# Why Naive $1/N$ Diversification Is Not So Naive, 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

**Abstract**

We show theoretically that the usual estimated investment strategies will not achieve the optimal Sharpe ratio when the dimensionality is high relative to sample size, and the $1/N$ rule is optimal in a one-factor model with diversifiable risks as dimensionality increases, which explains why it is difficult to beat the $1/N$ rule in practice. We also explore conditions under which it can be beaten, and find that we can outperform it by combining it with the estimated rules when $N$ is small, and by combining it with anomalies or machine learning portfolios, conditional on the profitability of the latter, when $N$ is large.

*We are extremely grateful to Thierry Foucault (the Editor) and two anonymous referees for their very detailed and insightful comments that have helped to improve the paper enormously. We are also grateful to Victor DeMiguel, John Dooley, Lorenzo Garlappi, Raymond Kan, Nathan Lassance, Ľuboš Pástor, Christopher Reilly, Landon Ross, Jun Tu, Raman Uppal, Xiaolu Wang, Alex Weissensteiner, Michael Wolf, Paolo Zaffaroni, Yingguang Zhang, Ge Zhe, and seminar and conference participants at Capital University of Economics and Business, Boston College, Fudan University, Louisiana State University, Peking University, University of Manitoba, University of Nottingham, Washington University in St. Louis, the 15th International Conference on Computational and Financial Econometrics and 2022 China International Conference for very helpful comments. We thank Songrun He for outstanding research assistance.

## I. Introduction

Portfolio choice is one of the most important aspects of investment theory, and the mean-variance framework pioneered by Markowitz (1952) is the major model used in asset allocation and active portfolio management.[1] However, to implement the mean-variance optimal portfolio, both the asset expected returns and covariance matrix must be estimated, introducing the well-known parameter uncertainty or estimation risk problem. Brown (1976), Bawa, Brown, and Klein (1979), Jorion (1986), Jagannathan and Ma (2003), MacKinlay and Ľuboš Pástor (2000), Ledoit and Wolf (2004) and Kan and Zhou (2007) are examples of early studies that provide various strategies to mitigate the estimation risk, but the performances of these strategies is sample size dependent. In their highly influential study, DeMiguel, Garlappi, and Uppal (2009) compare these advanced strategies with the naive $1/N$ 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.*" (p. 1915)

Their finding raises a serious issue on the value of investment theory as the sample size required is too large for the theory to be reliably 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, Martín-Utrera, and Nogales (2015), Ledoit and Wolf (2017),

[1] See, e.g., Grinold and Kahn (1999), Meucci (2005), Chincarini and Kin (2023) and references therein.

[2] Their paper has over 3809 Google citations. The $1/N$ strategy was known as the Talmud rule more than 1500 years ago (Duchin and Levy (2009)).

DeMiguel, Martín-Utrera, Nogales, and Uppal (2020), Shi, Shu, Yang, and He (2020), Lassance, Martín-Utrera, and Simaan (2022), and Kan, Wang, and Zhou (2022), among others, propose additional portfolio strategies that improve performances. However, it is still puzzling why those strategies can beat the $1/N$ rule in some data sets but not in others.

In this paper, we provide a deep theoretical understanding of both the $1/N$ rule and some of the major related portfolio strategies, which yields insights on why the estimated rules often cannot beat the $1/N$ rule. We also explore conditions under which we can beat the $1/N$ rule conditionally. Specifically, we make three major contributions to the literature. First, we obtain an asymptotic distribution of the Sharpe ratio of the three-fund rule proposed by Kan and Zhou (2007), which combines the popular plug-in rule with the widely used estimated global minimum variance (GMV) portfolio. The asymptotic distribution, extending Ao, Li, and Zhen (2019), concerns the case when $N$ is large, but $N < T$ and $N/T$ approaches a constant as the sample size $T$ increases to infinity. This large dimension case is quite relevant in practice. We show that the three-fund rule does not achieve the optimal Sharpe ratio in this large $N$ case as $T$ approaches infinity, providing insights on why usual estimated rules perform poorly in high dimensions.

The second contribution of our paper is to show that the $1/N$ rule is optimal when $N$ is large enough (irrespective of sample size), under the condition that the asset returns are governed by a one-factor model with diversifiable risks. This result seems surprising and powerful. If there is one factor that prices all the assets with diversifiable risks, then the $1/N$ portfolio must be equivalent to the factor plus idiosyncratic risks. Since the risks are diversifiable, they will not alter the Sharpe ratio of the $1/N$ portfolio in the limit as $N$ approaches infinity. Hence, the Sharpe ratio of the $1/N$ portfolio must approach that of the factor, and thus must be optimal.

We also illustrate the above theoretical finding on the $1/N$ by using two examples: one of which is a calibrated economy and the other is applying the $1/N$ rule to a set of randomly selected stocks from the S&P500. The calibrated economy shows that the $1/N$ converges fast to the optimal portfolio. In the second example, although the one-factor assumption is not necessarily true, the $1/N$ rule performs almost as well as the value-weighted index. However, when we compare it with an in-sample proxy of the unknown optimal portfolio of the S&P500, the $1/N$ rule underperforms, suggesting that the one-factor assumption is unlikely perfect for the stocks in the S&P500. But the in-sample proxy is infeasible in practice, consistent with the fact that the S&P500 is difficult to beat in the real world, so is the $1/N$ rule since its performance is similar to that of the S&P500.

Our third contribution is to propose tractable rules that can beat the $1/N$ rule under certain conditions. We consider two cases that require two fundamentally different solutions. First, when $N < T$, we consider combinations of the $1/N$ rule with either the plug-in rule or the GMV portfolio. In contrast to early studies such as Frahm and Memmel (2010) and Tu and Zhou (2011), we focus on maximizing the Sharpe ratio, which is a popular criterion in practice, instead of maximizing the expected mean-variance utility. We obtain both the exact and asymptotic distributions of the Sharpe ratios in the high dimensional case. We also solve explicitly the combination coefficients that maximize the asymptotic Sharpe ratios. We find that the combinations perform better than the $1/N$ rule when $N$ is small relative to $T$, but fail to do so when $N$ is relatively large.

In the case when $N > T$, since the sample covariance matrix is not invertible, existing studies focus on obtaining a suitable invertible matrix under various assumptions (e.g., Ledoit and Wolf, 2003, Chen and Yuan, 2016, and Bodnar, Okhrin, and Parolya, 2023). In contrast to these studies, we consider how to beat the $1/N$ rule with the use of *conditional information* that goes beyond the usual assumption of iid return data. In particular, we study the combination of the $1/N$ rule with an alpha portfolio measured against it. We examine two approaches. First, we combine it with anomalies, of which there is a large literature (e.g., Chen and Zimmermann, 2023, and references therein). Second, we combine it with the long-short portfolios from recent machine learning (ML) studies such as Gu, Kelly, and Xiu (2020), among others. We find that the combinations improve the performance of the $1/N$ rule substantially, conditional on the availability of significant anomalies or profitable ML portfolios. However, we caution with a caveat that there is theoretically no guarantee that the strong performance can persist in the future as the performance of the anomalies or ML portfolios is likely to change over time.

We next discuss the differences between our paper and existing studies, along with its limitations. We focus here on Sharpe ratios, while many studies focus on the expected mean-variance utility, which Lassance, Martín-Utrera and Simaan (2023) extend into a more robust framework by considering the uncertainty associated with utility maximization. Without estimation errors, both Sharpe ratio and utility maximization are equivalent. However, with estimation errors, the objectives are mathematically different, and the difference is analyzed by Lassance (2021). By contrast, we focus on the Sharpe ratio because it is widely used by both practitioners and researchers in comparing trading strategies and models. The Sharpe ratio is simpler and it does not require information on the risk aversion parameter (see, e.g., Barillas, Kan, Robotti, and Shanken, 2020). An additional reason is that it is analytically tractable in our context. We note that we have shown only the optimality of the $1/N$ rule in a restricted one-factor model, which is not true in a general multi-factor APT model as otherwise the APT will be reduced to a one-factor model with the $1/N$ portfolio as the factor. In the APT case, Raponi, Uppal, and

Zaffaroni (2021), not studying the $1/N$ rule, provide alternative portfolio strategies using alpha and beta portfolios. Interestingly, their strategy achieves the same Sharpe ratio as that of the $1/N$ rule under certain conditions. Finally, Ao, Li, and Zhen (2019) propose a new method to approach the unconditional efficient frontier portfolio when $N$ is large. By contrast, we add anomaly or ML portfolios, with conditional information, to obtain approximately conditional efficient portfolio to outperform the unconditional $1/N$ rule.

Our paper is also closely related to Pflug, Pichler, and Wozabal (2012) and Yan and Zhang (2017).[3] Pflug, Pichler, and Wozabal (2012) show that the $1/N$ rule is nearly optimal under high model ambiguity. Their result suggests that, if the estimation risk of an estimated rule is high enough, then the optimal combination of this rule with the $1/N$ rule will consist of almost entirely the $1/N$ rule (which is easy to show). Yan and Zhang (2017) find importantly that the $1/N$ rule is optimal in the absence of mispricing in the CAPM with diagonal idiosyncratic errors. Our one-factor optimality result, though developed independently, can be viewed as an extension of their work to the case of any asset returns as long as they can be well modeled by a one-factor model with diversifiable risks.

The rest of the paper is organized as follows. In Section 2, we discuss properties of the common estimated rules. In Section 3, we provide conditions under which the $1/N$ rule is optimal. In Section 4, we explore ways to outperform the $1/N$ rule, and we conclude in Section 5.

[3]We are grateful to an anonymous referee who informed us about these two important and related papers.

## II. Asymptotic Sharpe Ratios 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 $\mu$ and covariance matrix $\Sigma$.

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

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

where $\gamma$ is a risk aversion parameter, and the optimized Sharpe ratio is

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

However, the parameters $\mu$ and $\Sigma$ are unknown, and are usually have to be estimated from data.

Suppose there are $T$ periods of observed excess returns, then 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). In order to obtain exact distributional results, they are further assumed to be normal, that is, $R_t \sim N(\mu, \Sigma)$.

