{-1}e_1) + SR(e_1^\top W^{-1}X)}{\left(SR^2(e_1^\top W^{-2}e_1) + X^\top W^{-2}X + 2SR(e_1^\top W^{-2}X)\right)^{1/2}}, \tag{10}$$

Kan, Wang and Zheng (2020) provide the first explicit expression for the exact distribution of $SR_{\hat{w}}$. Equation (10) provides a complimentary alternative expression. It is interesting that, given the true Sharpe ratio SR, the Sharpe ratio performance of the plug-in rule has nothing to do any parameters of the model. Both N and T play important roles and they matter only through their impact on W and Z that summarizes the effects of estimation errors.

When $\lambda = 1$, we have

Corollary 2: For any $T > N+2$, the exact distribution of the Sharpe ratio of the GMV rule is

$$SR_{\hat{g}} =_d \frac{SR}{(e_1 W^{-2}e_1)^{1/2}} \left(\rho(e_1^\top W^{-1}e_1) + (1-\rho^2)^{1/2}(e_2^\top W^{-1}e_1)\right), \tag{11}$$

where

$$\rho = \frac{SR_g}{SR} = \frac{(1_N^\top \Sigma^{-1}\mu)}{(1_N^\top \Sigma^{-1}1_N)^{1/2}(\mu^\top \Sigma^{-1}\mu)^{1/2}}$$

is the fraction of the population Sharpe ratio of the GMV, SR_g, relative to the true Sharpe ratio.

Kempf and Memmel (2006) and Basak, Jagannathan and Ma (2009), among others, study various properties of the GMV. The exact distribution, (11), compliments their studies. It is interesting that the distribution of the GMV depends not only on SR, but also on ρ. In contrast to $SR_{\hat{w}}$, it requires one more parameter. Nevertheless, the exact distribution of $SR_{\hat{g}}$ is the simplest, and that of $SR_{\hat{w}}$ still looks simple. In contrast, the distribution of SR_λ is rather complex, whose complexity is introduced by the correlation between $\hat{w}$ and $\hat{w}_g$.

While it is well known that, for a fixed N, as the sample size increases to infinity, the Sharpe ratio of the three-fund rule, a special case of which is the plug-in rule, converges to the optimal one. But, when N is not small relative to T, this is no longer true, as shown below.

Proposition 2: As N/T approaches $0 < \eta < 1$, as T approaches infinity, then

$$SR_\lambda = \tau_\lambda SR + O_p(T^{-1/2}), \tag{12}$$

where

$$\tau_\lambda = \frac{((1-\lambda)SR^2 + \lambda\gamma\sigma_g^{-1}SR_g)\cdot(1-\eta)^{1/2}}{\sqrt{(1-\lambda)^2SR^2 + \lambda^2\gamma^2\sigma_g^{-2} + 2(1-\lambda)\lambda\gamma\sigma_g^{-1}SR_g + (1-\lambda)^2\eta}} < 1,$$

and $O_p(\cdot)$ denotes a bounded function in probability. .

In contrast to the exact distribution (9) with complex terms of A and B, (12) is much simplified with only a simple scalar τ_λ. For any given $\eta > 0$, τ_λ is always less than 1, implying that the three-fund rule will never approach optimal as T approaches infinity. Therefore, it is biased asymptotically regardless of the true parameters as long as the mean-variance frontier is well defined.

The asymptotic distribution allows us to choose λ to maximize the Sharpe ratio. If μ is not a proportional constant vector, then the optimal λ is given by

$$\lambda^* = \frac{SR^2 - SR_g^2}{\gamma(SR^2 - SR_g^2) + \eta\sigma_g SR_g}$$

Given an estimate of the three mean-variance frontier parameters, λ is straightforwardly solved from the above. Note that the optimized λ is strictly greater than 0 whenever GMV has a nontrivial expected return, i.e., $\mu_g \neq 0$, indicating that the combination is always beneficial.

In the special case of $\lambda = 0$, we have

Corollary 3: As N/T approaches $0 < \eta < 1$, and T approaches infinity, then

$$SR_{\hat{w}} = \tau SR + O_p\left(\frac{1}{\sqrt{T}}\right), \tag{13}$$

where

$$\tau = \sqrt{\frac{1-\eta}{1+\eta/SR^2}} < 1.$$

Ao, Li and Zhen (2019) is the first, to our knowledge, to obtain (13) for the plug-in rule. However, their focus is different from ours. To compare the plug-in with the $1/N$, let $SR_{1/N}$ be the Sharpe ratio of the latter, and

$$\delta = SR_{1/N}/SR \tag{14}$$

be the fraction of Sharpe ratio the $1/N$ can achieve. An immediate implication of Corollary 3 is that, if

$$\eta > \frac{1-\delta^2}{1+\delta^2/SR^2}, \tag{15}$$

then

$$SR_{\hat{w}} < SR_{1/N} + O_p\left(\frac{1}{\sqrt{T}}\right),$$

i.e., the naïve diversification is asymptotically superior to the plug-in portfolio. Condition (15) is useful for a quick check on the efficiency of $\hat{w}$.

It is of interest to examine the earlier quoted finding of DeMiguel, Garlappi and Uppal (2009) in light of (15). Assume that $\delta = 90\%$ (which is quite possible given results in the next section). With the same Sharpe ratio assumption as before, then η needs to be smaller than 30.78% in order for $SR_{\hat{w}}$ to be greater than $SR_{1/N}$, implying a sample size of $T > 3000$.

In the special case of $\lambda = 1$, we have

Corollary 4: As N/T approaches $0 < \eta < 1$, and T approaches infinity, then

$$SR_{\hat{g}} = \tau_g SR + O_p(T^{-1/2}), \tag{16}$$

where

$$\tau_g = \rho(1-\eta)^{1/2} < 1.$$

Bodnar, Parolya and Schmid (2018) provide estimation approaches to the GMV in the high dimensional case. In contrast, we compare the Sharpe ratios of $\hat{w}_g$ with that of the optimal one. Equation (16) shows strikingly that the effect of the estimation errors is captured solely by the term $(1-\eta)^{1/2}$. In contrast with $\hat{w}$, the absolute size of the true SR is irrelevant given ρ, and only the faction N/T matters. Intuitively, holding the covariance matrix constant, the smaller the SR, the smaller the means, and hence the more difficult to estimate μ. The larger estimation errors lead to the performance depending on the magnitude of SR. In opposition to this, $\hat{w}_g$ does not require the estimation of the mean, and so its asymptotic behavior is much simpler.

