Sample selection bias, return moments, and the performance of optimal versus naive diversification

Bowei Li*

15 August 2016

ABSTRACT

I examine the sample selection bias in portfolio horse race. Numerous studies propose mean-variance portfolio rules to outperform the naive 1/N portfolio rule. However, the outperformance is often justified by a small number of pre-selected datasets. Using a new performance test based on a large number of datasets, I compare thirteen "1/N outperformers" with the naive rule. Results show that not only a majority of the thirteen "1/N outperformers" on average underperform the 1/N rule significantly, but also none of these mean-variance rules significantly outperform the naive benchmark in more than 10% of the datasets. To improve mean-variance portfolio performance, I propose a switching strategy that selects between a mean-variance rule and the 1/N rule using predictability of portfolio performance based on assets' return moments. The new strategy outperforms not only its raw mean-variance rule but also the naive rule significantly out-of-sample.

JEL classification: G11, G14, G17.

*Bowei Li is a Ph.D. Candidate at Department of Finance, University of Melbourne, Australia. Email: boweil@student.unimelb.edu.au. I am greatly indebted to my principal supervisor, Joachim Inkmann, and my co-supervisors, Federico Nardari and Juan Sotes-Paladino for their extensive help and support. I also thank the participants at the University of Melbourne Finance Brown Bag Seminar.

I. Introduction

The mean-variance portfolio theory pioneered by Markowitz (1952, 1959) is widely applied in practice. However, DeMiguel, Garlappi, and Uppal (2009) question the value of the Markowitz theory relative to naive diversification, by comparing various mean-variance portfolio rules with the naive 1/N portfolio rule. They find that none of the sophisticated strategies can outperform the 1/N rule out-of-sample consistently because diversification benefit is more than offset by estimation error. This surprising finding has led to a trend of studies that attempt to develop better mean-variance strategies to challenge the naive rule[1].

Many of the newly proposed mean-variance strategies seem to do a good job outperforming the 1/N rule, an observation comforting and suspicious – comforting because it reaffirms the usefulness of the Markowitz theory, but suspicious because it may result from the datasets chosen. Existing studies involving portfolio horse races with the 1/N strategy invariably use a small number of pre-selected datasets of stock portfolios, primarily characteristic-based portfolios. Although empirical results suggest that these strategies are able to outperform the naive 1/N rule, the outperformance is likely attributed to the datasets chosen, and thus, create a sample selection bias in these studies. Motivated by the suspicion above, this paper examines whether the outperformance of numerous Markowitz rules involves such a selection bias.

To address the sample selection problem, I propose a new performance test based on a large number of datasets. Existing performance tests for portfolio horse races, such as the methods by Jobson and Korkie (1981) and Kirby and Ostdiek (2012), examine portfolio performance in a single dataset. To avoid the selection bias of a small number of datasets, I record portfolio performance in each of many datasets, and summarize the aggregate performance for each portfolio strategy. The new test provides statistical inference on aggregate portfolio performance while taking into account performance uncertainty in each dataset.

To examine the sample selection bias empirically, I compare the performance of the 1/N rule with 13 mean-variance rules that are claimed to outperform the naive benchmark under different empirical settings. First, these optimal strategies are tested against the naive strategy by using a typical characteristic-based portfolio dataset, the 25 size/book-to-market portfolio dataset of Fama

[1] for instance DeMiguel, Garlappi, Nogales, and Uppal (2009.), Tu and Zhou (2011), Kirby and Ostdiek (2012, 2013), Behr, Guettler, and Miebs (2013), Anderson and Cheng (2016), and Kan, Wang, and Zhou (2016)

and French (1993). The Fama-French dataset finds that all of the 13 "1/N outperformers" outperform the 1/N portfolio in terms of Sharpe ratio and certainty-equivalent (CEQ) return. However, a performance comparison based on a large number of datasets shows the opposite result. Each portfolio strategy is tested repeatedly across a large number of datasets that are randomly formed based on NYSE/AMEX/NASDAQ stocks on CRSP. The empirical comparison based on 6000 sampled datasets finds that none of the 13 "1/N outperformers" investigated is able to consistently outperform the 1/N benchmark in terms of Sharpe ratio or CEQ return. These Markowitz rules significantly outperform the 1/N portfolio in no more than 10% of the datasets. More surprisingly, when the aggregate performance is considered, a majority of these mean-variance strategies, in fact, underperform the naive benchmark in both statistically and economically significant manner. The sample selection bias is robust in several aspects, including different risk aversion, portfolio constraints, expanding versus rolling estimation window, as well as alternative sampling schemes of the datasets.

To understand the driving force in the performance of optimal versus naive diversification, I theoretically examine the relation between the mean-variance utility loss of the 1/N rule and the return moments of assets used for portfolio construction. First, the mean-variance loss of the 1/N rule with respect to the optimal Markowitz portfolio is an increasing function in the average and dispersion of expected returns (expected return effect). Second, the loss is a decreasing function in the average and dispersion of volatilities (volatility effect). Third, the loss of the naive portfolio tends to be large when the average and dispersion of correlations are either very large (heterogeneity effect) or very small (diversification effect). These mean-variance properties provide guidance for investors when selecting between optimal versus naive diversification.

To study the practical use of these mean-variance properties, I investigate the predictability of the performance of Markowitz rules with respect to the naive rule based on the in-sample return moments of the assets. Results show that the forecast ability of return moments is consistent with the theoretical prediction of portfolio performance. First, larger average expected return and wider expected return dispersion collectively predict better performance of a mean-variance strategy with respect to the naive 1/N strategy (expected return effect). Second, smaller average volatility and narrower volatility dispersion collectively predict better performance of a mean-variance strategy with respect to the naive rule (volatility effect). Third, the heterogeneity effect dominates the diversification effect in that larger average correlation and wider correlation dispersion collectively predict better performance of a mean-variance rule relative to the naive rule. More importantly, this predictability also delivers out-of-sample portfolio benefits. Specifically, I propose a portfolio switching strategy, which selects between a mean-variance rule and the 1/N rule based on the forecast from a predictive model. Results based on the 6000 sampled datasets show that, for all the 13 Markowitz rules examined, the switching strategy not only significantly outperforms its raw mean-variance counterpart, but also outperforms the 1/N rule in a statistically significant manner.

This paper contributes to the literature on sample selection bias in finance studies by documenting a selection bias in a novel context - portfolio horse race. Existing studies on sample selection bias are centered around three aspects, namely sample/database bias (Kothari, Shanken, and Sloan, 1995, Chan, Jegadeesh, and Lakonishok, 1995 and Kim, 1997) , delisting bias (Shumway, 1997 and Shumway and Warther, 1999) and survivorship bias (Carhart, Carpenter, Lynch, and Musto, 2002). Although the selection bias is well discussed in many areas of finance, rare studies examine this issue in a portfolio horse race. This paper fills the gap in the literature on sample selection issues by showing a unique selection bias in the datasets used by a number of mean-variance strategies to justify the outperformance over the 1/N benchmark.

This study also contributes to the recent debate between optimal mean-variance diversification and naive diversification. On one hand, theoretical and empirical studies in favor of naive diversification suggest that the naive rule has several advantages and attributes, including the absence of estimation error (DeMiguel et al., 2009) and model ambiguity (Pflug, Pichler, and Wozabal, 2012), and good empirical performance as a compensation for higher tail risk (Brown, Hwang, and In, 2013). On the other hand, advocates of mean-variance portfolio theory attempt to improve mean-variance strategies to challenge the naive rule via different channels, such as portfolio combination (Tu and Zhou, 2011), reducing portfolio turnover (Kirby and Ostdiek, 2012), and mitigating estimation risk (Kan et al., 2016). I find that optimal mean-variance rules rarely defeat the naive 1/N rule in a significant way. However, the Markowitz theory is still useful in appropriate investment scenarios. All in all, the debate between optimal versus naive diversification depends on the return moments of assets used for portfolio construction.