The popular plug-in rule is the estimated optimal portfolio rule, obtained by replacing the unknown parameters by their sample 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 estimation errors in $\hat{w}$, making the resulting portfolio 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 is:

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

The analytical expression states that $\hat{w}_g$ is affected by errors in estimating $\Sigma^{-1}$ only, and not by errors in estimating $\mu$ unlike the plug-in rule.

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 riskfree asset. Note that the scalar of the GMV, the inverse of $1_N^\top \hat{\Sigma}^{-1} 1_N$, is absorbed into $\lambda$ so that the formula for the optimal $\lambda$ looks simpler. Unlike Kan and Zhou (2007), who use two different combination coefficients, we scale them to make the presentation easier to understand. The scaling will not affect the Sharpe ratio and the optimal combination strategy. However, as discussed at the end of this section, it will have implications on risk control.

Let $SR_\lambda$ be the Sharpe ratio of the three-fund rule $\hat{w}_\lambda$. We are now interested in the exact distribution of $SR_\lambda$, and will analyze its asymptotic distribution later. To do so, let $SR_g$ be the Sharpe ratio of the GMV portfolio,

$$SR_g = \frac{\mu_g}{\sigma_g} = \frac{\mu^\top \Sigma^{-1} 1_N}{(1_N^\top \Sigma^{-1} 1_N)^{1/2}}, \tag{9}$$

where

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

which are the expected excess return and variance of the GMV portfolio.

Then the exact distribution is given by:

**Proposition 1:** Assume that $T > N + 2$. Then the exact distribution of $SR_\lambda$ is:

$$SR_\lambda =_d \frac{A}{B^{1/2}}, \tag{10}$$

where

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

and

$$
\begin{aligned}
B &= ((1-\lambda)^2SR^2+\lambda^2\gamma^2\sigma_g^{-2}+2(1-\lambda)\lambda\gamma\sigma_g^{-1}SR_g)(e_1W^{-2}e_1)+(1-\lambda)^2(X^\top W^{-2}X)\\
&\quad+2(1-\lambda)((1-\lambda)^2SR^2+\lambda^2\gamma^2\sigma_g^{-2}+2(1-\lambda)\lambda\gamma\sigma_g^{-1}SR_g)^{1/2}(e_1^\top W^{-2}X),
\end{aligned}
$$

with $W\sim Wishart(I_N/(T-1),T-1)$ and $X\sim N(0,I_N/T)$, which are independent of each other, $I_N$ being the identity matrix of order $N$, and $e_1$ 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 determined by the three usual mean-variance frontier parameters (all proofs are provided in the appendix). The role played by $N$ and $T$ are not explicit, but implicit in the definitions of $W$ and $Z$, which are independent of any parameters involving $\mu$ and $\Sigma$. As a result, $W$ and $Z$ are easily simulated. Hence, the exact distribution can be computed with arbitrary accuracy via Monte Carlo integration for any given value of the mean-variance frontier parameters.

Let us examine some special cases. When $\lambda=0$ or for the plug-in portfolio, we have the following:

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

$$(11) \qquad SR_{\hat{w}} =_d \frac{SR^2(e_1^\top W^{-1}e_1) + SR(e_1^\top W^{-1}X)}{\left(SR^2(e_1^\top W^{-2}e_1) + X^\top W^{-2}X + 2SR(e_1^\top W^{-2}X)\right)^{1/2}}.$$

In particular, if $N/T$ approaches $\eta$, $0 < \eta < 1$, when $T$ approaches infinity, then

$$(12) \qquad SR_{\hat{w}} = \tau SR + O_p\left(\frac{1}{\sqrt{T}}\right),$$

where $\tau = \sqrt{(1-\eta)/(1+\eta/SR^2)} < 1$, and $O_p(\cdot)$ denotes a bounded function in probability.

Kan, Wang, and Zheng (2020) provide the first explicit expression for the exact distribution of $SR_{\hat{w}}$. Equation (11) provides a complementary expression. It is interesting that, given the true Sharpe ratio, $SR$, the Sharpe ratio performance of the plug-in rule has nothing to do with the expected returns or the covariance matrix. Both $N$ and $T$ play important roles and they matter only through their impact on $W$ and $X$, which summarize the effects of estimation errors. On the asymptotic limit, Ao, Li, and Zhen (2019) is the first, to our knowledge, to obtain (12); however, their focus is different from ours.

The top panel of Figure 1 illustrates Corollary 1, where we set $SR = 1$ and compute the distribution of $SR_{\hat{w}}$ with one million simulated samples so that the distributions are little different from the exact ones.[4] We consider two cases, $N = 0.3T$ and $0.8T$, respectively. In the first case, even as sample size goes from 120, 240, 480 to approach 960, the mean is clearly below 0.8, biased away from 1. The bias in the second case, as suggested by the theory, becomes greater, with a mean

[4]The exact distribution is an integral multiplied by the density functions of the normal and Wishart, which is numerically difficult to evaluate via a grid method, but easy by using the Monte Carlo integration which approximates the exact value with the average of the integrand computed over the simulated samples of the random variables.

below 0.4. The vertical line is the asymptotic limit from equation (11). The plots clearly show that, as $T$ and $N$ both increase, the plug-in rule is still biased. That is, the plug-in rule is asymptotically biased in the large $N$ case.

Note that equation (12) provides a simple and quick comparison between the plug-in and the $1/N$ rules. To see this, let $SR_{1/N}$ be the Sharpe ratio of the $1/N$ rule, and $\delta = SR_{1/N}/SR$ be the fraction of the Sharpe ratio the $1/N$ rule can achieve. Then, if

$$\eta > \frac{1-\delta^2}{1+\delta^2/SR^2}, \tag{13}$$

we have

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

i.e., the naive diversification is asymptotically superior to the plug-in portfolio.

In light of equation (13), we can understand the earlier quoted statement of DeMiguel, Garlappi, and Uppal (2009) theoretically. Assume that $\delta = 85\%$ (which is quite possible given the results in the next section) and assume that the Sharpe ratio of the optimal portfolio is $0.5/\sqrt{12}$ per month (close to that of the S&P500). Then $N/T = \eta$ needs to be smaller than 0.78% in order for $SR_{\hat{w}}$ to be greater than $SR_{1/N}$, implying a sample size of $T > 3,205$ if $N = 25$, and $T > 6,410$ if $N = 50$. This provides a theoretical confirmation of their empirical results.

When $\lambda = 1$ or for the GMV portfolio, we have the following:

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

$$SR_{\hat{g}} =_d \frac{SR}{(e_1 W^{-2} e_1)^{1/2}} \left(\rho(e_1^\top W^{-1} e_1) + (1-\rho^2)^{1/2}(e_2^\top W^{-1} e_1)\right), \tag{14}$$

where $\rho \equiv SR_g/SR$, the fraction of the population Sharpe ratio of the GMV relative to the true Sharpe ratio, which is assumed to be less than 1. In particular, if $N/T$ approaches $\eta$, $0 < \eta < 1$, when $T$ approaches infinity, then

$$SR_{\hat{g}} = \tau_g SR + O_p\left(\frac{1}{\sqrt{T}}\right), \tag{15}$$

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

Kempf and Memmel (2006) and Basak, Jagannathan, and Ma (2009), among others, study various properties of the GMV. The exact distribution, (14), complements their studies. Bodnar, Parolya, and Schmid (2018) provide estimation approaches for 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 (15) shows that the effect of the estimation errors is captured solely by the term $(1-\eta)^{1/2}$. It is interesting that the distribution of the GMV depends not only on $SR$, but also on $\rho$. 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 $SR_{\hat{w}}$ is the next simplest. In contrast, the distribution of $SR_\lambda$ is complex, where the complexity is introduced by the correlation between $\hat{w}$ and $\hat{w}_g$.

The bottom panel of Figure 1 illustrates Corollary 2, where the distributions are computed similarly to $SR_{\hat{w}}$. Interestingly, the mean of $SR_{\hat{g}}$ is more biased away from 1 than $SR_{\hat{w}}$ for the same $T$ and $N$. Intuitively, this is expected. For a fixed $N$, $SR_{\hat{w}}$ is asymptotically unbiased but

$SR_{\hat{g}}$ is always biased as $T$ approaches infinity. Now $N$ grows too, so the bias of $SR_{\hat{g}}$ gets worse. However, $SR_{\hat{g}}$ has less volatility. This is expected too because it does not involve the expected returns, whereas $SR_{\hat{w}}$ has to estimate both the expected returns and the covariance matrix of the asset returns.

There are many studies on the GMV as it is very popular in practice due to its simplicity (and less volatility) as well as the fact that it does not require estimating the means. Hafner and Wang (2023), for example, impose sector restriction on the GMV to enhance its performance. In light of the asymptotic expression, equation (15), we can see that the GMV is better asymptotically than the plug-in rule if and only if

$$\rho > \frac{SR}{\sqrt{\eta + SR^2}}. \tag{16}$$

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, then the GMV is preferred. However, this is not always the case. When $\eta$ 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.

Next we examine the limiting behavior of the general three-fund rule. 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 converges to the optimal one. But when $N$ is not fixed and grows with $T$ at a constant rate $\eta$, then there is no longer the convergence, as shown below in Proposition 2.

**Proposition 2:** If $N/T$ approaches $\eta$, $0 < \eta < 1$, when $T$ approaches infinity, then

$$SR_\lambda = \tau_\lambda SR + O_p\left(\frac{1}{\sqrt{T}}\right), \tag{17}$$

where

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

In contrast to the exact distribution in equation (10) with complex terms of $A$ and $B$, equation (17) is much simplified with only a simple scalar $\tau_\lambda$. For any given $\eta > 0$, $\tau_\lambda$ is always less than 1, implying that the three-fund rule will never converge to the true optimal Sharpe ratio as $T$ approaches infinity. Therefore, as far as the Sharpe ratio is concerned, new methods are needed to improve the performance of the three-fund rule.