Indeed, the GMV is very popular in practice due to its reduced estimation errors without estimating the means. There are many studies on the GMV for its practical use. Hafner and Wang (2021), for example, impose sector restriction on the GMV to enhance its performance. In light of the asymptotic theory, we can see that the GMV is better asymptotically if and only if

$$\rho > \frac{SR}{\sqrt{\eta + SR^2}}. \tag{17}$$

For example, if $\eta = 0.20$ and $SR = 0.1443$, then $SR/\sqrt{\eta + SR^2} = 0.3071$. As it is quite likely that the population GMV should be able to achieve more than 30% of the true Sharpe ratio, and hence the GMV is preferred. However, this is not always the case. When η is very small, say, $\eta = 0.01$, then the right hand will be 82.19%, making the GMV unlikely to beat the plug-in rule asymptotically.

In summary, this section provides a thorough analysis of the three-fund rule, which is shown by a number of studies performing better than its special cases, the plug-in and the estimated GMV. The important message is that all of the estimated rules are asymptotically biased even as the sample size is large, as long as N is large too. This make them far from optimal. On the other hand, as we will show, the $1/N$ can be optimal under certain conditions. Then, combining these two observations, it is no wonder that $1/N$ is hard to beat in practice.

3. Optimality of $1/N$

Since the naïve $1/N$ investment strategy, which invests equally among N risky assets, has the portfolio weights

$$w_{1/N} = \frac{1}{N} 1_N, \tag{18}$$

which are parameter independent, and so there are no estimation errors whatsoever. However, since it differs from the true portfolio weights

$$w = \frac{1}{\gamma} \Sigma^{-1} \mu,$$

it will not be optimal in general unless w happens to be equal to $w_{1/N}$.

It is a common belief that the $1/N$ rule is close to optimal only when it is close to the true weights, which is clearly very rare. In fact, as we show below, the $1/N$ rule can in fact converge to the optimal Sharpe ratio as N increases, under seemingly general conditions.

Assume that we have the following one-factor model for all the assets,

$$R_i = \beta_i R_q + \varepsilon_i, \quad i = 1, 2, \ldots, N, \tag{19}$$

where R_q is the excess return on the tangency portfolio in the mean-variance frontier, and is uncorrelated with the ε_i's. Under the standard assumptions for the mean-variance portfolio theory, the Two-fund Separation Theorem implies the above (see, e.g., Huang and Litzenberger, 1988, p. 80). The key additional assumption we make is that idiosyncratic risks, the ε_i's, are diversifiable enough so that

$$(1_N^\top \Sigma_\varepsilon 1_N)/N^2 \longrightarrow 0, \qquad as \quad N \to \infty, \tag{20}$$

where Σ_ε is the covariance matrix of $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)'$. Consider a few well known cases. If the idiosyncratic risks has a bound, $\sigma^2_{\max}$, on their variances, then it is clear that

$$(1_N^\top \Sigma_\varepsilon 1_N)/N^2 \leq \sigma^2_{\max}/N \longrightarrow 0.$$

Another simple case is short-distance dependence:

$$cov(\varepsilon_i, \varepsilon_j) = 0, \qquad \text{if } |i-j| \geq \kappa_0$$

for some constant κ_0. A more general condition is

$$cov(\varepsilon_i, \varepsilon_j) \leq C|i-j|^{-\nu}$$

for some $\nu > 1$. Yet another one is block diagonal – the assets can be clustered (into industries) – so that there is no cross-dependence across clusters. Denote by $C(i)$ the cluster asset i falls into, then

$$cov(\varepsilon_i, \varepsilon_j) = 0, \qquad \text{if } C(i) \neq C(j).$$

The condition holds if the size of the largest cluster is of order smaller than N. This condition may be extended to allow short-distance dependence among clusters.

Interestingly, in their testing of beta-pricing models, Raponi, Robotti and Zaffaroni (2020) impose a very similar condition, their (22), that

$$\sum_{i \neq j} |E[\varepsilon_i \varepsilon_j]| = o(N),$$

and they note that it is a condition of sufficient weak correlation and a condition weaker than the one behind the arbitrage pricing theory (APT) of Ross (1976). The above condition together with bounded variances clearly imply our condition (20).

Assume further that the average beta converges,

$$\bar{\beta} = (\beta_1 + \cdots + \beta_N)/N \longrightarrow \beta_0 > 0, \tag{21}$$

and the cross sectional variation in the betas has a finite variance. Then, we have

Proposition 3: Under assumptions (19), (20) and (21), the $1/N$ is asymptotically optimal, i.e.,

$$SR_{1/N} = SR + O(N^{-1/2}), \tag{22}$$

where $SR_{1/N}$ is the Sharpe ratio of the $1/N$ rule.

It will be of interest to examine some special cases. Consider first the case in which the CAPM is true so that $R_q = R_m - r_f$. Then the excess return of holding an equal-weighted portfolio of the assets will the same as holding a portion of the market excess return plus an equal-weighted portfolio of the idiosyncratic risks,

$$\frac{1}{N} 1_N^\top R - r_f = \bar{\beta}(R_m - r_f) + \frac{1}{N} 1_N^\top \varepsilon.$$

Since the idiosyncratic component is uncorrelated with the market and its variance approaches zero due to enough diversification, then the $1/N$ portfolio excess return will be equivalent to holding the market excess return as the proportion constant will not affect the Sharpe ratio. Hence, the $1/N$ tends to optimal.

The intuition is that, if there is one factor that prices all the assets, the $1/N$ portfolio will be a portfolio of that factor and idiosyncratic noises. As long as the idiosyncratic risks are diversified away with large N, the $1/N$ is equivalent to the efficient portfolio, and so it achieves optimality.