The rest of the paper is structured as follows. In Section II, I theoretically show how return moments affect the performance of naive diversification, and discuss how they are related to the sample selection bias. Section III summarizes the data and methodologies used in the empirical study. Empirical results are presented in Section IV. Section V studies the predictability of portfolio performance based on assets' return moments, and Section VI concludes.

II. The Mean-Variance Properties of Naive Diversification

To illustrate how this sample selection bias is affected by the return moments of assets for portfolio construction, I theoretically show that the performance of naive diversification is a function in a number of mean-variance (or risk-return) characteristics of constituent assets. Next, I use a numerical example to confirm the relation between these return moments and the performance of the naive 1/N rule. Finally, I discuss how these mean-variance properties of naive diversification are related to the sample selection bias in a number of studies involving a portfolio horse race with the 1/N rule.

A. A Theoretical Framework

To formally examine how return moments affect performance of naive diversification relative to the optimal mean-variance portfolio, I study the properties of the 1/N rule in the mean-variance setting. In other words, how does the relative performance of the 1/N rule with respect to the optimal mean-variance portfolio varies with major return moments of assets, such as mean return, volatility, correlation and risk aversion.

I start from the standard mean-variance setup, where an investor chooses a vector of portfolio weights in order to maximize the following mean-variance utility:

$$\max_{x} \quad U(x) = x^T \mu - \frac{\gamma}{2} x^T \Sigma x \tag{1}$$

The optimal mean-variance portfolio weight is

$$x^* = \frac{1}{\gamma} \Sigma^{-1} \mu \tag{2}$$

The optimal mean-variance portfolio generates a utility of $\frac{1}{2\gamma}\mu^T \Sigma^{-1} \mu = \frac{1}{2\gamma} S_*^2$, where S_* is the Sharpe ratio of the portfolio. Any portfolio that deviates from this optimal portfolio suffers a mean-variance utility loss. Following closely the literature (Frost and Savarino (1986),Stambaugh (1997) and DeMiguel et al. (2009)), I define the expected utility loss of a particular portfolio $\hat{x}$ as:

$$L(x^*, \hat{x}) = U(x^*) - E[U(\hat{x})] = \frac{1}{2\gamma}\mu^T\Sigma^{-1}\mu - E[U(\hat{x})] \tag{3}$$

According to DeMiguel et al. (2009),the 1/N rule is $x^{ew} = \frac{c}{N}\mathbf{1}_N$, where $\mathbf{1}_N$ is a $N \times 1$ vector of ones and c is the fraction of wealth invested in the risky assets. The rest $1 - c$ is the fraction invested in the risk-less asset. The loss function of using such a 1/N rule is:

$$L(x^*, x^{ew}) = U(x^*) - \frac{c}{N}\mathbf{1}_N^T\mu + \frac{\gamma}{2}\frac{c^2}{N^2}\mathbf{1}_N^T\Sigma\mathbf{1}_N \tag{4}$$

The lowest bound of the loss is obtained by choosing $c^* = \frac{N\mathbf{1}_N^T\mu}{\gamma\mathbf{1}_N^T\Sigma\mathbf{1}_N}$, so the 1/N rule weight vector is:

$$x^{ew} = \frac{\mathbf{1}_N^T\mu\mathbf{1}_N}{\gamma\mathbf{1}_N^T\Sigma\mathbf{1}_N} \tag{5}$$

Therefore, the lowest bound for the loss function of the 1/N rule is:

$$L(x^*, x^{ew}) = \frac{1}{2\gamma}\left(\mu^T\Sigma^{-1}\mu - \frac{(\mathbf{1}_N^T\mu)^2}{\mathbf{1}_N^T\Sigma\mathbf{1}_N}\right) = \frac{1}{2\gamma}\left(S_*^2 - S_{ew}^2\right) \tag{6}$$

where S_* is the Sharpe ratio of the optimal mean-variance portfolio and S_{ew} is the Sharpe ratio of the 1/N portfolio.

Based on this loss function of the 1/N rule, the following four propositions summarize the performance of the 1/N according to changes in mean returns, volatilities, correlations and risk aversion, respectively.

Proposition 1. *Denote the mean vector for asset returns as $\mu = \phi\tilde{\mu}$ ($\phi > 0$), where $\tilde{\mu}$ is a fixed benchmark mean vector. For the loss function of the 1/N rule $L(x^*, x^{ew})$, we have:*

(1) $\frac{\partial L(x^*, x^{ew})}{\partial \phi} > 0$;

(2) $\frac{\partial^2 L(x^*, x^{ew})}{\partial \phi^2} > 0$;

Economic Intuition behind Proposition 1.

Proposition 1 uses a scalar ϕ to control the average and the dispersion of expected returns for the constituent assets used for portfolio construction. The proposition suggests that the mean-variance utility loss of the 1/N rule is an increasing function in the average and dispersion of expected returns of the constituent assets used to form portfolios.

Holding the covariance matrix unchanged, as the expected returns of assets increase proportionally (both average and dispersion of expected returns increase), the assets with initially high (low) expected returns will have a even higher (lower) expected returns. In this situation, a mean-variance investor tends to allocate more (less) weights to assets with higher (lower) expected return to improve the utility. However, the 1/N rule still allocates funds equally across assets, suffering more loss for not considering the proportional return change. $\frac{\partial^2 L(x^*,x^{ew})}{\partial \phi^2} > 0$ means this return effect magnifies with the increase of average expected return and return dispersion. The proposition shows that small average and dispersion of expected returns are in favor of the 1/N rule, while the naive rule tends to have higher loss in an environment with larger average and dispersion of expected returns.

Proposition 2. *Denote the variance-covariance matrix of asset returns as* $\Sigma = \lambda^2 \tilde{\Sigma}$ *(*$\lambda > 0$*), where* $\tilde{\Sigma}$ *is a fixed benchmark variance-covariance matrix. For the loss function of the 1/N rule* $L(x^*, x^{ew})$*, we have:*

(1) $\frac{\partial L(x^*,x^{ew})}{\partial \lambda} < 0$*;*

(2) $\frac{\partial^2 L(x^*,x^{ew})}{\partial \lambda^2} > 0$*.*

Economic Intuition behind Proposition 2.

By using a parameter λ to control the average and the dispersion of asset volatilities, the proposition suggests: if the volatilities change proportionally, the mean-variance utility loss of the 1/N rule is a decreasing function in the average and dispersion of volatilities. This is because when the average and dispersion of volatilities are small (large), a mean-variance investor relies more (less) on the return difference between assets, and tends to construct an optimal portfolio with more (less) extreme portfolio weights. The equally-weighted portfolio will then deviate more (less) from the optimal mean-variance portfolio. $\frac{\partial^2 L(x^*,x^{ew})}{\partial \lambda^2} > 0$ means this volatility effect diminishes as the average and dispersion of volatility increase.