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

$$\lambda^* = \frac{\eta\mu_g}{\gamma(SR^2 - SR_g^2) + \eta\mu_g}. \tag{18}$$

Given an estimate of the three mean-variance frontier parameters, $\lambda$ is straightforwardly solved from the above. Note that the optimized $\lambda$ 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.

Finally, we consider the general form of the three-fund rule of Kan and Zhou (2007),

$$\hat{w}_{\lambda_1,\lambda_2} = \lambda_1\frac{1}{\gamma}\hat{\Sigma}^{-1}\hat{\mu} + \lambda_2\hat{\Sigma}^{-1}1_N, \tag{19}$$

where $\lambda_1$ and $\lambda_2$ are free parameters. It is easy to show that their optimal choice does not matter to our earlier results, which impose constraints on their sum being 1. However, the choice does affect the leverage of the resulting portfolio. To see this, note that the Sharpe ratio is invariant to scaling, so it is only a function of $\lambda_1/\lambda_2$,

$$SR_{\hat{w}_{\lambda_1,\lambda_2}} = \frac{\frac{\lambda_1}{\lambda_2}\frac{1}{\gamma}\mu^\top\hat{\Sigma}^{-1}\hat{\mu} + \mu^\top\hat{\Sigma}^{-1}1_N}{\sqrt{\frac{\lambda_1^2}{\lambda_2^2}\frac{1}{\gamma^2}\mu^\top\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\mu} + 1_N^\top\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N + 2\frac{\lambda_1}{\lambda_2}\frac{1}{\gamma}\mu^\top\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N}}.$$

Now observe that the optimal ratio of $\lambda_1/\lambda_2$ that maximizes the asymptotic Sharpe ratio can accommodate at the same time a risk constraint,

$$\sqrt{\hat{w}_{\lambda_1,\lambda_2}^\top \Sigma \hat{w}_{\lambda_1,\lambda_2}} = \sigma_0,$$

where $\sigma_0$ is a prior risk level of the portfolio chosen by an investor. In other words, an investor can in fact achieve a double objective of maximizing the Sharpe ratio and meeting a risk constraint. This is a useful property of the three-fund rule.[5]

In summary, in this section we provide a thorough analysis on the three-fund rule, which is shown by a number of studies better than its special cases, the plug-in rule and the estimated GMV. Our message is that the three-fund rule and all of the common estimated rules are asymptotically not optimal even when the sample size is large, as long as $N$ is also large enough. On the other hand, as we show below, the $1/N$ rule can be optimal under certain conditions regardless of the sample size as long as $N$ is large enough. Then, combining these two observations, it is no wonder why the $1/N$ rule is generally hard to beat in practice when the $N$ is large.

[5] We are grateful to an anonymous referee for pointing this out.

## III. Optimality of $1/N$ Rule

The well-known naive $1/N$ investment strategy, which invests equally among $N$ risky assets, has the following constant portfolio weights,

$$w_{1/N} = \frac{1}{N} 1_N. \tag{20}$$

Since there are no parameters involved, there will be no estimation errors. However, because the $1/N$ rule generally differs from the true and usually unknown portfolio weights $w = \frac{1}{\gamma}\Sigma^{-1}\mu$, it will not be optimal unless it happens to be equal to $w$.

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

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

$$R_i = \beta_i R_q + \varepsilon_i, \quad i = 1, 2, \ldots, N, \tag{21}$$

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

$$(1_N^\top \Sigma_\varepsilon 1_N)/N^2 \longrightarrow 0, \qquad as \quad N \to \infty, \tag{22}$$

where $\Sigma_\varepsilon$ is the covariance matrix of $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)^\top$. Consider a special case. If the idiosyncratic risks are uncorrelated and have an upper bound, $\sigma^2_{\max}$, on their variances, then it is clear that

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

We will make this simple assumption here, although it can be relaxed to certain structural forms.

Interestingly, in their testing of beta-pricing models, Raponi, Robotti, and Zaffaroni (2020) impose a very similar condition, their equation (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 the bounded variance assumption will clearly imply our condition in equation (22).

Assume further that the average beta converges,

$$\bar{\beta} = (\beta_1 + \cdots + \beta_N)/N \longrightarrow \beta_0 > 0, \tag{23}$$

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

**Proposition 3:** Under the assumptions in equations (21), (22) and (23), the $1/N$ rule is asymptotically optimal when $N$ is large, i.e.,

$$SR_{1/N} = SR + O\left(\frac{1}{\sqrt{N}}\right), \tag{24}$$

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

We next 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 be 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 sufficient diversification, holding the $1/N$ portfolio will be equivalent to holding a proportion of the market portfolio, but the proportion constant will not affect the Sharpe ratio. Hence, the $1/N$ rule converges to the optimal one when $N$ approaches infinity.

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 noise. As long as the idiosyncratic risks are diversified away with large $N$, the $1/N$ rule is equivalent to the efficient portfolio, and so it achieves optimality.

Empirically, it is known that it is very difficult to beat the market (e.g., Harvey and Liu, 2021). Consistent with this, He and Zhou (2023) find that well-known factor models, such as the three-factor model of Fama and French (1993), provide little reduction in pricing errors for individual stocks. Also the CAPM is the factor model that investors and fund managers care the most (Berk and van Binsbergen, 2016). Therefore, conditional on the fact that the CAPM is difficult to beat, Proposition 3 explains why the $1/N$ rule is hard to too in practice.[6]

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

Note that we have only shown the optimality of the $1/N$ rule in our simple one-factor model that has no pricing errors. The optimality is clearly not true in a general multi-factor APT model. This is because if it were, then the $1/N$ portfolio would be an efficient portfolio pricing all the assets, reducing the model to a one-factor one and the factor is the $1/N$ portfolio, which is not true in general. Under the APT, Raponi, Uppal, and Zaffaroni (2021) provide a related general robust framework that decomposes portfolios into alpha and beta portfolios and exploits the different economic properties of each of the components. They show that, as $N$ increases, their strategy generates an economically substantial and statistically significant improvement in out-of-sample portfolio performance over existing methods.

We next provide a numerical example to see how fast the $1/N$ rule converges to the optimal portfolio in a one-factor model. To set the parameters to some realistic values, we assume that the market portfolio is the single factor and use the 25 size and book-market portfolios of Fama and French (1993) to calibrate the mean of asset betas and their standard deviation, $\bar{\beta}$ and $\sigma_\beta$; the mean

[6] 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.

of the residual volatility, $\bar{\sigma}$; and the volatility of volatility, $\sigma_\sigma$, which is the volatility of the residual volatilities across assets. The calibration is based on monthly data from 1926-07 to 2022-04. The sample mean of the market excess return and the volatility will be taken as the true parameters, $\mu_m$ and $\sigma_m$. Then we consider a series of economies with $N = 5, 10, 25, 50, 75$ and 100, respectively. Given $N$, we simulate the true beta parameters from

$$\beta_j \sim N(1, \sigma_\beta^2), \quad j = 1, 2, \ldots, N, \tag{25}$$

and then the $N$ asset expected returns will be given by $\mu = \beta\mu_m$. Draw

$$z \sim N(\bar{\sigma}, \sigma_\sigma^2). \tag{26}$$

If $z < \bar{\sigma}/2$, we re-draw, to make it realistic that no asset will have a volatility that is too small. Then we set

$$\sigma_{i,i} = z^2,$$

and for $1 \leq i \neq j \leq N$,

$$\sigma_{i,j} = \sqrt{\sigma_{i,i}\sigma_{j,j}}/(2(i-j)^2).$$

Hence, the covariance matrix of the residuals is now well defined. We simulate 10,000 sets of the parameters.

Figure 2 provides the ratio of the Sharpe ratio of the $1/N$ investment strategy to that of the optimal portfolio. The result is striking. When $N$ is as small as $N = 5$, the $1/N$ rule obtains about 90% of the true optional Sharpe ratio already. It obtains more than 95% when $N = 20$, and more than 99% when $N$ is merely 100. The numerical example shows that the $1/N$ rule converges very fast to the optimal rule in terms of the Sharpe ratio in a true one-factor model. Given that in vast applications, the first factor often dominates, which is yet another reason why the $1/N$ rule is hard to beat in practice.

Panel A of Table 1 provides the annualized Sharpe ratios from the simulations in which the true optimal Sharpe ratio is scaled to be equal to 0.5, the rough value of the US stock market (see Table 2). Consistent with Figure 2, when $N = 5$, the $1/N$ Sharpe ratio is already 0.4073. As $N$ increases to 50, the value is 0.4889, remarkably close to 0.5. When $N = 100$, the value is 0.4967, not much different from 0.5.

Now we examine how the Sharpe ratio of the $1/N$ rule is affected as the dispersion among the model parameters increases. Panel B of Table 1 shows that, once we double the volatility of the betas while keeping all else the same, the performance deteriorates. This is expected as more variations in the betas will make the diversification more difficult. However, the deterioration is small. For example, when $N = 5$, it goes down only to 0.3835 from 0.4073. When $N = 50$, it has a very small drop to 0.4850 from 0.4889. For comparison, we also examine the results when the volatility of the covariances of the residuals are doubled, with all else being the same. Panel C shows that the performance deteriorates a bit more than the beta case, but the changes are still very minor. Overall, we find that the rate of convergence of the $1/N$ rule in a true one-factor model is quite fast and robust. In short, when the one-factor model assumption is true, the $1/N$ rule seems to have a fast rate of convergence to optimality.

Consider another example in which we compare the $1/N$ rule to the S&P500 index. This comparison is of interest even if the one-factor model assumption is not true. In this case, if the

$1/N$ rule performs well relative to the index, it will suggest that it is difficult to beat the $1/N$ in practice, because the S&P500 index is hard for fund managers to beat in the real world (e.g., Harvey and Liu, 2021). Now, if the one-factor model assumption is true, then the optimal portfolio, or the maximum Sharpe ratio portfolio, will be the benchmark of interest for comparing with the $1/N$ rule. We analyze both cases below.

First, empirically, we apply the $1/N$ rule to the real data by investing into $N$ stocks randomly selected from the S&P500 index, which are also randomly replaced if some of the stocks are merged or removed from the index. The data are monthly, from 1957–03 to 2021–12, and there are 778 months in total.