Empirically, it is known that it is very difficult to beat the market (see, e.g., Harvey and Liu (2020)). Consistent with this, He, He, Huang and Zhou (2021) find that factor models beyond the CAPM provide little reduction in pricing errors for individual stocks. Also the CAPM is the major factor model investors and fund managers care the most (Berk and van Binsbergen (2016)). Therefore, conditional on the CAPM is difficult to beat, Proposition 3 explains why the $1/N$ is hard to beat too in practice.[4]

In the case of applying the $1/N$ to a large set of representative assets, the market portfolio is likely the main factor if not the only one. Then the above argument will be approximately true. Another case is to apply the $1/N$ within a sector. The sector index has typically most of the systematic risk. In the absence of obvious alphas relative to the index, the $1/N$ is also difficult to beat.

Note that we have only shown the optimality of the $1/N$ rule in our simple one-factor model without pricing errors. This conclusion is clearly not true in a general multi-factor APT model. Raponi, Uppal and Zaffaron (2021) provide a related general robust framework that decompose portfolios into alpha and beta portfolios, and exploit the different economic properties of each of the components. Building on this, they show that, as N increases, their new strategy generates an economically substantial and statistically significant improvement in out-of-sample portfolio performance over existing methods.

[4]The same argument as in Proposition 3 shows that the CAPM is theoretically valid (the market index converges to the tangency portfolio) in a large economy if the value-weighted idiosyncratic risks are diversifiable enough, without the usual market-clearing and representative agent assumptions.

4. Beating the $1/N$

When $N < T$, existing rules, such as the plug-in or the GMV, may be used to combine with the $1/N$ to beat it.[5] However, when $N > T$, none of the estimated rules can be applied due to singularity in $\hat{\Sigma}^{-1}$, and a completely new approach is needed. Hence, we consider these two cases separately in this section.

4.1. Case 1: $N < T$

To improve over the $1/N$, we consider first the combination of the GMV with it,

$$\hat{w}_{g,\lambda} = \lambda \hat{\Sigma}^{-1} 1 + (1-\lambda) 1_N / N, \tag{23}$$

Recall that $\rho = SR_g/SR$ and $\delta = SR_{1/N}/SR$. Write

$$\theta_1 = N\sigma_{1/N}\left(\delta e_1 + \left(1-\delta^2\right)^{1/2} \cdot e_2\right)$$

and

$$\theta_2 = \frac{\rho}{\sigma_g} \cdot e_1 + \frac{\sigma_{1/N}^{-1} - \sigma_g^{-1}\rho\delta}{\left(1-\delta^2\right)^{1/2}} \cdot e_2 + \left(\frac{1}{\sigma_g^2}(1-\rho^2) - \frac{\left(\sigma_{1/N}^{-1} - \sigma_g^{-1}\rho\delta\right)^2}{(1-\delta^2)}\right)^{1/2} \cdot e_3.$$

We obtain first the exact distribution of $SR_{g,\lambda}$, the Sharpe ratio of the combination.

[5] The three-fund may be used for the combination too. But the exact and asymptotic distributions are too complex that limit their practical use. Hence, we omit presenting them.

Proposition 4: Assume that $T > N+2$. Then

$$SR_{g,\lambda} =_d \frac{A}{B^{1/2}},$$

where

$$A = \lambda \cdot SR \cdot (e_1^\top W^{-1}\theta_2) + (1-\lambda)\mu_{1/N}$$

and

$$B = \lambda^2 \theta_2^\top W^{-2}\theta_2 + (1-\lambda)^2\sigma_{1/N}^2 + \frac{2\lambda(1-\lambda)}{N}\theta_1^\top W^{-1}\theta_2.$$

where $W \sim Wishart(I_N/(T-1), T-1)$.

Based on the exact distribution, we can obtain its behavior in high dimension.

Proposition 5: If N/T approaches $0 < \eta < 1$ as T approaches infinity, then

$$SR_{g,\lambda} = \tau_{g,\lambda} \cdot SR + O_p(T^{-1/2})$$

where

$$\tau_{g,\lambda} = \frac{\lambda \sigma_g^{-1}\sigma_{1/N}^{-1} SR_g/(1-\eta) + (1-\lambda)SR_{1/N}}{\left(\lambda^2\sigma_g^{-2}\sigma_{1/N}^{-2}/(1-\eta)^3 + (1-\lambda)^2 + 2\lambda(1-\lambda)\sigma_{1/N}^{-2}/(1-\eta)\right)^{1/2}}.$$

It is interesting that the fraction of the achievable Sharpe ratio, $\tau_{g,\lambda}$, is independent of SR conditional on SR_g and $SR_{1/N}$. This is an extension of such a relation for τ_g of the GMV rule $\hat{w}_g$, but is different from the plug-in rule $\hat{w}$.

To implement $\hat{w}_{g,\lambda}$, we choose λ to maximize the asymptotic Sharpe ratio, namely $\tau_{g,\lambda}$. Based on Proposition 5, if $\sigma_g SR_{1/N} \neq \sigma_{1/N} SR_g$, then $\tau_{g,\lambda}$ is uniquely maximized at

$$\lambda^* = \left(1 + \frac{(1-\eta)\sigma_{1/N}^{-1} SR_g - \sigma_g^{-1} SR_{1/N}}{(1-\eta)^2(\sigma_g SR_{1/N} - \sigma_{1/N} SR_g)}\right)^{-1}$$

which is straightforward to estimate in practice.

Now we consider the combination of the plug-in with naive diversification,

$$\hat{w}_\lambda = \lambda \hat{w} + (1-\lambda) 1_N / N, \tag{24}$$

where λ is a parameter between 0 and 1.[6] Tu and Zhou (2011) is the first to study the performance of such a combination. There are two major differences between our focus and theirs. First, they focus on maximizing the expected utility, and we focus on maximizing the expected Sharpe ratio. Second, they do not solve λ in terms of a few key parameters and do not have analytical results on the performance. In contrast, we characterize both the exact distribution and the asymptotic one of the Sharpe ratio, and solve λ explicitly in the high dimensional case. In the low dimensional case, λ can be solved numerically by line search to maximize the expected Sharpe ratio based on its exact distribution which can be computed easily.