The change in λ can also be regarded as a change in effective risk aversion. The proposition uses a scalar λ to adjust volatilities, and the mean-variance portfolio rule under the adjusted volatilities is $x^* = \frac{1}{\gamma}\Sigma^{-1}\mu = \frac{1}{\gamma\lambda^2}\tilde{\Sigma}^{-1}\mu = \frac{1}{\gamma^*}\tilde{\Sigma}^{-1}\mu$. This mean-variance solution is the optimal mean-variance portfolio with covariance matrix $\tilde{\Sigma}$, mean vector μ, and effective risk aversion of $\gamma^* = \gamma\lambda^2$. According to this formula, a mean-variance investor adjusts the proportion of funds invested in the risky assets according to the changes in volatilities (λ). With a higher (lower) λ, the investor tends to be more (less) effectively risk averse, allocating more (less) extreme portfolio weights across risky assets. Thus, the 1/N rule tends to deviate less (more) from the optimal mean-variance portfolio. The relation between risk aversion and the performance of 1/N rule is will again be discussed in Proposition 4 in this section.

This proposition can also be related to the estimation error in return moments. The estimation errors are generally small (large) when the volatilities are small (large), and thus the 1/N rule that involves no estimation error is less (more) attractive. In general, the proposition shows that large average and dispersion of volatilities are in favor of the 1/N rule, while the naive rule tends to suffer from higher loss when these quantities are smaller.

Proposition 3. *Denote the variance-covariance matrix of the asset returns as* $\Sigma = \delta\tilde{\Sigma} + (1-\delta)\tilde{\Sigma}_d$, *where* $0 < \delta \leq 1$. $\tilde{\Sigma}$ *is a benchmark variance-covariance matrix of the asset returns, and* $\tilde{\Sigma}_d$ *is an N by N diagonal matrix of* $\tilde{\Sigma}$. *For the loss function of the 1/N rule* $L(x^*, x^{ew})$, *we have:*

(1) $\frac{\partial L(x^*,x^{ew})}{\partial \delta} \propto \left(\sum_{i \neq j} \sigma_{ij} x_i^{ew} x_j^{ew} - \sum_{i \neq j} \sigma_{ij} x_i^* x_j^*\right)$;

(2) $\frac{\partial^2 L(x^*,x^{ew})}{\partial \delta^2} \geq 0$;

where σ_{ij} *is the covariance between asset i and asset j, the portfolio weight in asset i for the mean-variance portfolio is denoted as* x_i^*, *and that for the 1/N rule is denoted as* x_i^{ew}.

Economic Intuition behind Proposition 3.

The proposition suggests that assuming the correlations between asset returns change proportionally (affecting both average and dispersion of correlations), the performance of the 1/N rule (relative to the mean-variance portfolio) depends on the difference between two quantities, $\sum_{i \neq j} \sigma_{ij} x_i^{ew} x_j^{ew}$ and $\sum_{i \neq j} \sigma_{ij} x_i^* x_j^*$. The first quantity is the variance of the 1/N portfolio attributed to covariances between asset returns, while the second quantity is the variance of the optimal mean-variance portfolio attributed to covariances between asset returns. If $\sum_{i \neq j} \sigma_{ij} x_i^* x_j^* > \sum_{i \neq j} \sigma_{ij} x_i^{ew} x_j^{ew}$, the covariances in the mean-variance portfolio are more important for overall portfolio variance compared to those in the 1/N portfolio. In this situation, diminishing correlations (towards zero) will benefits more to the mean-variance portfolio rather than the 1/N portfolio, so that $\frac{\partial L(x^*,x^{ew})}{\partial \delta} < 0$. If, on the other hand, $\sum_{i \neq j} \sigma_{ij} x_i^* x_j^* < \sum_{i \neq j} \sigma_{ij} x_i^{ew} x_j^{ew}$, the covariances in the 1/N portfolio are more important for overall portfolio variance compared to those in the optimal mean-variance portfolio. In this case, diminishing correlations (towards zero) will benefits more to the naive rule in contrast to the optimal rule, so that $\frac{\partial L(x^*,x^{ew})}{\partial \delta} > 0$. In general, the correlation effect on the performance of the 1/N rule relative to the mean-variance portfolio depends on the magnitudes of the two portfolio variances attributed to the covariances between asset returns.

$\frac{\partial^2 L(x^*,x^{ew})}{\partial \delta^2} \geq 0$ suggests that the loss of the 1/N rule is a strictly convex function in δ, a parameter that controls the average and dispersion of correlations. This implies that a large loss of the 1/N portfolio occurs at either low or high correlation regimes. When correlations are low, there is more diversification effect by using the optimal mean-variance portfolio in contrast to the 1/N rule. When correlations are high, lower diversification benefits calls for the mean-variance portfolio to rely more on the mean and variance difference between assets to improve utility. In both cases, the naive rule tends to have higher mean-variance utility loss.

Proposition 4. *Assume the mean vector μ and the variance-covariance matrix of asset returns Σ are fixed. For the loss function of the 1/N rule $L(x^*, x^{ew})$, we have:*

(1) $\frac{\partial L(x^*,x^{ew})}{\partial \gamma} < 0$*;*

(2) $\frac{\partial^2 L(x^*,x^{ew})}{\partial \gamma^2} > 0$*.*

Economic Intuition behind Proposition 4.

Proposition 4 suggests that the mean-variance utility loss of the 1/N portfolio is a decreasing function in the coefficient of relative risk aversion. For a low (high) risk aversion, the mean-variance portfolio tends to have more (less) extreme portfolio weights across assets. Hence, the 1/N rule tends to deviate more (less) from the optimal mean-variance portfolio when the risk aversion is low (high). This explains why the 1/N suffers more mean-variance utility loss when the coefficient of risk aversion decreases. This risk aversion effect will magnify as the risk aversion coefficient increases, as $\frac{\partial^2 L(x^*,x^{ew})}{\partial \lambda^2} > 0$. The loss profile of the 1/N rule regarding risk aversion is consistent with the finding by Duchin and Levy (2009) who discover good performance of the 1/N rule compared to the mean-variance portfolio when risk aversion is high.

Overall, the 4 propositions theoretically show that the relative performance of the 1/N rule with respect to the optimal mean-variance portfolio is a function of return moments of constituent assets used to form portfolios. In general, the naive portfolio tends to have higher mean-variance utility loss when: (1) the average and dispersion of expected return are larger; (2) the average and dispersion of asset volatilities are smaller; (3) average correlation and correlation dispersion are either very large or very small; and (4) the coefficient of risk aversion is smaller.

B. A Numerical Example

To depict the mean-variance properties of the 1/N rule in a clearer manner, I use a numerical example to demonstrate the 4 propositions discussed earlier. The setup of the numerical example is similar to the simulation setup in DeMiguel et al. (2009) and MacKinlay and Pastor (2000). In the setup, N risky assets including one factor portfolio are considered. The excess returns of the $N-1$ risky assets are generated by the factor model $R_{it} = \alpha_i + \beta_i R_{mt} + \epsilon_{it}$. R_{it} is the excess return on asset i, α_i is the abnormal return for asset i, β_i is the factor loading for asset i, R_{mt} is the excess returns on the factor portfolio, and ϵ_{it} is the noise for asset i, which is independent with respect to the factor portfolio. The factor portfolio has an expected annual excess return of 8% and standard deviation of 16%. The mis-pricing of all the assets is assumed to be zero, and the factor loadings, β_i, are evenly spread between 0.5 and 1.5. This allows the expected excess returns vary from 4% to 12% per year. The noises are cross-sectionally independent normal random variables with zero means, and their volatilities are drawn from a uniform distribution with a range of [0.10, 0.30]. The coefficient of relative risk aversion is assumed to be one, and the annual risk-free rate follows a normal distribution with mean of 2% and standard deviation of 2%.