Table 2 provides the results, which seem striking. When $N = 5$, the monthly return of the $1/N$ rule is 1.02% and the monthly volatility is 5.27, whose performance is already quite comparable to that of the value-weighted S&P500 index, reported in the last column. Their annualized Sharpe ratios are 0.44 and 0.49, respectively, which are remarkably close already. As the number of assets increases to 10, the Sharpe ratio is even slightly greater at 0.54. But the Sharpe ratios are in the close range as $N$ further increases to 25, and up to 500. Overall, it seems that it takes a much smaller number of stocks to replicate the performance of the value-weighted S&P500 in a frictionless world without trading costs.

It is important to emphasize that, in the real world, the $1/N$ portfolio has to be rebalanced each month, which incurs transaction costs. On the other hand, the value-weighted S&P500 index has virtually no such costs except when reinvesting dividends and when adjusting components due to occasional composition changes in the index. Hence, when transaction costs are considered, the $1/N$ portfolio is unlikely to have similar performance as the S&P500 index. But our focus here is on the comparison of the $1/N$ rule with the estimated portfolio rules which are subject the transaction costs. The point is, then, since the S&P500 index is hard to beat in the real world and $1/N$ rule is close to it, the $1/N$ rule must be hard to beat too by the estimated portfolio rules when both ignore transaction costs. Proposition 3 states that it is indeed difficult as the performance of the $1/N$ rule, which does not depend on sample size, converges to optimality when $N$ is large in the one-factor case. In contrast, the performance of any estimated rule depends on sample size $T$, and it will not converge to the optimal Sharpe ratio for large $N$ (Proposition 2) whether in a one-factor model or not. As $N/T$ is large in practice and a one-factor structure is a good proxy in many applications, both reasons help explain why the $1/N$ rule is hard to beat by common estimated rules.

In the presence of transaction costs for both the $1/N$ rule and the estimated portfolio rules, existing studies show that the $1/N$ rule has generally lower trading costs (e.g., DeMiguel, Garlappi, and Uppal, 2009, and Kan, Wang and Zhou, 2022). Hence, if both types of rules are subtracted by roughly the same amount of trading fees, their ranking will be unchanged. Once again, even with realistic transaction costs, the $1/N$ rule is still hard to beat by common estimated rules.

Now consider the hypothesis that the one-factor model is true for the 500 stocks in the S&P500 index. Based on Proposition 3, the Sharpe ratio of the $1/N$ rule should converge to that of the optimal portfolio or the maximum Sharpe ratio portfolio under the null. However, the maximum Sharpe ratio portfolio is unknown and has to be estimated. A proxy will be the in-sample optimal portfolio, the one with the unknown plug-in weights estimated using all the data, and then used to invest in all the periods. This portfolio is infeasible in practice as one does not have future data for the parameter estimation. But all the data are likely to provide a more accurate estimate of the true and unknown parameters, and hence the in-sample optimal portfolio weights based on the 778 months of data should provide a good estimate of the unknown optimal weights.

There is, however, one important complication. If the same stocks are in the index all the time, then it is straightforward to compute the in-sample optional portfolio and use it as a proxy for the true optimal portfolio. But stocks come in and out of the S&P500 index from time to time. For example, Tesla, Inc. entered the S&P 500 on December 21, 2020, making the covariance matrix difficult to estimate. While ad hoc sparsity assumptions may be imposed to make the estimation and the covariance matrix inversion possible, there are also issues in dealing with stocks that are moved out of the index, which requires additional ad hoc assumptions. Since the missing data problem is not the focus of our paper, we provide a simple solution by using the size-sorted portfolios of the 500 stocks as the underlying assets, which is robust to the index composition changes.

Specifically, we sort the 500 stocks of the S&P500 index into $N$ size portfolios and rebalance them each month, then there will no missing data problem on their sample mean and covariance matrix. Hence, the in-sample optimal portfolio is easily computed. For $N = 10$, 25, 50, and 100, the Sharpe ratios of the $N$ size portfolios are $0.97, 1.07, 1.24$ and 1.63, respectively. They range from 2 to 3 times over the Sharpe ratio of the $1/N$ rule reported in Table 2. If we impose no-short sells, the Sharpe ratios become 0.87, 0.89, 0.92, and 0.98, respectively, still substantially greater than that of the $1/N$ rule. We interpret the results as evidence that the S&P500 stocks do not follow a one-factor model. Because if it did, the Sharpe ratios would be close to that of the $1/N$ rule based on Proposition 3.

However, despite the gap between the in-sample optimal portfolio (which is infeasible) and the

$1/N$ rule, it seems very difficult to beat the $1/N$ rule in real time unconditionally (relying on the return data only) due to estimation risk (and perhaps other risks too). If it were easy to beat the $1/N$ rule, and then it would be easy to beat the market (as the market is close to the $1/N$ rule for the S&P500 stocks). But beating the market is difficult, a problem we leave as future research. In short, the $1/N$ is difficult to beat in general, when only the return data are available (without conditional information).

## IV. Beating the $1/N$ Rule in Special Cases or Conditionally

Although the $1/N$ rule is optimal in a one-factor model when $N$ is large, it is possible to improve it when $N$ is small or when the one-factor model is not true. When $N$ is small, existing rules may be combined with the $1/N$ rule. When $N$ is large, adding an alpha portfolio can be helpful in improving the performance.

### A. Combinations when $N < T$

To improve the $1/N$ rule, we examine its combinations with the GMV and the plug-in rules, respectively. Consider first the combination with the GMV,

$$\hat{w}_{g,\lambda} = \lambda \hat{\Sigma}^{-1} 1 + (1 - \lambda) 1_N / N, \tag{27}$$

where $\lambda$ is the combination parameter. Frahm and Memmel (2010), among others, provide properties of such a combination, although their objective is different from ours. Note that the combination coefficients can be unconstrained, similar to equation (19), to have a risk control target.

To present our results succinctly, we introduce two parameters,

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

and

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

which will be useful later for characterizing distributions.

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

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

$$SR_{g,\lambda} =_d \frac{A}{B^{1/2}}, \tag{28}$$

where

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

and

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

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

Based on the exact distribution, we can then derive its behavior in large sample.

**Proposition 5:** If $N/T$ approaches $\eta$, $0 < \eta < 1$, when $T$ approaches infinity, we have

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

where

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

From Proposition 5, we see that the fraction of the achievable Sharpe ratio, $\tau_{g,\lambda}$, is not a function of SR given $SR_g$ and $SR_{1/N}$. This is an extension of such a relation for $\tau_g$ 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 $\lambda$ 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

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

which is straightforward to compute in practice.

We next study the combination of the plug-in rule with the naive diversification,

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

where $\lambda$ is a parameter between 0 and 1.[7] Tu and Zhou (2011) are the first to study the performance

[7] Again, the combination coefficients can be unconstrained similar to (19) to have a risk control target.

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 $\lambda$ 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 $\lambda$ explicitly in the high dimensional case here.

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

**Proposition 6:** Assume that $T > N+2$. Then

$$SR_\lambda =_d \frac{A}{B^{1/2}}, \tag{31}$$

where

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

and

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

where $W \sim Wishart(I_N/(T-1), T-1)$ and $X \sim N(0, I_N/T)$, which are independent of each other.

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 cross-section average expected asset returns, $1^\top \mu/N$. This is not surprising because it matters to the $1/N$ rule. Nevertheless, similar to the three-fund case, the exact distribution can be evaluated with arbitrary accuracy via Monte Carlo integration.

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

**Proposition 7:** If $N/T$ approaches $\eta$, $0 < \eta < 1$, when $T$ approaches infinity, we have

$$SR_\lambda = \tau_\lambda SR + O_p\left(\frac{1}{\sqrt{T}}\right), \tag{32}$$

where

$$\tau_\lambda = \frac{\lambda \cdot SR^2/(1-\eta) + (1-\lambda)\gamma\mu_{1/N}}{\left((1-\lambda)^2\gamma^2\sigma^2_{1/N} + 2\lambda(1-\lambda)\gamma\mu_{1/N}/(1-\eta) + \lambda^2 SR^2/(1-\eta)^3 + \lambda^2\eta/(1-\eta)^3\right)^{1/2}}. \tag{33}$$

To implement $\hat{w}_\lambda$, we choose $\lambda$ to maximize the asymptotic Sharpe ratio. Based on Proposition 7, if the $1/N$ rule is not optimal, i.e., $SR_{1/N} < SR$, then $\tau_\lambda$ is uniquely maximized at

$$\lambda^* = \left(1 + \frac{1}{\gamma} \cdot \frac{\eta SR_{1/N}(1+SR^2)}{(1-\eta)^2\sigma_{1/N}(SR^2 - SR^2_{1/N})}\right)^{-1}, \tag{34}$$

which is straightforward to compute in practice.

There are interesting implications from the theoretically results. First, equations (29) and (34) make the implementation of the combination rules easy. Otherwise, complex algorithms may have to be devised to find them. Second, they provide insights on beating the $1/N$ rule. When the $\eta$ is close to 1, $\lambda$ will be close to 0, making it impossible to beat the $1/N$ rule. However, when $\eta$ is close to 1, $\lambda$ will be close to 1. If the true Sharpe ratio or the population Sharpe ratio of the GMV is greater than the population Sharpe ratio of the $1/N$ rule, then the combined portfolio will beat the $1/N$ rule when $T$ is sufficiently large.

Next we illustrate the message of the Propositions by using an example. To deviate from a one-factor model, we consider $N$ portfolios sorted by firm size, which adds effectively a size factor in additional to the usual market exposure of the stocks. With monthly data 1973-01 to 2022-03, we use an estimation window of $T = 360$,[8] leaving an out-of-sample investment period from 2003-01 to 2022-03. We examine the performance of the two combination portfolios in terms of their annualized Sharpe ratios, along with the annualized Sharpe ratios of the individual rules.