Consider first the exact distribution of the Sharpe ratio of $\hat{w}_\lambda$. We have

Proposition 6: Assume that $T > N+2$. Let $W \sim Wishart(I_N/(T-1), T-1)$ and $X \sim N(0, I_N/T)$ be independent of each other. Then

$$SR_\lambda =_d \frac{A}{B^{1/2}}, \tag{25}$$

where

$$A = \lambda SR^2 (e_1^\top W^{-1} e_1) + \lambda SR (e_1^\top W^{-1} X) + (1-\lambda)\mu_{1/N}$$

[6] Similar to the three-fund case, two different combination coefficients may be used. But for our purposes, they are simplified to one.

and

$$
\begin{aligned}
B = \lambda^2 SR^2 (e_1^\top W^{-2} e_1) \quad &+ \quad \lambda^2 X^\top W^{-2} X + 2\lambda^2 SR (e_1^\top W^{-2} Z) + (1-\lambda)^2 \sigma_{1/N}^2 \\
&+ \quad 2\lambda(1-\lambda)\delta\sigma_{1/N}(e_1^\top W^{-1} X) \\
&+ \quad 2\lambda(1-\lambda)\sigma_{1/N}(1-\delta^2)^{1/2}(e_2^\top W^{-1} X) \\
&+ \quad 2\lambda(1-\lambda)\mu_{1/N}(e_1^\top W^{-1} e_1) \\
&+ \quad 2\lambda(1-\lambda)\sigma_{1/N}(1-\delta^2)^{1/2} \cdot SR \cdot (e_2^\top W^{-1} e_1),
\end{aligned}
$$

Unlike the case for the three-fund rule, the exact distribution here depends not only on the usual mean-variance frontier parameters, but also on the across section average expected asset returns, $1^\top \mu/N$. This is in fact not surprising because it matters to the $1/N$ rule. Nevertheless, similar to the three-fund case, the exact distribution can be evaluated at arbitrary accuracy with Monte Carlo integration.

Based on Proposition 6, we can derive further the asymptotic distribution in the high dimensional case.

Proposition 7: Assume that N/T approaches $0 < \eta < 1$ as T approaches infinity. Then

$$SR_\lambda = \tau_\lambda SR + O_p\left(\frac{1}{\sqrt{T}}\right), \tag{26}$$

where

$$\tau_\lambda = \frac{\lambda \cdot SR^2/(1-\eta) + (1-\lambda)\mu_{1/N}}{\left((1-\lambda)^2\sigma_{1/N}^2 + 2\lambda(1-\lambda)\mu_{1/N}/(1-\eta) + \lambda^2 SR^2/(1-\eta)^3 + \lambda^2\eta/(1-\eta)^3\right)^{1/2}}. \tag{27}$$

To implement $\hat{w}_\lambda$, we choose λ to maximize the asymptotic Sharpe ratio. Based on Proposition

7, if $1/N$ is not optimal, i.e., $SR_{1/N} < SR$, then τ_λ is uniquely maximized at

$$\lambda^* = \left(1 + \frac{\eta SR_{1/N}(1+SR^2)}{(1-\eta)^2\sigma_{1/N}(SR^2 - SR_{1/N}^2)}\right)^{-1},$$

which is straightforward to estimate in practice.

Note that, when $\eta < 1$, the Sharpe ratio of the optimized combination rule, $\max_\lambda SR_\lambda$, is always theoretically better than the naïve diversification if the latter is suboptimal already, i.e., $\delta < 1$. However, in practice, λ^* is estimated and the performance will depend on the estimation accuracy.

4.2. Case 2: $N > T$

To improve the $1/N$, the early combination idea is still useful, though the plug-in and the like cannot be used. What we need to do is to find a new portfolio of assets to combine with the $1/N$ instead of using existing estimated portfolios which require $N < T$, but now it is the case $N > T$.

The question is what assets that are likely to add investment value relative to the $1/N$. Taking $1/N$ portfolio as the benchmark and denoting its excess return as R_B, we consider the following projection of asset excess returns on R_B,

$$R_i = \alpha_i + \beta_i R_B + \varepsilon_i, \quad i = 1, 2, \ldots, N. \tag{28}$$

It is clear that those assets with zero alphas cannot add any value to R_B. If $\alpha_i > 0$ for some i, then adding the pure alpha-generating asset

$$R_{\alpha_i} = R_i - \beta_i R_B$$

will surely improve R_B in yielding a greater Sharpe ratio. If $\alpha_i < 0$, shorting of the asset properly will improve the Sharpe ratio too. Hence, an idea to improve over $1/N$ is to long positive alpha assets and short negative alpha ones.

To make the investment strategy feasible, we define it in terms of the estimated alphas from (28). Let

$$\Pi = \{\hat{\alpha}_i > \xi \cdot se(\hat{\alpha}_i)\} \cup \{\hat{\alpha}_i < -\xi \cdot se(\hat{\alpha}_i)\} \tag{29}$$

be the set of extreme estimated alphas, $se(\hat{\alpha})$ is the standard error, and $\xi > 0$ is the threshold. If Π is not an empty set, we define

$$R_{\hat{\varphi}} = \sum_{\hat{\alpha}_i > \xi \cdot se(\hat{\alpha}_i)} R_{\hat{\alpha}_i} - \sum_{\hat{\alpha}_i < -\xi \cdot se(\hat{\alpha}_i)} R_{\hat{\alpha}_i}, \tag{30}$$

where $R_{\hat{\alpha}_i}$ is the asset that generates a pure alpha of $\hat{\alpha}_i$ relative to benchmark R_B. It is clear that $R_{\hat{\varphi}}$ is a trading strategy that long positive alpha assets above the threshold, and short those below the threshold. It is a simple method to exploit the large alphas after adjusted by their standard errors.