I regard the setup above as a baseline case, and set the number of assets to be 25. Denote the mean vector and the covariance matrix of asset return in the baseline base as $\tilde{\mu}$, and $\tilde{\Sigma}$, respectively. Define the new mean vector as $\mu = \phi\tilde{\mu}$ and covariance matrix of asset return as $\Sigma = \lambda^2(\delta\tilde{\Sigma} + (1-\delta)\tilde{\Sigma}_d)$ (where $\tilde{\Sigma}_d$ is the diagonal matrix of the baseline covariance matrix $\tilde{\Sigma}$). This design allows me to use unique parameters to adjust one single return moment of assets at a time, while holding other factors constant. I change ϕ and λ from zero to two, while δ varies from zero to one to ensure positive semi-definite for the covariance matrix. The mean-variance utility loss of the 1/N rule with respect to the optimal mean-variance portfolio is calculated, and plotted against each return moment.

[Insert Figure 1 here]

Figure 1 plots the annual mean-variance utility loss of the 1/N rule against the average return and return dispersion of assets used to construct portfolios. As both average return and return dispersion increase, the loss of the 1/N portfolio increase exponentially. This is due to the reason that the naive rule tends to deviate more (less) from the mean-variance portfolio that is more (less) likely to have extreme portfolio weights when the average expected return and return dispersion are larger (smaller). And this return effect magnifies with the increase of the average and dispersion of expected returns. The relation between the loss of the 1/N rule with returns is consistent with the prediction of Proposition 1.

[Insert Figure 2 here]

Figure 2 shows the relation between the annualized mean-variance utility loss of the 1/N rule with the average volatility and volatility dispersion of constituent assets. The loss of the naive rule monotonically decreases at a diminishing rate with the average and dispersion of volatilities. This is because when the volatilities are larger (smaller), the mean-variance portfolio tends to rely less (more) on the return difference between assets, and thus the 1/N rule tends to deviate less (more) from the mean-variance portfolio. This volatility effect is consistent with the prediction of Proposition 2.

[Insert Figure 3 here]

To show the correlation effect, Figure 3 plots the loss of the 1/N portfolio for different average and dispersion of correlations of asset returns. In general, the loss of the 1/N is a non-constant convex function in δ (reflecting both average and dispersion of correlations). As the average correlation decreases from 0.43 to about 0.3 (correlation dispersion decreases from 0.6 to 0.4), the loss of the 1/N rule decreases. This is when $\sum_{i \neq j} \sigma_{ij} x_i^* x_j^* < \sum_{i \neq j} \sigma_{ij} x_i^{ew} x_j^{ew}$. In other words, the portfolio variance attributed to covariances is larger for the naive rule compared to the mean-variance portfolio. In this case, decreasing correlations will provide more benefits to the naive rule. If we further reduce the average correlation from 0.3 towards zero (correlation dispersion decreases from 0.4 towards 0), the loss of the 1/N rule increases. In this situation, $\sum_{i \neq j} \sigma_{ij} x_i^* x_j^* > \sum_{i \neq j} \sigma_{ij} x_i^{ew} x_j^{ew}$, which means the covariances between assets in the mean-variance portfolio, in contrast to the 1/N rule, tend to play a more important role in total portfolio variance. Therefore, by further decreasing the correlations, it will benefit the mean-variance portfolio more. The correlation effect, in accordance with Proposition 3, suggests that the naive portfolio tends to have higher mean-variance utility loss when the average and dispersion of correlations are either very large or very small.

[Insert Figure 4 here]

The loss of the 1/N portfolio with regard to changing risk aversion is shown in Figure 4. The graph shows that as the coefficient of risk aversion increases, the loss of the 1/N rule decreases at a diminishing rate. This is because for a lower (higher) risk aversion, the mean-variance portfolio is constructed with more (less) extreme portfolio weights, so that the equally-weighted portfolio tends to deviate more (less) from this mean-variance portfolio. The loss profile of the naive strategy regarding risk aversion is consistent with Proposition 4 discussed earlier.

To sum up, the mean-variance properties of the 1/N rule shown in the numerical example are consistent with the 1/N loss profiles discussed in Proposition 1 to 4. Generally, the mean-variance utility loss of the 1/N rule increases with the average and dispersion of expected returns. In addition, small average and dispersion of asset volatility are less in favor of the 1/N rule. In terms of correlations, very large and very small average correlation and correlation dispersion tend to associate with more 1/N loss. Lastly, the loss of the naive portfolio increases as investors become less risk averse.

C. The Sample Selection Bias for "1/N Outperformers"

The mean-variance properties of the 1/N rule derived earlier point to the potential sample selection bias in a number of studies that propose new portfolio strategies to challenge the 1/N portfolio. The loss function of the 1/N portfolio in a number of return moments suggest that datasets with different return moments of constituent assets may produce very different results of mean-variance strategies compared to the 1/N rule. However, there is no clear guideline in the literature on the appropriateness of datasets used to justify empirical outperformance. Advocates of a new mean-variance rule could engage in a reverse engineering process of testing many datasets and "cherry-pick" the ones that shows the best performance of the proposed strategy. This creates a sample selection bias for the outperformance of mean-variance strategies over the naive benchmark. Existing literature on sample selection issues in financial studies are centered around three aspects, namely sample/database bias (Kothari et al., 1995, Chan et al., 1995 and Kim, 1997) , delisting bias (Shumway, 1997 and Shumway and Warther, 1999) and survivorship bias (Carhart et al., 2002), although they can overlap is some cases. To distinguish between the traditional sample selection bias and this new selection bias, the latter will be referred to as the dataset selection bias.

This dataset selection bias is well reflected by the datasets used in a number of studies that claim to invent new mean-variance strategies to outperform the naive 1/N rule. First, these studies invariably use a limited number (often 6-7 at most) of pre-selected datasets to empirically show the outperformance of proposed strategies over the 1/N benchmark. As suggested by Propositions 1-4, the loss of the 1/N rule is not constant with respect to a number of return moments, thus, empirical results from a limited number of selected datasets are likely to be biased. It is possible that the pre-selected datasets may consist of assets whose mean-variance properties fall at the extreme ends of the loss function of the 1/N rule, and hence, place the naive rule at an unfair disadvantage relative to mean-variance strategies. Second, these studies consider primarily characteristic-based portfolios, such as portfolios sorted on size, book-to-market, and momentum. The large dispersion of expected returns for these characteristic-based portfolios is in favor of the mean-variance strategies compared to the 1/N rule, because the utility loss of the 1/N portfolio is an increasing function in the expected return dispersion of constituent assets as shown in Proposition 1. Third, studies involving a portfolio horse race with the 1/N rule usually apply datasets of stock portfolios without considering empirical datasets of individual stocks. Proposition 2 demonstrates that the loss of the naive rule is generally smaller when volatilities are larger. As individual stocks are more volatile than stock portfolios, empirical horse races with the naive strategy that only use portfolio datasets introduce potential selection bias. It is likely that large volatilities of individual stocks could erode the outperformance of newly proposed strategies over the naive benchmark. All in all, the analysis of mean-variance properties of the 1/N portfolio reveals that the pre-selected datasets can potentially bias the results of mean-variance strategies outperforming the naive 1/N rule.

III. Data and Methodology

The discussions of mean-variance properties of the 1/N portfolio in the previous section point to the potential dataset bias in a number of studies that propose new strategies to outperform the 1/N portfolio. To examine whether such a bias exist, I intend to analyze a large number of datasets that are free from selection bias in this paper. To facilitate the analysis, I propose a new procedure that aggregates overall portfolio performance across datasets and tests performance built upon statistics in each dataset. The data, portfolio strategies and measures, as well as statistical inference methodologies are discussed in details in the rest of this section.