Table 3 reports the results. When $N$ is small, $N = 5$, both $\hat{w}_{g,\lambda^*}$ and $\hat{w}_{\lambda^*}$ outperform the $1/N$ rule, with annualized Sharpe ratios, 0.73 and 0.61, greater than 0.54 of the $1/N$ rule. When $N = 20$, both remain better. However, when $N = 50$, only the combination with the GMV can outperform the $1/N$ rule. This is not surprising as the estimation errors grows with $N$, and the plug-in rule suffers more than the GMV, because it has to estimate the means. Nevertheless, when $N$ is equal to 100, all the estimated rules, including the combinations, are worse than the $1/N$ rule as the estimation errors become large in the high dimensional case, consistent with what is shown theoretically by Proposition 2. Although not reported, it is clear that the estimated rules will perform much worse when $N$ increases beyond 100 (but less than 300 for the invertibility of

[8]Smaller estimation window size is not sufficient for the superior performance reported below.

the sample covariance matrix).

The asymptotic theory discussed in Section 2 can provide some insight on the above results. Given Corollary 4, we have that $SR_{\hat{g}} > SR_{1/N}$ asymptotically if

$$\eta < 1 - \left(\frac{SR_{1/N}}{SR_g}\right)^2.$$

Suppose, for example, that $SR_{1/N}/SR_g = 0.8$ (which appears reasonable given Table 3) and $N = 50$, then only a sample size $T > 138$ is needed. This sample size is not very large and the result seems to explain why the combination of the GMV and $1/N$ rules outperforms the $1/N$ rule when $N = 50$. It also suggests that the sample size required by DeMiguel, Garlappi, and Uppal (2009) may be lowered for the GMV. However, this will depends critically on $SR_{1/N}/SR_g$, and it is possible when the ratio is sufficiently below one.

Table 3 suggests that combing the GMV with the $1/N$ rule does perform better than combining the plug-in with it, when $N$ is small. However, Lassance, Vanderveken, and Vrins (2022) find that it is often preferable to combine the plug-in rule with $1/N$ rule instead, to obtain greater diversification. Consistent with their finding, the combination has another advantage in that it will converge to the optimal portfolio if $N$ is fixed and $T$ is large. In contrast, there is no guarantee that combining the GMV with the $1/N$ rule will converge except for some special cases.

DeMiguel, Garlappi and Uppal (2009) show the difficulties of estimated rules in beating the $1/N$ rule. Arguing for the value of investment theory, Tu and Zhou (2011) show that combining the $1/N$ rule with the plug-in rule can beat the $1/N$ for $T = 240$, but still not when $N = 120$, in factor models with $N = 25$ assets (their Table 2). Recently, Kan, Wang, and Zhou (2022), and Bodnar, Okhrin, and Parolya (2023), among others, consider alternative strategies that can improve upon the $1/N$ rule. Overall, the message is consistent with Table 3 that beating the $1/N$ rule is possible when $N$ is small and $T$ is large, but it is generally difficult. One way to mitigate this difficulty is to improve $\hat{w}_{g,\lambda}$ by using a sophisticated estimator of the covariance matrix instead of the sample covariance,[9] which is an interesting avenue for future research. But for this rule, or any other estimated rule, the estimation error can be large when the sample size is not large enough, and so it will still underperform the $1/N$ rule when the latter is not too far away from the true optimal portfolio, which depends on the choice of data sets.[10] Theoretically, Proposition 3 tells us that if the data set admits approximately a one-factor structure, the $1/N$ rule should be close to being optimal, making it almost impossible to beat by any estimated rule. Empirically, given that the $1/N$ performs as well as the S&P500 index as shown by Table 2, it appears difficult to beat the $1/N$ rule in general, since fund managers typically fail to beat the S&P500 index.

### B. The Case when $N > T$

When $N > T$, the sample covariance matrix is no longer invertible. Existing studies focus on obtaining a suitable invertile matrix under various sparsity assumptions, so that the plug-in rule and the like can still be used. For examples, Ledoit and Wolf (2003, 2017) impose a factor model first and later a more general condition. Chen and Yuan (2016) also impose a factor structure, while Bodnar, Okhrin, and Parolya (2021) use the pseudo-inverse. When $T/N < 1$ and $T$ is sufficiently large, these methods can potentially improve the $1/N$ rule, but the gain is typically small.

In this paper, we take a new direction. We consider how to beat the $1/N$ rule using *conditional*

[9] We thank an anonymous referee for pointing out this idea to us.

[10] Alternatively, it can also be the case in which both the $1/N$ rule and the estimated rule are far away from the true one, then the latter still cannot beat the $1/N$ rule if it does not converge to the true one fast enough.

*information* that goes beyond the usual unconditional set-up in which only the return data are used. Ferson and Siegel (2001) show how to construct an optimal portfolio from the conditional mean based on a signal. Lassance and Martín-Utrera (2022) propose a methodology that exploits investor sentiment. By contrast, we simply combine an alpha portfolio with the $1/N$ one. We use two alpha portfolios, respectively. The first is a portfolio of anomalies (e.g., Chen and Zimmermann, 2023) and the second is the long-short portfolio from ML studies (e.g., Gu, Kelly and Xiu, 2020).

### 1.. Adding anomalies when $N > T$

Since anomalies usually have alphas relative to the market, which is highly correlated with the $1/N$ portfolio, we expect they also have positive alphas relative to $1/N$ rule (adjusting the sign if necessary).[11] Let $R_{\hat{\alpha}_i}$ be the estimated alpha of an anomaly. Then

$$R_{\hat{\varphi}} = \frac{1}{N_\omega} \sum_{\hat{\alpha}_i \geq \hat{\alpha}_\omega} R_{\hat{\alpha}_i} \tag{35}$$

is likely to have a positive alpha if $\hat{\alpha}_\omega$ is the top $\omega\%$ of the $N$ alphas. For example, if $\omega = 5$, $R_{\hat{\varphi}}$ is simply an equal-weighted portfolio of those anomalies whose alphas are in the top 5% percentile. The reason that we ignore other anomalies is to minimize the estimation errors as the low alpha anomalies may have too small alphas to be statistically different from zeros.

To combine $R_{\hat{\varphi}}$ with the $1/N$, we can use the plug-in rule. Since this is a two asset case, the plug-in rule should work well as the estimation error now should be small. Theoretically, as long as the two assets are not perfectly correlated, the combined portfolio must outperform the $1/N$

[11] In the stock market, for example, the $1/N$ portfolio of the 500 stocks in the S&P500 has a correlation of 95% with the value-weighted S&P 500 index for the data period used for Table 2.

rule.

Table 4 reports the results with the use of 151 anomalies from the Chen-Zimmermann dataset. The time period is from 1973-01 to 2021-12. We use a typical 120 month rolling window to estimate the alphas. So the out-of-sample portfolio performance is from 1983-01 to 2021-12. The time starts from 1973-01 because this is the time when CRSP has more than 3,000 stocks, so that $N$ is large enough. We have 151 anomalies because these anomalies have full history from 1973-01 to 2021-12.

The (annualized) Sharpe ratio of the $1/N$ portfolio that invests in all stocks is 0.73. If we consider the portfolio of only the top 5% anomalies, $R_{\hat{\varphi}}$, the Sharpe ratio is 1.88, performing surprising well. The combination of the two has a Sharpe ratio of 1.97, improving the individual Sharpe ratios as expected. The gain is driven by the diversification benefit that reduces the risk to 2.81% per month. Consider now the top 10% alpha portfolios. The corresponding $R_{\hat{\varphi}}$ has smaller risk, yielding a greater Sharpe ratio than before. As a result, the combination performs slightly better than the top 5% case. Overall, combining the $1/N$ rule with the anomalies help the portfolio to outperform the $1/N$ rule substantially.

However, there is one caveat. The performance depends critically on the performance of the anomalies, which utilize information beyond the return data. If we are restricted to the use of the realized return data only, it is unclear how to outperform the $1/N$ rule when $N$ is large. Another caveat is that, as shown by McLean and Pontiff (2016), among others, the performance of the anomalies deteriorates over time. As a result, theoretically, there is no guarantee that the above combinations will perform as well in the future as they do now.

## 2.. Adding machine learning portfolios when $N > T$

There is a growing machine learning (ML) literature in finance, including Chinco, Clark-Joseph, and Ye (2019), Freyberger, Neuhierl, and Weber (2020), Gu, Kelly, and Xiu (2020), Kozak, Nagel, and Santosh (2020), Chen, Pelger, and Zhu (2023), and Han, He, Rapach and Zhou (2023).[12] Typically, these studies obtain long-short portfolios from the information of large sets of firm characteristics. In particular, Gu, Kelly, and Xiu (2020) show that gradient boosted regression trees, random forest, and neural networks provide some of the most profitable long-short portfolios. We extend their out-of-sample period from 1997-01 to 2020-12, and then combine the long-short portfolios with the $1/N$ rule where we estimate the weights based on 120 months of data recursively.

Table 5 reports the results. As Panel A shows, when all stocks are included, the $1/N$ rule has a Sharpe ratio of 0.52, but increases to 0.85 after combining with the long-short portfolio from the gradient boosted regression trees, and rises further to 1.04 from the random forest method. The best Sharpe ratios are obtained by combining the $1/N$ rule with those long-short portfolios from the neural networks. The one layer network has the lowest Sharpe ratio value of 1.05, and the three-layer one has the greatest of 1.26.

Panel B of Table 5 reports the results excluding microcap stocks, which is to address an important concern of Avramov, Cheng, and Metzker (2023), who find that ML methods tend to pick up small and illiquid stocks. Interestingly, the performance of the $1/N$ rule is only slightly weakened. Together with the weakened performance of the machine learning methods,

[12] Hutchinson, Lo, and Poggio (1994) and Rapach, Strauss, and Zhou (2013) are earlier studies that apply learning networks and the elastic net into finance, respectively.

the combinations perform worse than before. However, the Sharpe ratios decrease at most about 20%. Overall, the combinations outperform the $1/N$ rule substantially.[13]

## V. Conclusion

The modern portfolio theory pioneered by Markowitz (1952) is widely used in practice and extensively taught in schools. Yet, the estimated Markowitz portfolio rule and most of its extensions underperform the naive $1/N$ 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 $1/N$ rule is hard to beat.