If $\xi = 1.96$, each selected asset is likely to have nonzero alphas, but there are multiple assets, and so 5% of them will likely to have zero alphas. If the latter assets are selected, they are likely to contribute only to the estimation errors, making the combination strategy worse than otherwise. On the other hand, the $1/N$ is very difficult to beat when N is large under the one-factor model (as shown in the previous section). Hence, in this paper, we focus only on assets with extreme alphas, whose presence indicates that the $1/N$ is not optimal. If $1/N$ is indeed optimal, then all alphas are zero. It is well known that in this case

$$\max_i \hat{\alpha}_i / se(\hat{\alpha}_i) = \sqrt{2 \log N}(1 + o_p(1)),$$

assuming that they are mutually independent. See, e.g., de Haan and Ferreira (2006). Motivated by this fact, we shall therefore set ξ to be slightly larger than $\sqrt{2 \log N}$. In doing so, we improve the $1/N$ only when there is a high likelihood of beating it or when there exists extreme alphas. In this case, we ensure that $E[R_{\hat{\varphi}}] > 0$.

Consider now adding $R_{\hat{\varphi}}$ to $1/N$, or a portfolio of R_B and $R_{\hat{\varphi}}$,

$$R_p = w_1 R_B + w_2 R_{\hat{\varphi}}. \tag{31}$$

Since they are uncorrelated, the optimal portfolio weights are

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \frac{1}{\gamma} \begin{bmatrix} \frac{E[R_B]}{\mathrm{var}[R_B]} \\ \frac{E[R_{\hat{\varphi}}]}{\mathrm{var}[R_{\hat{\varphi}}]} \end{bmatrix}, \tag{32}$$

and the resulting squared Sharpe ratio is

$$(\text{Sharpe Ratio})^2 = \frac{(E[R_B])^2}{\mathrm{var}[R_B]} + \frac{(E[R_{\hat{\varphi}}])^2}{\mathrm{var}[R_{\hat{\varphi}}]}. \tag{33}$$

Hence, we have the following

Proposition 8: Let $\xi = \sqrt{2(1+\varepsilon)\log N}$ for an arbitrarily fixed $\varepsilon > 0$ (say, $\varepsilon = 0.005$). If Π is not an empty set, then

$$SR_{\xi,\hat{\varphi}} > SR_{1/N}$$

with probability approaching one as N increases, where $SR_{\xi,\hat{\varphi}}$ is the Sharpe ratio of adding $R_{\hat{\varphi}}$ to the $1/N$ strategy.

Admittedly, the condition of Proposition 6 may be weakened substantially. Nevertheless, it identifies one important condition under which the $1/N$ can be beaten. The large threshold $\xi = \sqrt{2.01 \log N}$ ensures that there is virtually a probability of 100% that the $E[R_{\hat{\varphi}}] > 0$. In this case, the $1/N$ cannot be an efficient portfolio, and so there is a high likelihood that we can improve it.

5. Conclusion

The modern portfolio theory pioneered by Markowitz (1952) is widely used in practice and extensively taught at schools. Yet, the estimated Markowitz portfolio rule and most of its extensions underperform the naïve $1/N$ rule (that invests equally across N risky assets) in many practical data sets. In this paper, we provide a number of analytical insights on why the estimated rules perform poorly and why the $1/N$ is hard to beat.

First, as long as the dimensionality is high relative to sample size, we show that the usual estimated rules are biased even asymptotically due to estimation errors. Second, we show that the $1/N$ rule is optimal in a one-factor model with diversifiable risks as dimensionality increases, irrespective of the sample size, making investment theory-based rules inadequate as they suffer from estimation errors. Third, we provide strategies that can outperform the $1/N$ under suitable conditions. As future research, while our study offers improved understanding of the $1/N$, much work remains to be done to show the value of investment theory by developing more theory-based strategies to outperform the naïve diversification.

A Proofs

Proof of Proposition 1: Write $\alpha = (1-\lambda)/\gamma$, then

$$\hat{w} = \alpha \hat{\Sigma}^{-1} \hat{\mu} + \lambda \hat{\Sigma}^{-1} 1. \tag{A.1}$$

Now we write

$$\mathbf{r}_t = \Sigma^{1/2} X_t + \mu, \qquad t = 1, \ldots, T. \tag{A.2}$$

Then, we have

$$\hat{w}^\top \mu = \alpha \mu_*^\top S^{-1} \mu_* + \alpha \mu_*^\top S^{-1} \bar{X} + \lambda \mu_*^\top S^{-1} \Sigma^{-1/2} 1$$

and

$$\begin{aligned} \hat{w}^\top \Sigma \hat{w} \;\; = \;\; & \lambda^2 1^\top \Sigma^{-1/2} S^{-2} \Sigma^{-1/2} 1 \\ & + 2\alpha\lambda \mu_*^\top S^{-2} \Sigma^{-1/2} 1 + 2\alpha\lambda \bar{X}^\top S^{-2} \Sigma^{-1/2} 1 \\ & + \alpha^2 \mu_*^\top S^{-2} \mu_* + \alpha^2 \bar{X}^\top S^{-2} \bar{X} + 2\alpha^2 \mu_*^\top S^{-2} \bar{X}, \end{aligned}$$

where $(T-1)S \sim \text{Wishart}(I, T-1)$ and $\bar{Z} \sim N(0, I/T)$ are mutually independent. Let U be a orthonormal matrix whose first column is

$$\frac{\Sigma^{-1/2}(\alpha\mu + \lambda 1)}{\sqrt{(\alpha\mu + \lambda 1)^\top \Sigma^{-1} (\alpha\mu + \lambda 1)}}$$

and second column is $a/\|a\|$ if $\lambda \neq 0$, where

$$a = \mu_* - \frac{\alpha \mu^\top \Sigma^{-1} \mu + \lambda \mu^\top \Sigma^{-1} 1}{(\alpha\mu + \lambda 1)^\top \Sigma^{-1} (\alpha\mu + \lambda 1)} \cdot (\Sigma^{-1/2}(\alpha\mu + \lambda 1)).$$