A. Data

Sample selection bias is generally not resolvable within the sample that is subject to the bias. Studies that investigate sample selection bias in previous research usually introduce other samples or data to argue against the original findings. For instance, Kothari et al. (1995) examine the book-to-market anomaly in S & P data, and conjecture the selection bias of COMPUSTAT data in book-to-market results by Fama and French (1992). In this paper, I intend to quantify the selected dataset bias in a number of mean-variance portfolio strategies by using datasets that are free from the bias. Specifically, in contrast to previous studies that use datasets of primarily characteristic-sorted portfolios, this paper tries to eliminate sample selection bias by randomly selecting both stocks and equity portfolios from NYSE/AMEX/NASDAQ to form datasets[2]. Each datasets consists of a block of monthly return series of 20 years randomly drawn from January 1926 to December 2014. Thus, the empirical result is not biased by any specific historical period. The data are obtained from the Center for Research in Security Prices (CRSP), and the risk-free rate is the 90-day T-bill return.

For stock datasets, a fixed number of stocks with the full 20-year data are randomly chosen. For portfolio datasets, stocks with the full 20-year data are randomly formed into a fixed number of groups (the number of portfolios). Each portfolio is equally weighted across the stocks within the group, so that naive diversification on these portfolios corresponds to an equally-weighted combination of individual stocks. Each portfolio strategy is then applied to the pre-selected stocks/portfolios

[2]Alternative sampling schemes are tested for the robustness of the results in Section IV.D

in each dataset.

Unlike previous studies that focus on only a limited number of datasets (often 6-7 at most), I test the portfolio performance across a large number of datasets. Specifically, I consider 6 types of datasets, including datasets of 10, 25 and 50 stocks, and those of 10, 25 and 50 portfolios. For each category, 1000 datasets are randomly generated, so that a summary of portfolio performance is based on a large sample of 6000 datasets.

B. Portfolio Strategies

A number of portfolio strategies have been proposed in order to challenge the 1/N puzzle that optimal diversification cannot consistently outperform naive 1/N diversification rule. In this paper, I consider 13 "1/N outperformers" proposed in the literature, namely the optimal risky portfolio by Kan et al. (2016), 3 combination strategies by Tu and Zhou (2011), and 9 timing strategies by Kirby and Ostdiek (2012). Not only because these studies are direct responses to the findings by DeMiguel et al. (2009), but also they represent a variety of both constrained and unconstrained advanced strategies under the modern portfolio theory of Markowitz (1952, 1959).

The combination strategies by Tu and Zhou (2011) aim at reducing estimation error of a sophisticated portfolio by combining the portfolio with the 1/N rule. These strategies stem from the unconstrained Markowitz mean-variance portfolio that allows levered positions. The combined portfolios are theoretically better than their uncombined counterparts, and are empirically shown to outperform the 1/N rule. Mathematically, the portfolio weight at time t is:

$$w_t^C = \hat{\delta}_t \hat{w}_t^S + (1 - \hat{\delta}_t) w_{1/N} \tag{1}$$

where $\hat{\delta}$ is the combination coefficient, $\hat{w}_t^S$ is the uncombined sophisticated portfolio, and $w_{1/N}$ is the 1/N portfolio.

Three combination strategies of Tu and Zhou (2011) are tested against the 1/N rule, namely the combination of 1/N with the sample mean-variance portfolio, the combination of 1/N with the Bayes-Stein mean-variance portfolio of Jorion (1986), and the combination of 1/N with the "three-fund" model of Kan and Zhou (2007)[3] .

[3]There is an additional combination strategy considered by Tu and Zhou (2011), which is the combination of 1/N with the "missing factor" model of MacKinlay and Pastor (2000). This strategy empirically underperform the 1/N

Kan et al. (2016) point out that the failure of sophisticated strategies tested in DeMiguel et al. (2009) is partly due to the exclusion of the risk-free asset. In order to have a direct comparison with the 1/N rule of risky asset only, they propose a optimal risky portfolio (called QL) that performs well relative to the 1/N rule in both calibrations and real datasets. The optimal risky portfolio is calculated as:

$$w_t^{QL} = \hat{w}_{g,t} + \frac{\hat{c}_t}{\gamma}\hat{w}_{z,t} \tag{2}$$

where $\hat{w}_{g,t}$ is the sample global minimum-variance portfolio, $\hat{w}_{z,t} = \hat{\Sigma}_t^{-1}(\hat{\mu}_t - 1_N)\hat{\mu}_{g,t}$ is a zero position long-short portfolio, and $\hat{c}_t$ is estimated to mitigate the estimation error.

Kirby and Ostdiek (2012) propose 9 timing strategies based on aggressive shrinkage estimators of the covariance matrix to reduce portfolio turnovers. They represent a range of constrained strategies that rely on different sets of information, such as the mean vector, covariance matrix, and conditioning variables capturing cross-sectional returns. The timing strategies are shown to outperform the 1/N portfolio empirically in presence of high transaction costs. Specifically, they develop two broad class of timing strategies, the volatility timing strategy, and the reward-to-risk timing strategy. The volatility timing strategy uses an aggressive shrinkage estimator, the diagonal matrix of the sample covariance matrix, for the global minimum-variance portfolio. An exogenously determined tuning parameter is also introduced that measures the timing aggressiveness of volatility change. Let $\Sigma_{d,t}$ denotes the diagonal matrix of the sample covariance matrix at time t, and $\mathbf{1}_N$ is a column vector of ones of size N. The portfolio weight for the volatility timing strategy with a given timing aggressiveness η is shown below:

$$w_t^{VT} = \frac{(\Sigma_{d,t}^{-1}\mathbf{1}_N)^\eta}{\mathbf{1}_N'(\Sigma_{d,t}^{-1}\mathbf{1}_N)^\eta} \tag{3}$$

The second category of the timing strategy is the reward-to-risk timing strategy which uses the diagonal of the sample covariance matrix as the covariance matrix for the tangency portfolio with shortsale constraint. Similar to the volatility timing strategy, the timing aggressiveness parameter is also introduced. Let μ_t^+ denotes the excess return vector at time t where each element return is no less than zero. The portfolio weight for the reward-to-risk timing strategy with a given timing

rule in all the real datasets in their paper, so it is excluded from the comparison in this study.

aggressiveness η is shown below:

$$w_t^{RRT} = \frac{(\Sigma_{d,t}^{-1}\mu_t^{+})^{\eta}}{\mathbf{1}_N'(\Sigma_{d,t}^{-1}\mu_t^{+})^{\eta}} \tag{4}$$

To further improve the performance of the reward-to-risk timing strategy, Kirby and Ostdiek (2012) replace the mean vector by $\bar{\beta}_t^{+} = max(\bar{\beta}_t, 0)$, where $\bar{\beta}_t$ is the average beta estimated from the Carhart (1997) 4-factor extension of the Fama and French (1993) 3-factor model. To distinguish between these two types of reward-to-risk timing strategies, I call the former the return-to-risk timing strategy and the latter the beta-to-risk timing strategy. The beta-to-risk timing strategy is listed below:

$$w_t^{BRT} = \frac{(\Sigma_{d,t}^{-1}\bar{\beta}_t^{+})^{\eta}}{\mathbf{1}_N'(\Sigma_{d,t}^{-1}\bar{\beta}_t^{+})^{\eta}} \tag{5}$$

For each category of the timing strategy, Kirby and Ostdiek (2012) consider timing aggressiveness of 1, 2 and 4. Thus, there are 9 timing strategies to be tested in the study.

[Insert Table I here]

To compare with the 1/N portfolio benchmark, I also consider three traditional approaches, the classical mean-variance portfolio, the tangency portfolio, and the global-minimum-variance portfolio. All the portfolio strategies and their abbreviations are summarized in the table above. Consistent with the literature, each portfolio is estimated every month based on a rolling estimation window of 120 months. Since each dataset has a sample period of 20 years, the evaluation period for each dataset is 10 years.