First, as long as the dimensionality is high relative to sample size, we show that the usual estimated rules are not approach to the maximum Sharpe ratio even asymptotically due to estimation errors. Second, we show that the $1/N$ rule is optimal in terms of the Sharpe ratio in a one-factor model with diversifiable risks as dimensionality (number of assets) increases, irrespective of the sample size, making well-known investment theory-based rules inadequate as they suffer from estimation errors.

Third, we explore strategies that can outperform the $1/N$ rule *conditionally*, that is, conditional on the availability of profitable anomalies or machine learning (ML) portfolios. However, when $N$ is large, we caution that the outperformance over the $1/N$ rule is likely to decrease over time because the profitability of anomalies and the ML portfolios shows a decreasing pattern. While our study offers new insights on the $1/N$ rule, much work remains to be done. To possibly outperform the $1/N$ rule, our study indicates that a possible line of future research is to incorporate more

[13] The earlier caveats apply here. The performance of ML portfolios can deteriorate over time, and so there is no guarantee that the combinations will perform as well in the future.

conditional information into the traditional portfolio optimization framework.

## 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

$$\hat{\mu} = \mu + \Sigma^{1/2}\bar{X}, \qquad \text{and} \qquad \hat{\Sigma} = \Sigma^{1/2}S\Sigma^{1/2}, \tag{A.3}$$

where $\bar{X}$ and $S$ are the sample mean and covariance matrix of $X_t$s. Denoting by $\mu_* = \Sigma^{-1/2}\mu$, the standardized mean, we get

$$\hat{w}^\top\mu = \alpha\mu_*^\top S^{-1}\mu_* + \alpha\mu_*^\top S^{-1}\bar{X} + \lambda\mu_*^\top S^{-1}\Sigma^{-1/2}1 \tag{A.4}$$

and

$$\begin{aligned}\hat{w}^\top\Sigma\hat{w} &= \lambda^2 1^\top\Sigma^{-1/2}S^{-2}\Sigma^{-1/2}1 \\ &\quad +2\alpha\lambda\mu_*^\top S^{-2}\Sigma^{-1/2}1 + 2\alpha\lambda\bar{X}^\top S^{-2}\Sigma^{-1/2}1 \\ &\quad +\alpha^2\mu_*^\top S^{-2}\mu_* + \alpha^2\bar{X}^\top S^{-2}\bar{X} + 2\alpha^2\mu_*^\top S^{-2}\bar{X}.\end{aligned} \tag{A.5}$$

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

$$\begin{aligned}
\hat{w}^\top\Sigma\hat{w} &= (\alpha^2\mu^\top\Sigma^{-1}\mu + \lambda^2 1^\top\Sigma^{-1}1 + 2\alpha\lambda\mu^\top\Sigma^{-1}1)(e_1 W^{-2} e_1) + \frac{\alpha^2}{T}X^\top W^{-2}X \\
&\quad + \frac{2\alpha}{\sqrt{T}}(\alpha^2\mu^\top\Sigma^{-1}\mu + \lambda^2 1^\top\Sigma^{-1}1 + 2\alpha\lambda\mu^\top\Sigma^{-1}1)^{1/2} e_1^\top W^{-2}X
\end{aligned} \tag{A.6}$$

and

$$\begin{aligned}
\hat{w}^\top\mu &= (\alpha\mu^\top\Sigma^{-1}\mu + \lambda\mu^\top\Sigma^{-1}1)\cdot(e_1^\top W^{-1}e_1) \\
&\quad + \lambda\left((\mu^\top\Sigma^{-1}\mu)(1^\top\Sigma^{-1}1) - (\mu^\top\Sigma^{-1}1)^2\right)^{1/2}\cdot(e_2^\top W^{-1}e_1) \\
&\quad + \frac{\alpha}{\sqrt{T}}\frac{\alpha\mu^\top\Sigma^{-1}\mu + \lambda\mu^\top\Sigma^{-1}1}{(\alpha^2\mu^\top\Sigma^{-1}\mu + \lambda^2 1^\top\Sigma^{-1}1 + 2\alpha\lambda\mu^\top\Sigma^{-1}1)^{1/2}}\cdot(e_1^\top W^{-1}X) \\
&\quad + \frac{\alpha\lambda}{\sqrt{T}}\left(\frac{(\mu^\top\Sigma^{-1}\mu)(1^\top\Sigma^{-1}1) - (\mu^\top\Sigma^{-1}1)^2}{\alpha^2\mu^\top\Sigma^{-1}\mu + \lambda^2 1^\top\Sigma^{-1}1 + 2\alpha\lambda\mu^\top\Sigma^{-1}1}\right)^{1/2}\cdot(e_2^\top W^{-1}X).
\end{aligned} \tag{A.7}$$

This proves the exact distribution. Note that the limiting distributions in Corollaries 1 and 2 follow from the next proof. $\square$

*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

$$e_1^\top W^{-1} e_2 = O_p(T^{-1/2}). \tag{A.8}$$

Similarly, we have

$$e_1^\top W^{-1} e_1 = \frac{1}{1-\eta} + O_p(T^{-1/2}), \tag{A.9}$$

and

$$e_1^\top W^{-2} e_1 = \frac{1}{(1-\eta)^3} + O_p(T^{-1/2}). \tag{A.10}$$

In light of (A.10), with probability tending to 1, we have

$$e_1^\top W^{-2} e_1 \le \frac{2}{(1-\eta)^3}. \tag{A.11}$$

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 we get

$$\frac{e_1^\top W^{-1} X}{(e_1^\top W^{-2} e_1)^{1/2}} = O_p(1). \tag{A.12}$$

Together, we have

$$e_1^\top W^{-1} X = O_p(1). \tag{A.13}$$

Similarly, we have

$$e_1^\top W^{-2} X = O_p(1), \qquad \text{and} \qquad e_2^\top W^{-1} X = O_p(1). \tag{A.14}$$

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.8) and (A.9), 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

$$\text{(A.15)} \qquad \mu^\top 1/N = \bar{\beta}\mu_q, \qquad \text{and} \qquad 1^\top\Sigma 1/N^2 = \sigma_q^2\bar{\beta}^2 + 1^\top(\Sigma_\varepsilon)1/N^2.$$

In particular, if $\beta_0 > 0$ and the betas has a finite variance $\sigma_\beta^2$, 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

$$\text{(A.16)} \qquad SR_{1/N} = \frac{\mu_q}{\sigma_q} + O(N^{-1/2}).$$

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 $\bar{X} \sim N(0, I/T)$ such that

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

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

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

where $\mu_* = \Sigma^{-1/2}\mu$ as before. Now let $U$ be an orthonormal matrix whose first column is $\mu_*/\|\mu_*\|$, second column is $a/\|a\|$ where

$$a = \Sigma^{1/2} 1 - \frac{\mu^\top 1}{\mu^\top \Sigma^{-1} \mu} \cdot \mu_*, \tag{A.17}$$

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

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

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

$$\mu^\top \hat{w} = \lambda SR(e_1^\top W^{-1} \theta_2) + (1-\lambda)\mu_{1/N}, \tag{A.18}$$

and

$$\hat{w}^\top \Sigma \hat{w} = \lambda^2 \theta_2^\top W^{-2} \theta_2 + (1-\lambda)^2 \sigma_{1/N}^2 + 2\lambda(1-\lambda)\theta_1^\top W^{-1} \theta_2 / N, \tag{A.19}$$

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

$$X^\top W^{-2} X = \frac{\eta}{(1-\eta)^3} + O_p(T^{-1/2}), \tag{A.20}$$

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 $SR_\lambda$ is also the Sharpe ratio of

$$\hat{w} = \hat{\Sigma}^{-1}\hat{\mu} + \omega 1, \tag{A.21}$$

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

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

Then, we have

$$\hat{w}^\top \mu = \mu_*^\top S^{-1}\mu_* + \mu_*^\top S^{-1}\bar{X} + \omega\mu^\top 1 \tag{A.22}$$

and

$$\begin{aligned} \hat{w}^\top \Sigma \hat{w} &= \omega^2 1^\top \Sigma 1 + 2\omega\mu_*^\top S^{-1}\Sigma^{1/2}1 + 2\omega\bar{X}^\top S^{-1}\Sigma^{1/2}1 \\ &\quad + \mu_*^\top S^{-2}\mu_* + \bar{X}^\top S^{-2}\bar{X} + 2\mu_*^\top S^{-2}\bar{X}, \end{aligned} \tag{A.23}$$

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

$$a = \Sigma^{1/2}1 - \frac{\mu^\top 1}{\mu^\top \Sigma^{-1}\mu} \cdot \mu_*. \tag{A.24}$$

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

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

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

$$\delta = \frac{SR_{1/N}}{SR} = \frac{\mu^\top 1}{(\mu^\top \Sigma^{-1} \mu)^{1/2} (1^\top \Sigma 1)^{1/2}}. \tag{A.25}$$

□

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

## References


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

Avramov, D.;, S. Cheng; L. Metzker; and S. Voigt. "Machine learning versus economic restrictions: evidence from stock return predictability." *Management Science*, 69 (2023), 2547–3155.

Barillas, F.; R. Kan; C. Robotti; and J. Shanken. "Model comparison with Sharpe ratios." *Journal of Financial and Quantitative Analysis*, 55 (2020), 1840–1874.

Barillas, F.; and J. Shanken. "Which alpha?" *Review of Financial Studies*, 30 (2017), 1316–1338.

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

Bawa, V.; S. Brown; and R. Klein. *Estimation Risk and Optimal Portfolio Choice* (1979), Amsterdam: North-Holland.

Berk, B.; and J. van Binsbergen. "Assessing asset pricing models using revealed preference." *Journal of Financial Economics*, 119 (2016), 1–23

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

Bodnar, T.; Y. Okhrin; and N. Parolya. "Optimal shrinkage-based portfolio selection in high dimensions." *Journal of Business and Economic Statistics* 41 (2023), 140–156.