Observe that $W = U^\top S U \sim \text{Wishart}(I/(T-1), T-1)$ and $X = \sqrt{T} U^\top \bar{X} \sim N(0, I_N)$. Furthermore, we have

$$\begin{aligned}
\hat{w}^\top \Sigma \hat{w} &= (\alpha^2 \mu^\top \Sigma^{-1} \mu + \lambda^2 1^\top \Sigma^{-1} 1 + 2\alpha\lambda \mu^\top \Sigma^{-1} 1)(e_1 W^{-2} e_1) + \frac{\alpha^2}{T} X^\top W^{-2} X \\
&\quad + \frac{2\alpha}{\sqrt{T}} (\alpha^2 \mu^\top \Sigma^{-1} \mu + \lambda^2 1^\top \Sigma^{-1} 1 + 2\alpha\lambda \mu^\top \Sigma^{-1} 1)^{1/2} e_1^\top W^{-2} X
\end{aligned}$$

and

$$\begin{aligned}
\hat{w}^\top \mu &= (\alpha \mu^\top \Sigma^{-1} \mu + \lambda \mu^\top \Sigma^{-1} 1) \cdot (e_1^\top W^{-1} e_1) \\
&\quad + \lambda \left((\mu^\top \Sigma^{-1} \mu)(1^\top \Sigma^{-1} 1) - (\mu^\top \Sigma^{-1} 1)^2 \right)^{1/2} \cdot (e_2^\top W^{-1} e_1) \\
&\quad + \frac{\alpha}{\sqrt{T}} \frac{\alpha \mu^\top \Sigma^{-1} \mu + \lambda \mu^\top \Sigma^{-1} 1}{(\alpha^2 \mu^\top \Sigma^{-1} \mu + \lambda^2 1^\top \Sigma^{-1} 1 + 2\alpha\lambda \mu^\top \Sigma^{-1} 1)^{1/2}} \cdot (e_1^\top W^{-1} X) \\
&\quad + \frac{\alpha\lambda}{\sqrt{T}} \left(\frac{(\mu^\top \Sigma^{-1} \mu)(1^\top \Sigma^{-1} 1) - (\mu^\top \Sigma^{-1} 1)^2}{\alpha^2 \mu^\top \Sigma^{-1} \mu + \lambda^2 1^\top \Sigma^{-1} 1 + 2\alpha\lambda \mu^\top \Sigma^{-1} 1} \right)^{1/2} \cdot (e_2^\top W^{-1} X).
\end{aligned}$$

This accomplishes the proof. □

Proof of Proposition 2: Noting that

$$\mathbb{E}(e_1^\top W^{-1} e_2) = 0$$

and

$$var(e_1^\top W^{-1} e_2) = \frac{(T-1)^2}{(T-N-1)(T-N-2)(T-N-4)},$$

we have

$$e_1^\top W^{-1} e_2 = O_p(T^{-1/2}). \tag{A.3}$$

Similarly, we have

$$e_1^\top W^{-1} e_1 = \frac{1}{1-\eta} + O_p(T^{-1/2}), \tag{A.4}$$

and

$$e_1^\top W^{-2} e_1 = \frac{1}{(1-\eta)^3} + O_p(T^{-1/2}). \tag{A.5}$$

In light of (A.5), with probability tending to 1,

$$e_1^\top W^{-2} e_1 \le \frac{2}{(1-\eta)^3}.$$

On the other hand, if $X \sim N(0,I)$ then conditional on W, $e_1^\top W^{-1} X \sim N(0, e_1^\top W^{-2} e_1)$, and so

$$\frac{e_1^\top W^{-1} X}{(e_1^\top W^{-2} e_1)^{1/2}} = O_p(1). \tag{A.6}$$

Together, we have

$$e_1^\top W^{-1} X = O_p(1). \tag{A.7}$$

Similarly, we have

$$e_1^\top W^{-2} X = O_p(1), \qquad \text{and} \qquad e_2^\top W^{-1} X = O_p(1). \tag{A.8}$$

We now consider the term $X^\top W^{-2} X$. Note that

$$\frac{1}{T}\mathbb{E}[X^\top W^{-2} X] = \frac{1}{T}\mathbb{E}[\mathbb{E}[X^\top W^{-2} X|W]] = \frac{1}{T}\mathbb{E}[\mathrm{tr}(W^{-2})] \to \frac{\eta}{(1-\eta)^3}.$$

On the other hand,

$$\frac{1}{T}\mathbb{E}[X^\top W^{-2} X - \mathbb{E}[X^\top W^{-2} X]]^2 = \frac{2}{T}\mathbb{E}[\mathrm{tr}(W^{-4})] \to \frac{\eta}{(1-\eta)^5}.$$

Together with (A.3) and (A.4), we derive that

$$A = ((1-\lambda)SR^2 + \lambda\gamma\sigma_g^{-1}SR_g)\cdot(1-\eta)^{-1} + O_p(T^{-1/2})$$

and

$$B = ((1-\lambda)^2SR^2 + \lambda^2\gamma^2\sigma_g^{-2} + 2(1-\lambda)\lambda\gamma\sigma_g^{-1}SR_g)(1-\eta)^{-3} + (1-\lambda)^2\eta(1-\eta)^{-3} + O_p(T^{-1/2})$$

where A and B are defined in Proposition 1. This concludes the proof. □

Proof of Proposition 3: Based on the factor model, we evaluate the expected return and variance risk of the $1/N$ portfolio as

$$\mu^\top 1/N = \bar{\beta}\mu_q, \qquad \text{and} \qquad 1^\top\Sigma 1/N^2 = \sigma_q^2\bar{\beta}^2 + 1^\top(\Sigma_\varepsilon)1/N^2.$$

In particular, if $\beta_0 > 0$ and the betas has a finite variance σ_β^2, we have

$$\bar{\beta} = \beta_0 + O(N^{-1/2}), \qquad \text{and} \qquad N\bar{\beta}^2 = \mu_\beta^2 + O(N^{-1/2}).$$

This implies that

$$SR_{1/N} = \frac{\mu_q}{\sigma_q} + O(N^{-1/2}). \tag{A.9}$$

The proof of the statement then follows. □

Proof of Proposition 4: Recall that there exist two independent random variables $S \sim Wishart(I, T-1)$ and $\bar{Z} \sim N(0, I/T)$ such that

$$\hat{\Sigma} = \Sigma^{1/2}S\Sigma^{1/2} \qquad \text{and} \qquad \hat{\mu} = \Sigma^{1/2}\bar{Z} + \mu.$$