C. Performance Measures

For performance evaluations in the empirical test, I consider the Sharpe ratio, the certainty-equivalent (CEQ) return, the portfolio turnover, and outperformance frequency over the naive 1/N rule. The first three measures are considered in DeMiguel et al. (2009), while the last ones is to examine whether an "1/N outperformer" can consistently outperform the benchmark.

The out-of-sample Sharpe ratio of strategy k is defined as:

$$\hat{SR}_k = \frac{\hat{\mu}_k}{\hat{\sigma}_k} \tag{6}$$

where $\hat{\mu}_k$ and $\hat{\sigma}_k$ out-of-sample mean and volatility of excess return of strategy k net of transaction cost, respectively.

The certainty-equivalent return of the strategy k is calculated as:

$$C\hat{E}Q_k = \hat{\mu}_k - \frac{\gamma}{2}\hat{\sigma}_k^2 \tag{7}$$

where γ is the coefficient of risk aversion. For the empirical comparison, I set γ equal to 2[3], but the robustness of the results will be checked for different γ values.

The third portfolio performance measure is the portfolio turnover, which is an important measure in presence of transaction costs. The turnover of the strategy k is defined as:

$$Turnover_k = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{N}(|\hat{w}_{k,j,t+1} - \hat{w}_{k,j,t+}|) \tag{8}$$

where $\hat{w}_{k,j,t+}$ is the portfolio weight on asset i for strategy k before rebalancing at $t+1$, and $\hat{w}_{k,j,t+1}$ is the portfolio weight on asset i for strategy k after rebalancing at $t+1$.

DeMiguel et al. (2009) pose the 1/N puzzle that sophisticated mean-variance strategies can hardly outperform the 1/N portfolio in a consistent manner. To investigate the consistency of outperformance over the 1/N benchmark for a number of "1/N outperformers", I also consider the outperformance frequency as the last portfolio performance measure. This new performance metric is recently applied by Plyakha, Uppal, and Vilkov (2014) who compare the performance of equally- versus-value weighted portfolios. The outperformance frequency for the strategy k relatively to the 1/N benchmark is defined as:

$$f_k = \frac{m}{M} \tag{9}$$

where m stands for the number of datasets in which the strategy k outperforms the benchmark, and M is the total number of datasets. I consider the frequency of performance in Sharpe ratio and CEQ return, in terms of both point estimate and significant outperformance.

[1]The choice of γ in studies involving portfolio horse race varies. For instance, DeMiguel et al. (2009) set $\gamma = 1$, Tu and Zhou (2011) and Kan et al. (2016) consider 1 and 3, and Kirby and Ostdiek (2012) look at 1 and 5. I set $\gamma = 2$ because it is calculated from the implied market risk aversion by using monthly value-weighted CRSP returns for the period from Jan. 1926 to Dec. 2014. As suggested by Proposition 4, a smaller risk aversion is associated with larger loss of the 1/N rule. Therefore, a small γ is not a concern, because it biases against the 1/N rule, which biases against the results in this paper.

D. Statistical Inference

I follow the approach by Kirby and Ostdiek (2012) to conduct inferences about the relative performance of different strategies in each dataset. Let $\hat{\lambda}_{im}$ and $\hat{\lambda}_{jm}$ denote the estimated Sharpe ratios for strategies i and j in dataset m, respectively. If the 2 strategies have the same population Sharpe ratio, then we have the large-sample approximation as below:

$$\frac{\hat{\lambda}_{jm} - \hat{\lambda}_{im}}{\hat{V}_m^{1/2}} \overset{a}{\sim} N(0, 1) \tag{10}$$

where $\hat{V}_m$ denotes a consistent estimator of the asymptotic variance of $\hat{\lambda}_{jm} - \hat{\lambda}_{im}$[4].

Because there is no evidence on the quality of the approximation, the p-value for $H_0 : \lambda_{jm} - \lambda_{im} = 0$ is computed using a block bootstrap approach. Each bootstrap trial consists of 2 steps. Let $y_m = (y_{m1}, y_{m2}, ..., y_{mT})$, where $y_{mt} = (r_{imt}, r_{jmt})$, denote the set of out-of-sample excess returns for strategies i and j in dataset m. First, I construct a resample $y_m^* = (y_{m1}^*, y_{m2}^*, ..., y_{mT}^*)$ using the stationary bootstrap of Politis and Romano (1994). The resample is such that, in general, if $y_{st}^* = y_{mt}$, then $y_{s,t+1}^* = y_{m,t+1}$ with probability π and $y_{s,t+1}^*$ is drawn randomly from y_m with probability $1 - \pi$. This delivers an expected block length of $1/(1 - \pi)$. Second, I calculate

$$\hat{\theta_m^*} = \frac{(\hat{\lambda}_{jm}^* - \hat{\lambda}_{im}^*) - (\hat{\lambda}_{jm} - \hat{\lambda}_{im})}{\hat{V}_m^{*1/2}} \tag{11}$$

where $\hat{\lambda}_{im}^*$, $\hat{\lambda}_{jm}^*$ and $\hat{V}_m^*$ denote the estimates for the resample. After carrying out B bootstrap trials in total, we compute the p-values for the t-statistic in Equation 10 using the observed percentiles of $\hat{\theta_m^*}$. Due to computational reasons[5], I set B = 1000 and π= 0.9 for an expected block length of 10. I use a similar approach to assess the statistical significance of the estimated CEQ return.

---

[4]The generalized method of moments is used to construct this estimator. Let $e_t(\hat{\theta}_m) = (r_{imt} - \hat{\sigma}_{im}\hat{\lambda}_{im}, r_{jmt} - \hat{\sigma}_{jm}\hat{\lambda}_{jm}, (r_{imt} - \hat{\sigma}_{im}\hat{\lambda}_{im})^2 - \hat{\sigma}_{im}^2, (r_{jmt} - \hat{\sigma}_{jm}\hat{\lambda}_{jm})^2 - \hat{\sigma}_{jm}^2)^T$, where $\hat{\theta}_m = (\hat{\lambda}_{im}, \hat{\lambda}_{jm}, \hat{\sigma}_{im}^2, \hat{\sigma}_{jm}^2)^T$. $\hat{\theta}_m \overset{a}{\sim} N(\theta_m, \hat{D}_m^{-1}\hat{S}_m\hat{D}_m^{-1'}/T)$, where $\hat{D}_m = 1/T\sum_{t=1}^{t=T} \partial e_t(\hat{\theta}_m)/\partial\hat{\theta}_m'$ and $\hat{S}_m = \hat{\Gamma}_{0m} + \sum_{l=1}^{s}(1 - l/(s+1))(\hat{\Gamma}_{lm} + \hat{\Gamma}_{lm}')$ with $\hat{\Gamma}_{lm} = 1/T\sum_{t=l+1}^{t=T} e_t(\hat{\theta}_m)e_{t-l}(\hat{\theta}_m)'$. For the empirical analysis, s is set to be 5 and $\hat{V}_m = \hat{U}_{22} - 2\hat{U}_{21} + \hat{U}_{11}$, where $\hat{U} = \hat{D}_m^{-1}\hat{S}_m\hat{D}_m^{-1'}/T$.

[5]Kirby and Ostdiek (2012) set the number of bootstrap trial B=10000 to test performance in each dataset, but they only consider 4 datasets in their empirical study. In this study, a bootstrap trail of 10000 requires a typical computer with 24G RAM to operate 8000-9000 seconds to test one dataset. Given the large number of datasets in this study, I reduce the number of bootstrap trial to 1000. However, based on a non-tabulated robustness check, this treatment does not qualitatively affect the results.