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

Chen, L.; M. Pelger; and J. Zhu. "Deep learning in asset pricing." *Management Science*, forthcoming (2023).

Chen, J.; and M. Yuan. "Efficient portfolio selection in a large market." *Journal of Financial Econometrics*, 14 (2016), 496–524.

Chen, A.; and T. Zimmermann. "Open source cross-sectional asset pricing." *Critical Finance Review*, forthcoming (2023).

Chincarini, L.; and D. Kim. *Quantitative Equity Portfolio Management* (2023), 2ed., New York, NY: McGraw Hill.

Chinco, A.; A. Clark-Joseph; and M. Ye. "Sparse signals in the cross-section of returns." *Journal of Finance*, 74 (2019), 449–492.

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

DeMiguel, V.; L. Garlappi; F. Nogales; and R. Uppal. "A generalized approach to portfolio optimization: improving performance by constraining portfolio norms." *Management Science*, 55 (2009), 798–812.

DeMiguel, V.; A. Martín-Utrera; and F. Nogales. "Parameter uncertainty in multiperiod portfolio optimization with transaction costs." *Journal of Financial and Quantitative Analysis*, 50 (2015), 1443–1471.

DeMiguel, V.; A. Martín-Utrera; F. Nogales; and R. Uppal. "A transaction-cost perspective on the multitude of firm characteristics." *Review of Financial Studies*, 33 (2020), 2180–2222.

Duchin, R.; and H. Levy. "Markowitz versus the Talmudic portfolio diversification strategies." *Journal of Portfolio Management*, 35 (2009), 71–74.

Fama, E.; and K. French. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics*, 33 (1993), 3–56.

Ferson, W.; and A. Siegel. "The efficient use of conditioning information in portfolios." *Journal of Finance*, 56 (2001), 967–982.

Frahm, G.; and C. Memmel. "Dominating estimators for minimum-variance portfolios." *Journal of Econometrics*, 159 (2010), 289–302.

Freyberger, J.; A. Neuhierl; and M. Weber. "Dissecting characteristics nonparametrically." *Review of Financial Studies*, 33 (2020), 2326–2377.

Gu, S.; B. Kelly; and D. Xiu. "Empirical asset pricing via machine learning." *Review of Financial Studies*, 33 (2020), 2223–227.

Grinold, R.; and R. Kahn. *Active Portfolio Management: Quantitative Theory and Applications* (1999), New York, NY: McGraw Hill.

Hafner, C.; and L. Wang. "Dynamic portfolio selection with sector-specific regularization." *Econometrics and Statistics*, forthcoming (2023).

Han, Y.; A. He; D. Rapach; and G. Zhou. "Cross-sectional expected returns: New Fama-MacBeth regressions in the era of machine learning" Working Paper, Washington University in St. Louis (2023).

Harvey, C. R.; and Liu, Y. "Lucky factors." *Journal of Financial Economics*, 141 (2021), 413–435.

He, A.; and G. Zhou. "Diagnostics of asset pricing models." *Financial Management*, forthcoming (2023).

Huang, C.; and R. Litzenberger. *Foundations for Financial Economics* (1988), Amsterdam: North-Holland.

Hutchinson, J.; A. Lo; and T. Poggio. "A nonparametric approach to pricing and hedging derivative securities via learning networks." *Journal of Finance*, 49 (1994), 851–889.

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

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

Kan, R.; X. Wang; and X. Zheng. "In-sample and out-of-sample Sharpe ratios of multi-factor asset pricing models." Working Paper, University of Toronto (2020).

Kan, R.; X. Wang; and G. Zhou. "Optimal portfolio choice with estimation risk: No risk-free asset case." *Management Science*, 22 (2022) 2047–2068.

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

Kempf, A.; and C. Memmel. "Estimating the global minimum variance portfolio." *Schmalenbach Business Review*, 58 (2006), 332–348.

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

Kozak, S.; S. Nagel; and S. Santosh. "Shrinking the cross section." *Journal of Financial Economics*, 135 (2020), 271–292.

Lassance, N. "Maximizing the out-of-sample Sharpe ratio." Working Paper, UCLouvain (2021).

Lassance, N.; and A. Martin-Utrera. "Shrinking against sentiment: Exploiting latent asset demand in portfolio selection." Working Paper, UCLouvain (2022).

Lassance, N.; A. Martín-Utrera.; and M. Simaan. "The risk of expected utility under parameter uncertainty." *Management Science*, forthcoming (2023).

Lassance, N.; R. Vanderveken. and F. Vrins. "On the optimal combination of naive and mean-variance portfolio strategies." Working Paper, UCLouvain (2022).

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

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

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

Markowitz, H. "Portfolio selection." *Journal of Finance*, 7 (1952), 77–91.

McLean, R.; and J. Pontiff. "Does academic research destroy return predictability." *Journal of Finance*, 71 (2016), 5–32.

Meucci, A. *Risk and Asset Allocation* (2005), New York, NY: Springer-Verlag.

Pflug, G.; A. Pichler; and D. Wozabal. "The $1/N$ investment strategy is optimal under high model ambiguity." *Journal of Banking and Finance*, 36 (2012), 410–7.

Rapach, D.; J. Strauss; and G. Zhou. "International stock return predictability: What is the role of the United States?" *Journal of Finance*, 68 (2013), 1633–1662.

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

Raponi V.; R. Uppal; and P. Zaffaroni. "Robust Portfolio Choice" Working Paper, Imperial College (2021).

Ross, S. "The arbitrage theory of capital asset pricing." *Journal of Economic Theory*, 13 (1976), 341–360.

Shi, F.; L., Shu; A., Yang; and F., He. "Improving minimum-variance portfolios by alleviating overdispersion of eigenvalues." *Journal of Financial and Quantitative Analysis*, 55 (2020), 2700–2731.

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

Yan, C.; and H. Zhang. "Mean-variance versus naïve diversification: The role of mispricing" *Journal of International Financial Markets, Institutions and Money*, 48 (2017), 61–81.

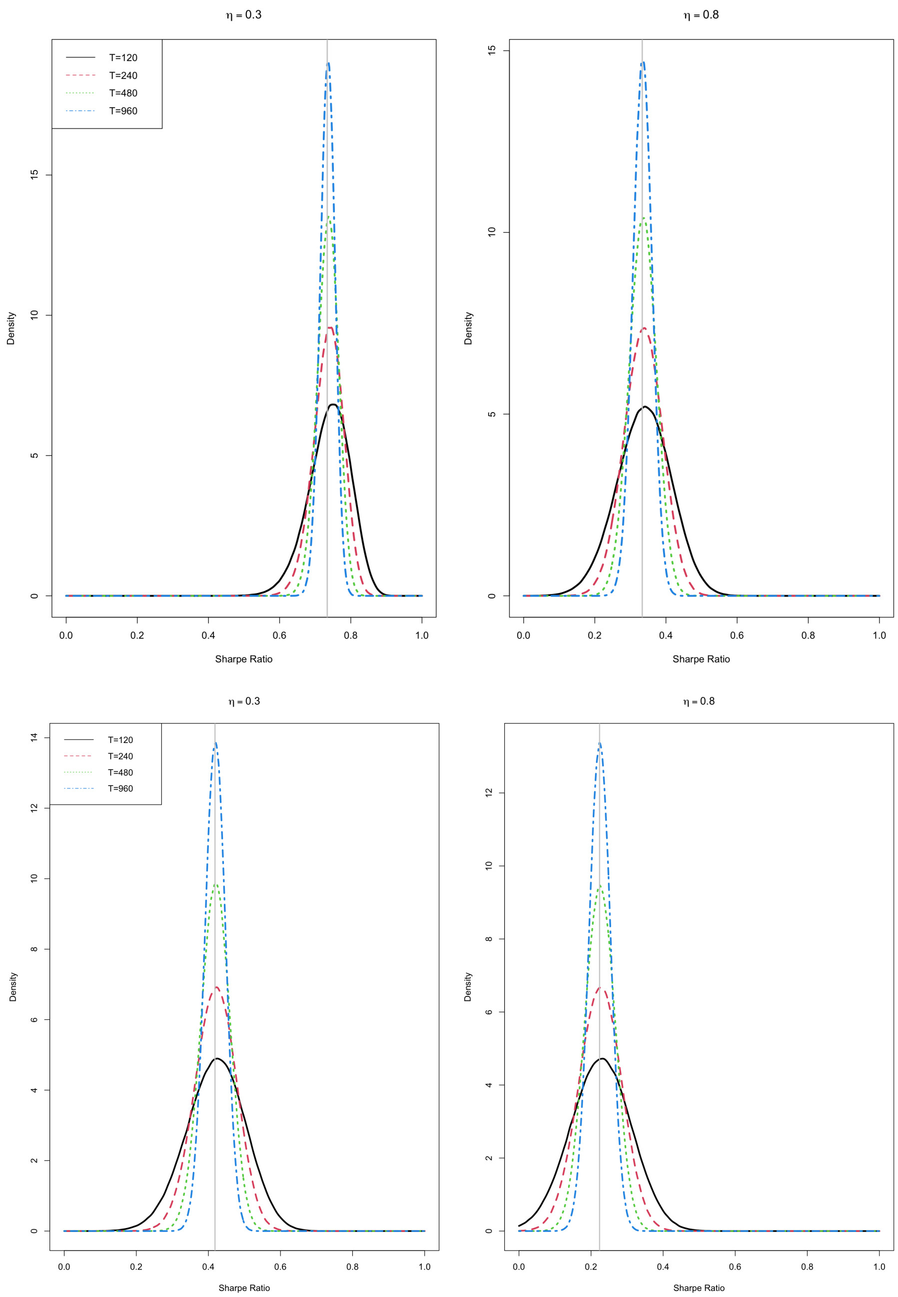


**Figure 1.** Densities of $SR_{\hat{w}}$ and $SR_{\hat{g}}$ estimated from 1000000 simulations for different values of $\eta$ and $T$. In each panel, the vertical grey line corresponds to the limiting value of $SR_{\hat{w}}$ and $SR_{\hat{g}}$ as $T \to \infty$.