Then

$$\hat{w} = \lambda \Sigma^{-1/2} S^{-1} \Sigma^{-1/2} 1 + (1-\lambda) 1/N.$$

It can be derived that

$$\mu^\top \hat{w} = \lambda \mu_*^\top S^{-1} \Sigma^{-1/2} 1 + (1-\lambda) \mu_{1/N}.$$

and

$$\hat{w}^\top \Sigma \hat{w} = \lambda^2 1^\top \Sigma^{-1/2} S^{-2} \Sigma^{-1/2} 1 + (1-\lambda)^2 \sigma_{1/N}^2 + 2\lambda(1-\lambda) 1^\top \Sigma^{1/2} S^{-1} \Sigma^{-1/2} 1$$

Now let U be an orthonormal matrix whose first column is $\mu_*/\|\mu_*\|$, second column is $a/\|a\|$ where

$$a = \Sigma^{1/2} 1 - \frac{\mu^\top 1}{\mu^\top \Sigma^{-1} \mu} \cdot \mu_*,$$

and third column is $b/\|b\|$ where

$$\begin{aligned} b = \Sigma^{-1/2} 1 - &\frac{((1^\top \Sigma 1)(\mu^\top \Sigma^{-1} 1) - N\mu^\top 1)\mu_* + (N(\mu^\top \Sigma^{-1} \mu) - (\mu^\top 1)(\mu^\top \Sigma^{-1} 1))\Sigma^{1/2} 1}{(\mu^\top \Sigma^{-1} \mu)(1^\top \Sigma 1) - (\mu^\top 1)^2} \\ &= \Sigma^{-1/2} 1 - \frac{\mu^\top \Sigma^{-1} 1}{\mu^\top \Sigma^{-1} \mu} \cdot \mu_* - \frac{N(\mu^\top \Sigma^{-1} \mu) - (\mu^\top \Sigma^{-1} 1)(\mu^\top 1)}{(1^\top \Sigma 1)(\mu^\top \Sigma^{-1} \mu) - (\mu^\top 1)^2} \cdot a. \end{aligned}$$

It can be derived that $U^\top \Sigma^{-1/2} \mu = \|\mu_*\| e_1$, $U \Sigma^{1/2} 1 = \theta_1$ and $U \Sigma^{-1/2} 1 = \theta_2$. Write $W = U^\top S U$ and $X = \sqrt{T} U^\top \bar{Z}$. Then

$$\mu^\top \hat{w} = \lambda SR(e_1^\top W^{-1} \theta_2) + (1-\lambda) \mu_{1/N},$$

and

$$\hat{w}^\top \Sigma \hat{w} = \lambda^2 \theta_2^\top W^{-2} \theta_2 + (1-\lambda)^2 \sigma_{1/N}^2 + 2\lambda(1-\lambda) \theta_1^\top W^{-1} \theta_2 / N,$$

which completes the proof. □

Proof of Proposition 5: With what is shown before, we have

$$e_1^\top W^{-1}e_1 = \frac{1}{1-\eta} + O_p(T^{-1/2}), \qquad e_1^\top W^{-2}e_1 = \frac{1}{(1-\eta)^3} + O_p(T^{-1/2})$$

$$e_1^\top W^{-1}e_2 = O_p(T^{1/2}), \qquad e_1^\top W^{-1}X = O_p(1), \qquad e_1^\top W^{-2}X = O_p(1),$$

and

$$X^\top W^{-2}X = \frac{\eta}{(1-\eta)^3} + O_p(T^{-1/2}),$$

where $X \sim N(0,I)$ and $W \sim Wishart(I/(T-1), T-1)$ are independent of each other. The statement follows from these equalities. □

Proof of Proposition 6: Note that SR_λ is also the Sharpe ratio of

$$\hat{w} = \hat{\Sigma}^{-1}\hat{\mu} + \omega 1,$$

where $\omega = (1-\lambda)/(N\lambda)$. As before, we write

$$\mathbf{r}_t = \Sigma^{1/2}X_t + \mu, \qquad t = 1, \ldots, T$$

Then, we have

$$\hat{w}^\top \mu = \mu_*^\top S^{-1}\mu_* + \mu_*^\top S^{-1}\bar{X} + \omega\mu^\top 1$$

and

$$\begin{aligned} \hat{w}^\top \Sigma \hat{w} &= \omega^2 1^\top \Sigma 1 + 2\omega\mu_*^\top S^{-1}\Sigma^{1/2}1 + 2\omega\bar{X}^\top S^{-1}\Sigma^{1/2}1 \\ &\quad + \mu_*^\top S^{-2}\mu_* + \bar{X}^\top S^{-2}\bar{X} + 2\mu_*^\top S^{-2}\bar{X}, \end{aligned}$$

where $\mu_* = \Sigma^{-1/2}\mu$.

Now let U be an orthonormal matrix whose first column is $\mu_*/\|\mu_*\|$, and second column is $a/\|a\|$, where

$$a = \Sigma^{1/2}1 - \frac{\mu^\top 1}{\mu^\top \Sigma^{-1}\mu} \cdot \mu_*. \tag{A.10}$$

Let $W = U^\top S U$ and $X = \sqrt{T} U^\top \bar{Z}$, then we have

$$\hat{w}^\top \mu = \|\mu_*\|^2 e_1^\top W^{-1} e_1 + \frac{\|\mu_*\|}{\sqrt{T}} \cdot e_1^\top W^{-1} X + \omega \mu^\top 1$$

and

$$\begin{aligned}
\hat{w}^\top \Sigma \hat{w} \quad = \quad & \omega^2 1^\top \Sigma 1 + 2\omega(\mu^\top 1) e_1^\top W^{-1} e_1 + 2\omega \left(\|\mu_*\|^2 (1^\top \Sigma 1) - (\mu^\top 1)^2 \right)^{1/2} e_1^\top W^{-1} e_2 \\
& + \frac{2\omega(\mu^\top 1)}{\sqrt{T}\|\mu_*\|} \cdot e_1^\top W^{-1} X + \frac{2\omega \left(\|\mu_*\|^2 (1^\top \Sigma 1) - (\mu^\top 1)^2 \right)^{1/2}}{\sqrt{T}\|\mu_*\|} \cdot e_2^\top W^{-1} X \\
& + \|\mu_*\|^2 \cdot e_1^\top W^{-2} e_1 + \frac{1}{T} \cdot X^\top W^{-2} X + \frac{2\|\mu_*\|}{\sqrt{T}} \cdot e_1^\top W^{-2} X.
\end{aligned}$$