E. *Aggregate Performance Across Datasets*

To summarize portfolio performance across datasets, I propose a new approach that aggregates overall portfolio performance and test performance built upon statistics in each dataset. This new test provides statistical inference on aggregate portfolio performance while taking into account uncertainties of performance in each dataset. To illustrate, let $\bar{\lambda}_i = 1/M \sum_{m=1}^{M} \hat{\lambda}_{im}$ and $\bar{\lambda}_j = 1/M \sum_{m=1}^{M} \hat{\lambda}_{jm}$ denote the average Sharpe ratios for strategy i and j across M datasets, where $\hat{\lambda}_{im}$ and $\hat{\lambda}_{jm}$ are the estimated Sharpe ratios for strategies i and j in dataset m.

Under the null hypothesis $H_0 : \lambda_j - \lambda_i = 0$, we have the following large-sample approximation:

$$\frac{\bar{\lambda}_j - \bar{\lambda}_i}{\hat{V}^{1/2}} \overset{a}{\sim} N(0,1) \tag{12}$$

where $\hat{V}$ denotes the variance of $\bar{\lambda}_j - \bar{\lambda}_i$.

Given the random nature of the sampled datasets, I assume that the relative performance of strategy j with respect to strategy i are uncorrelated across datasets. In other words, the covariance between $\hat{\lambda}_{jp} - \hat{\lambda}_{ip}$ and $\hat{\lambda}_{jq} - \hat{\lambda}_{iq}$ $(p \neq q)$ is assumed to be zero. Therefore:

$$\begin{aligned}
\hat{V} &= var\left[\bar{\lambda}_j - \bar{\lambda}_i\right] \\
&= var\left[\frac{1}{M}\sum_{m=1}^{M}(\hat{\lambda}_{jm} - \hat{\lambda}_{im})\right] \\
&= \frac{1}{M^2} var\left[\sum_{m=1}^{M}(\hat{\lambda}_{jm} - \hat{\lambda}_{im})\right] \\
&= \frac{1}{M^2}\sum_{m=1}^{M} var\left[\hat{\lambda}_{jm} - \hat{\lambda}_{im}\right] \\
&= \frac{1}{M^2}\sum_{m=1}^{M} \hat{V}_m \\
&= \frac{1}{M}\bar{V}
\end{aligned}$$

where $\bar{V}$ is the average of $\hat{V}_m$ across M datasets, with $\hat{V}_m$ being the estimated variance of $\hat{\lambda}_{jm} - \hat{\lambda}_{im}$ via the bootstrap method discussed in Section III.D.

Therefore, the test statistic in Equation 12 can be calculated as:

$$\sqrt{M}\left(\frac{\bar{\lambda}_j - \bar{\lambda}_i}{\bar{V}^{1/2}}\right) \overset{a}{\sim} N(0,1) \tag{13}$$

Given the large sample nature of the datasets, the Central Limit Theorem applies and the standard normal distribution is considered to draw inference on the aggregate Sharpe ratio. I use a similar approach to assess the statistical significance of the aggregate CEQ return for each strategy.

IV. Empirical Results

In this section, I will show empirical evidence of sample selection issues for 13 advanced portfolio strategies in claiming their outperformance over the 1/N rule. I will start by presenting the empirical results from a typical dataset of characteristic-based portfolios - the classic Fama-French 25 size/book-to-market portfolio dataset. The Fama-French dataset has been used in numerous portfolio studies (for instance Pástor and Stambaugh (2000), Wang (2005) and DeMiguel et al. (2009)), and Tu and Zhou (2011), Kan et al. (2016) and Kirby and Ostdiek (2012) apply the dataset to empirically show the outperformance of their strategies over the naive 1/N rule. Such a dataset involves a larger return dispersion and smaller volatilities than datasets of individual stocks. These properties place the naive 1/N portfolio at a comparative disadvantage with respect to sophisticated mean-variance strategies, which create a potential sample selection bias for each "1/N outperformer". Testing portfolio performance using a large number of sampled datasets is one way to tackle this sample selection problem. To analyze this selection problem, I will mainly focus on the results from sampled datasets of stocks and portfolios, and show how they differ from the results based on Fama-French dataset.

A. Results from the Fama-French Dataset

Table II summarizes the annualized performance of each portfolio strategy based on the Fama-French 25 size/book-to-market portfolio dataset. For non-naive portfolios, Sharpe ratios and CEQ returns are tested against the naive 1/N benchmark which delivers an annualized Sharpe ratio of 0.523 and an annualized CEQ return of 6.14%. Regarding traditional strategies, the classic mean-variance portfolio (MV) generates very huge return and volatility, and requires to turn over its total portfolio wealth by over 65 times each year. Although it outperforms the 1/N rule with respect to the Sharpe ratio, it significantly underperforms the benchmark based on CEQ return. The tangency portfolio (TP) involves huge estimation errors reflected by huge turnover and volatility. It delivers negative Sharpe ratio and CEQ return, underperforming the 1/N portfolio significantly. The sample minimum-variance portfolio (MIN), on the hand hand, requires less turnover than MV and TP, and outperforms the 1/N benchmark in terms of Sharpe ratio and CEQ return in a statistically significant manner.

[Insert Table II here]

With respect to the 13 "1/N outperformers", results generally suggest that they outperform the 1/N rule on a risk adjusted basis. Three combination strategies (CML, CPJ and CKZ) outperform the 1/N rule by Sharpe ratio. In terms of economic significance, they deliver annualized CEQ returns of 24-28%, about four times larger than that for the naive rule. Although performing worse than the combination strategies, the optimal risky portfolio also outperforms the naive rule in terms of both Sharpe ratio and CEQ return. Characterized by portfolio constraints and small turnovers, the nine timing strategies also show consistent outperformance over the benchmark. They generally have annualized Sharpe ratios from 0.54 to 0.61, and average CEQ returns about 7%. Six timing strategies statistically outperform the naive 1/N rule, while all nine strategies outperform the benchmark in terms of the point estimates of the Sharpe ratio and CEQ return (except for VT4 which deliver statistically indifferent CEQ return as the 1/N rule). The Fama-French dataset indeed demonstrates the outperformance over the 1/N rule for these "1/N outperformers".

B. Results from Sampled Datasets of Stocks

Table III shows the annualized out-of-sample portfolio performance based on 1000 datasets of 10 stocks randomly selected from NYSE/AMEX/NASDAQ. The 1/N benchmark on average delivers an annualized Sharpe ratio of 0.582 (0.562) before (after) transaction cost of 50 basis points. The annual certainty-equivalent (CEQ) return is 6.94% on average for the naive strategy, and it reduces to 6.57% when transaction cost is considered. The average portfolio turnover for the 1/N rule is 74.5% on an annual basis. Compared with the naive benchmark, traditional portfolios all generate unfavorable results in terms of all the performance measures. Both classic mean-variance portfolio (MV) and the minimum-variance portfolio (MIN) deliver significantly lower Sharpe ratios and CEQ returns before and after transaction costs compared to the 1/N benchmark. Due to huge portfolio turnover and estimation error, the tangency portfolio (TP) possesses a small Sharpe ratio and huge negative CEQ return. For the 13 "1/N outperformers", the combination strategies perform better than the sample mean-variance and the tangency portfolio. However, all three combination strategies deliver significantly lower average Sharpe ratio and CEQ return (both with and without transaction costs) than the 1/N portfolio. The optimal risky portfolio also underperforms the naive benchmark significantly. The timing strategies generally perform better than the combination strategies and the optimal risky portfolio. In particular, two volatility time strategies (VT1 and VT2) and two beta-to-risk timing strategies (BRT1 and BRT2) deliver better Sharpe ratio and CEQ return on average than the 1/N benchmark for both with and without transaction costs. However, there still exist five out of nine timing strategies that underperform the naive rule significantly in terms of both Sharpe ratio and CEQ return with and without transaction cost. With respect to portfolio turnover, only a series of volatility timing strategies delivers lower turnover than the 1/N portfolio. All other portfolio strategies require higher portfolio turnover than the naive benchmark.