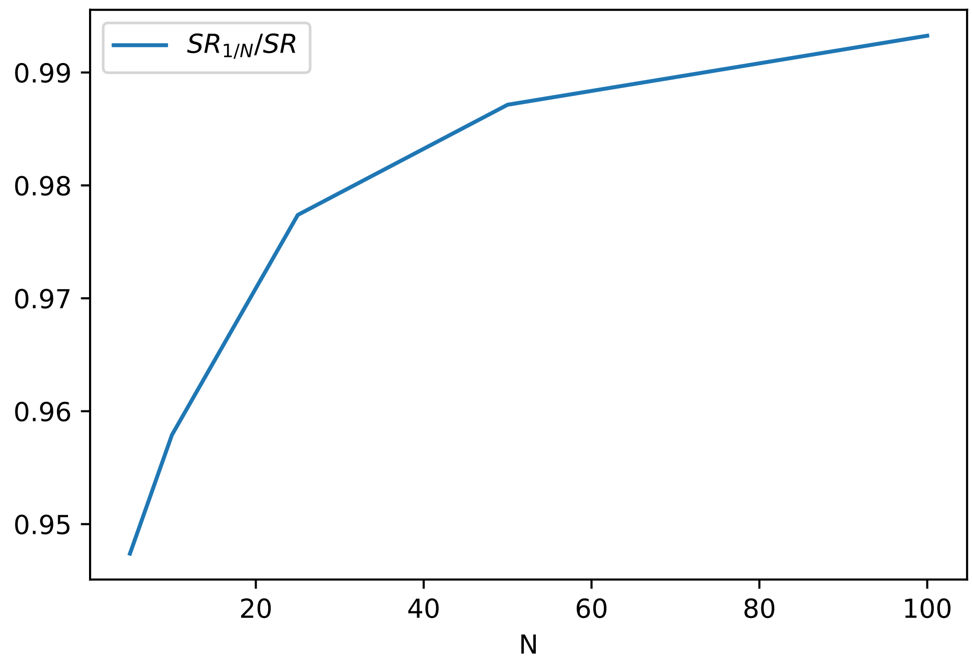


**Figure 2.** Comparison of Sharpe ratios of $1/N$ with the Optimal

This figure plots the ratio of the Sharpe ratio of the $1/N$ investment strategy to the Sharpe ratio of the mean-variance optimal portfolio. The population parameters are from a one-factor model with the number of assets $N$ increases from 5 to 100.

## TABLE 1

### $1/N$ Portfolio in a one-factor model

The first row of the table reports the annualized Sharpe ratio of the $1/N$ portfolio when the asset returns follow a one-factor model,

$$R_i = \beta_i R_q + \varepsilon_i, \quad i = 1, 2, \ldots, N,$$

when $N$, the number of assets, equals to $5, 10, 25, 50$ and 100, respectively. The beta parameters are randomly drawn from a normal distributions,

$$\beta_i \sim N(1, \sigma_\beta^2), \qquad i = 1, 2, \ldots, N,$$

with $\sigma_\beta$ calibrated from the 25 size and book-market portfolios of Fama and French (1993) with monthly data from 1926-7 to 2022-04 relative to the stock market index. The $\sigma_{ij}$'s, parameters of the residual covariance matrix, are drawn similarly. The number of simulations is 10,000. The second row of the table reports the results when just the beta volatility $\sigma_\beta$ is doubled, and the third row reports the results when just the volatility of the $\sigma_{ij}$'s is doubled. The true Sharp ratio, True in the table, is scaled to be 0.5.

| | 5 | 10 | 25 | 50 | 100 | *True* |
|---|---|---|---|---|---|---|
| Calibrated model | 0.4073 | 0.4418 | 0.4750 | 0.4889 | 0.4967 | 0.5 |
| Beta vol doubled | 0.3835 | 0.4286 | 0.4698 | 0.4850 | 0.4932 | 0.5 |
| Sigma vol doubled | 0.3406 | 0.3855 | 0.4444 | 0.4752 | 0.4948 | 0.5 |

## TABLE 2

### $1/N$ Portfolio of Random Stocks in the S&P500

Panel A of the table reports the monthly mean, standard deviation (both in percentage points) and annualized Sharpe ratio of the equal-weighted portfolio of $N$ stocks, with $N = 5, 10, 25, 50, 100, 250$ and 500, respectively. The stocks are selected randomly from the S&P500 index and are replaced randomly if any of them are removed from the index. The data is monthly, from 1957–03 to 2021–12, and there are 778 months in total. The last column, VW, reports the same performance measures for the value-weighted S&P500 index. Panels B and C report the same results except that the random stocks were chosen 10, and 100 times, respectively.

| | 5 | 10 | 25 | 50 | 100 | 250 | 500 | VW |
|---|---|---|---|---|---|---|---|---|
| | | | | *Panel A: 1 set of Random Stocks* | | | | |
| mean | 1.02 | 1.12 | 1.10 | 1.10 | 1.10 | 1.14 | 1.13 | 0.95 |
| std | 5.27 | 4.93 | 4.71 | 4.70 | 4.61 | 4.60 | 4.86 | 4.22 |
| Sharpe ratio | 0.44 | 0.54 | 0.55 | 0.55 | 0.56 | 0.59 | 0.55 | 0.49 |
| | | | | *Panel B: 10 Sets of Random Stocks* | | | | |
| mean | 1.16 | 1.12 | 1.12 | 1.11 | 1.12 | 1.13 | 1.13 | 0.95 |
| std | 5.76 | 5.05 | 4.86 | 4.7 | 4.64 | 4.67 | 4.86 | 4.22 |
| Sharpe ratio | 0.49 | 0.53 | 0.55 | 0.56 | 0.58 | 0.58 | 0.55 | 0.49 |
| | | | | *Panel C: 100 Set of Random Stocks* | | | | |
| mean | 1.10 | 1.11 | 1.12 | 1.12 | 1.12 | 1.12 | 1.13 | 0.95 |
| std | 5.73 | 5.19 | 4.83 | 4.71 | 4.63 | 4.66 | 4.86 | 4.22 |
| Sharpe ratio | 0.46 | 0.51 | 0.55 | 0.56 | 0.57 | 0.58 | 0.55 | 0.49 |

**TABLE 3**

**Combined Rules and Comparison**

This table reports the annualized Sharpe ratios of the estimated optimal combination of the GMV and Plug-in rule with the $1/N$ rule, and those of these rules individually. The asset returns are the $N$ size portfolios of all the stocks. The data is monthly, and the rolling estimation window is fixed at 360 months. The data starts from 1973-01, and the out-of-sample period is from 2003-01 to 2022-03.

| $N$ | GMV + 1/N | Plug-in + 1/N | 1/N | GMV | Plug-in |
|---|---|---|---|---|---|
| 5 | 0.73 | 0.61 | 0.54 | 0.73 | 0.72 |
| 10 | 0.84 | 0.67 | 0.60 | 0.84 | 0.72 |
| 20 | 0.92 | 0.77 | 0.72 | 0.81 | 0.77 |
| 50 | 0.71 | 0.55 | 0.68 | 0.72 | 0.39 |
| 100 | 0.73 | 0.66 | 0.73 | 0.70 | 0.44 |

## TABLE 4

### Combining with Anomalies

This table reports the monthly mean, standard deviation (both in percentage points) and annualized Sharpe ratio of the $1/N$ portfolio of all stocks with prices greater than $5, the equal-weighted portfolio of anomaly alpha portfolios $R_\omega$ and their equal combinations, where $\omega$ equal 5% or 10%, respectively, that is, of only those top 5% or 10% alpha anomalies. The data is monthly, and the rolling estimation window is fixed at 120 months. The out-of-sample period is from 1983-01 to 2021-12.

| | *Panel A: top 5%* | | | *Panel B: top 10%* | | |
|---|---|---|---|---|---|---|
| | 1/N | $R_\omega$ | 1/N+$R_\omega$ | 1/N | $R_\omega$ | 1/N+$R_\omega$ |
| mean | 1.08 | 1.70 | 1.59 | 1.08 | 1.59 | 1.52 |
| std | 5.10 | 3.14 | 2.81 | 5.10 | 2.64 | 2.49 |
| Sharpe Ratio | 0.73 | 1.88 | 1.97 | 0.73 | 2.08 | 2.12 |

**TABLE 5**
**Combining with Machine Learning Portfolios**

The table reports the monthly mean, standard deviation (both in percentage points) and annualized Sharpe ratio of the $1/N$ portfolio of all stocks (with prices greater than \$5), its optimal combination with a machine learning portfolio, which is computed by using gradient boosted regression trees (GBR), random forest (RF) and neural networks (NN, with 1 to 5 layers), respectively. The data starts from 1957-01, and the out-of-sample period is from 1997-01 to 2020-12.

| | $\frac{1}{N}$ | $\frac{1}{N}$+GBR | $\frac{1}{N}$+RF | $\frac{1}{N}$+NN1 | $\frac{1}{N}$+NN2 | $\frac{1}{N}$+NN3 | $\frac{1}{N}$+NN4 | $\frac{1}{N}$+NN5 |
|---|---|---|---|---|---|---|---|---|
| | | | | *Panel A: All Stocks* | | | | |
| mean | 1.04 | 1.42 | 1.40 | 2.88 | 4.06 | 4.27 | 4.8 | 4.24 |
| std | 5.82 | 5.78 | 4.68 | 9.47 | 12.11 | 11.75 | 13.28 | 12.24 |
| Sharpe ratio | 0.52 | 0.85 | 1.04 | 1.05 | 1.16 | 1.26 | 1.25 | 1.20 |
| | | | | *Panel B: All but Microcaps* | | | | |
| mean | 0.96 | 1.26 | 1.40 | 2.09 | 2.25 | 2.88 | 2.66 | 2.66 |
| std | 5.43 | 5.38 | 5.00 | 7.74 | 8.04 | 9.25 | 8.69 | 9.31 |
| Sharpe ratio | 0.51 | 0.81 | 0.97 | 0.93 | 0.97 | 1.08 | 1.06 | 0.99 |