The proof then follows from the facts that $\|\mu_*\| = SR$ and

$$\delta = \frac{SR_{1/N}}{SR} = \frac{\mu^\top 1}{(\mu^\top \Sigma^{-1}\mu)^{1/2} (1^\top \Sigma 1)^{1/2}}. \tag{A.11}$$

□

Proof of Proposition 7: The proof is similar to that of Proposition 2 and 5, and therefore omitted for brevity. □

Proof of Proposition 8: Denote by $\mathscr{A}$ the collection of the indices of the assets whose ω is zero or negative. Note that, if $\omega_i = 0$, we have

$$\mathbb{P}[\hat{\omega}_i > \sqrt{2(1+\varepsilon)\log N} \cdot se(\hat{\omega}_i)] \leq N^{-(1+\varepsilon)}. \tag{A.12}$$

Therefore, with probability tending to one, for all $i \in \mathscr{A}$, we obtain

$$\hat{\omega}_i < \sqrt{2(1+\varepsilon)\log N} \cdot se(\hat{\omega}_i).$$

Denote by $\mathscr{E}$ this event. Let $\mathscr{S}$ the indices of assets such that

$$\hat{\omega}_i > \sqrt{2(1+\varepsilon)\log N} \cdot se(\hat{\omega}_i).$$

Then under $\mathscr{E}$, for any $i \in \mathscr{S}$, $\omega_i > 0$.

Observe that $\hat{\varphi}$ is an equal-weighted portfolio over all assets from $\mathscr{S}$, and so its alpha is given by

$$\frac{1}{|\mathscr{S}|} \sum_{i \in \mathscr{S}} \omega_i > 0. \tag{A.13}$$

In light of (33), we conclude that

$$SR_{\lambda,\hat{\varphi}} \geq SR_{1/N}, \tag{A.14}$$

where the equality holds if and only $\mathscr{S}$ is an empty set. □

References

Ao, M., Y. Li, X. Zhen, 2019, Approaching mean-variance efficiency for large portfolios, Review of Financial Studies 32, 2890–2919.

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

Bawa, V. S., S. J. Brown, R. W. Klein, 1979, Estimation risk and optimal portfolio choice. North-Holland, Amsterdam.

Berk, B., J. H. van Binsbergen, 2016, Assessing asset pricing models using revealed preference, Journal of Financial Economics 119, 1–23

Bodnar, T., N. Parolya, W. Schmid, 2018, Estimation of the global minimum variance portfolio in high dimensions, European Journal of Operational Research 266, 371–390.

Brown, S. J., 1976, Optimal portfolio choice under uncertainty: A Bayesian approach. Ph.D. dissertation, University of Chicago.

DeMiguel, V., L. Garlappi, R. Uppal, 2009, Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy? Review of Financial Studies 22, 1915–1953.

Duchin, R., H. Levy, 2009, Markowitz versus the Talmudic portfolio diversification strategies, Journal of Portfolio Management 35, 71–74.

Grinold, R. C., R. N. Kahn, 1999, Active Portfolio Management: Quantitative Theory and Applications. McGraw-Hill, New York.

de Haan, L., A. Ferreira, 2006, Extreme Value Theory: An Introduction. Springer, New York.

Hafner, C., L. Wang, 2021, Dynamic portfolio selection with sector-specific regularization, working paper.

Harvey, C. R., Liu, Y., 2020, Lucky factors, Journal of Financial Economics (forthcoming).

He, A., S. He, D. Huang, G. Zhou, 2021, An economic specification test of asset pricing models, working paper.

Huang, C., R. Litzenberger, 1988, Foundations for Financial Economics, North-Holland.

Jaganathan, R., T. Ma, 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, Journal of Finance 58, 1651–1683.

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

Kan, R., X. Wang, X. Zheng, 2020, In-sample and out-of-sample sharpe ratios of multi-factor asset pricing models, working paper.

Kan, R., X. Wang, G. Zhou, 2021, Optimal portfolio choice with estimation risk: No risk-free asset case, Management Science (forthcoming)

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

Kempf, A., C. Memmel, 2006, Estimating the global minimum variance portfolio, Schmalenbach Business Review 58, 332–348.

Kirby, C., B. Ostdiek, 2012, It's all in the timing: simple active portfolio strategies that outperform naïve diversification, Journal of Financial and Quantitative Analysis 47, 437–467.

Lassance, N., 2021, Maximizing the out-of-sample Sharpe ratio, working paper.

Lassance, N. Martin-Utrera, A., Simaan, M., 2021, A robust approach to optimal portfolio choice with parameter uncertainty, working paper.

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

Ledoit, O., M. Wolf, 2017, Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks, Review of Financial Studies 30, 4349–4388.

Litterman, B., 2003, Modern Investment Management: An Equilibrium Approach. Wiley, New York.

MacKinlay, A. C., Ľ. Pástor, 2000, Asset pricing models: Implications for expected returns and portfolio selection, Review of Financial Studies 13, 883–916.

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

Meucci, A., 2005, Risk and Asset Allocation, Springer-Verlag, New York.

Qian, E., R. Hua, E. Sorensen, 2007, Quantitative Equity Portfolio Management: Modern Techniques and Applications, Chapman & Hall, New York.

Ross, S. A., 1976, The arbitrage theory of capital asset pricing, Journal of Economic Theory 13, 341–360.

Raponi, V., C. Robotti, P. Zaffaroni, 2020, Testing beta-pricing models using large cross-sections, Review of Financial Studies 33, 2796–2842.

Raponi V., R. Uppal, P. Zaffaroni, 2021, Robust Portfolio Choice, working paper.

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