[Insert Table III here]

When the outperformance frequency for each portfolio strategy is considered, no strategy shows consistent outperformance over the naive benchmark. The combination strategies on average outperform (in terms of both Sharpe ratio and CEQ return) the 1/N portfolio by less than 30% of the datasets. The optimal risky portfolio is better, with outperformance frequency about 30% with transaction cost. The timing strategies are the best in general. The VT1 strategy outperforms the 1/N rule in terms of Sharpe ratio (CEQ return) with 61% (48%) of the chance. Nonetheless, even the best one show very poor results when significant outperformance is considered. All the portfolios considered cannot significantly outperform the 1/N rule with more than 10% chance in terms of Sharpe ratio, or with more than 5% chance in terms of CEQ return. Overall, none of the 13 "1/N outperformers" investigated can consistently outperform the naive diversification strategy.

[Insert Table IV here]

Table IV summarizes the results of portfolio performance based on 1000 datasets of 25 stocks. Similar to results in datasets of 10 stocks, combination strategies and the optimal risky portfolio deliver significantly lower Sharpe ratios and CEQ returns than the 1/N portfolio. Only 4 (VT1, VT2, BRT1 and BRT2) out of the 13 strategies generate on average significantly higher Sharpe ratio than the naive benchmark, but they fail to outperform the benchmark in terms of the CEQ return. In fact, all (12) of the 13 strategies delivers significantly lower average CEQ return than the 1/N rule with (without) transaction cost. In terms of outperformance frequency, none of 13 "1/N outperformers" can outperform the 1/N portfolio consistently out-of-sample. Sophisticated portfolios significantly outperform the naive benchmark by no more than 10% on average, which suggests that in presence of estimation error, advanced mean-variance portfolios are hard to justify superior performance over naive diversification.

[**Insert Table V here**]

Results from datasets of 50 stocks presented in Table V are qualitatively the same as results from datasets of 10 and 25 stocks. The optimal risky portfolio still underperforms the naive benchmark significantly in terms of both Sharpe ratio and CEQ return. The combination strategies are getting worse compared to the performance when N=10 or 25, reflected by the reduced outperformance frequency over the naive benchmark. This is because when N is larger, there is more estimation error involved in the matrix inversion, so that mean-variance portfolios that involve inverting covariance matrices tend to perform worse. This is consistent with the finding by DeMiguel et al. (2009) who show theoretically that sample mean-variance portfolios need longer samples to estimate valid means and covariance matrix in competing with 1/N when N gets larger. For timing strategies that involves no matrix inversion, the results are not so much affected by the increased N. However, no timing strategy shows true outperformance over the naive rule. The best performing strategy BRT1 delivers a net annualized CEQ return of 6.87%, while that for the 1/N portfolio is 6.94%. Moreover, all the non-naive strategies cannot consistently outperform the benchmark, and they generally fail to significantly outperform the 1/N rule in terms of CEQ return by more than 10% chance.

In summary, results from the sampled datasets of individual stocks can be summarized as follows. First, about 8 out of the 13 "1/N outperformers" underperform the 1/N portfolio significantly in terms of both average Sharpe ratio and CEQ return. Second, only 3-4 out of the 13 advanced portfolio strategies deliver on average higher Sharpe ratio than the 1/N benchmark, but they underperform the naive rule significantly in terms of CEQ return. Third, when outperformance frequency is considered, none of the 13 "1/N outperformers" can consistently outperform the 1/N portfolio. Last but not least, statistically significant outperformance is only achieved in no more than 10% of datasets investigated for all the sophisticated strategies.

C. Results from Sampled Datasets of Portfolios

The results for 1000 datasets of 10 randomly formed portfolios are presented in Table VI. The 1/N portfolio on average delivers 0.682 (0.677) annualized Sharpe ratio and 7.93% (7.86%) annualized CEQ return before (after) transaction costs. It requires 14.1% annual turnover. Combination strategies and the optimal risky portfolio all underperform the naive strategy significantly in terms of both Sharpe ratio and CEQ return. The timing strategies deliver very similar performance as the naive benchmark in terms of all the three performance measures. None of the timing strategies is able to beat the naive benchmark in terms of both statistical and economic significance. For instance, the VT1 strategy on average generates 0.683 (0.678) annualized Sharpe ratio and 7.93% (7.86%) annual CEQ return before (after) transaction costs. It cannot distinguish itself from the naive rule in terms of both statistical and economical significance. The turnover of the strategy is 15.2% on an annual basis, slightly higher than that of the naive benchmark. Similar to the results from the stocks datasets, all portfolio rules in the datasets of 10 portfolios cannot consistently outperform the 1/N benchmark. Moreover, none of the portfolios perform very well in terms of significant outperformance frequency over the naive diversification. "1/N outperformers" significantly outperform the benchmark by no more than 10-11% (6-7%) chance in terms of Sharpe ratio before (after) transaction costs. The significant outperformance frequency is even as low as almost 0% for CEQ return when transaction cost is considered.

[Insert Table VI here]

The results from datasets of 25 and 50 portfolios are presented in Table VII and VIII below. All the combination strategies still have significantly smaller Sharpe ratios and CEQ returns than the 1/N benchmark. They tend to produce very poor results when transaction cost is considered, due to very large turnovers. The optimal risk portfolio significantly underperforms the naive strategy in terms of both Sharpe raio and CEQ return when transaction cost is considered. Although there exist some portfolio strategies such as volatility timing and beta-to-risk timing strategies having higher Sharpe ratio than the naive benchmark, the outperformance is not economically significant. Most of these strategies underperform the naive rule in terms of CEQ return. For instance, for datasets of 25 portfolios, while the 1/N rule delivers a net CEQ return of 7.61%, these volatility timing and beta-to-risk timing strategies generate by no more than this return.

[Insert Table VII and VIII here]

The outperformance frequency over the naive benchmark for the optimal risky strategy and various combination strategies are no more than 40% for datasets of 25 and 50 portfolios. They even produce close to zero outperformance frequency when transaction cost is considered. The large turnovers in these portfolios hamper portfolio performance. Although the outperformance frequencies for several timing strategies increase when portfolios rather than stocks are used as asset for portfolio construction, they are still far away from consistent outperformance over naive diversification. Significant outperformance is generally achieved in less than 10 % of the datasets for most sophisticated portfolio strategies.

To sum up, switching from datasets of stocks to stock portfolios, the results are qualitatively the same. Results based on sampled portfolio datasets can be summarized below. First, the majority of the 13 "1/N outperformers" significantly underperforms the 1/N portfolio. Second, only 3-4 out of these 13 strategies deliver on average slightly higher Sharpe ratio than the 1/N benchmark, but their CEQ returns almost always underperform that of the naive rule. Last, not only the 13 "1/N outperformers" are unable to outperform the 1/N portfolio consistently out of sample, but also they fail to outperform the naive benchmark significantly in more than 10% of datasets. And some strategies almost never significantly outperform the 1/N rule.

Comparing the results from the sampled datasets with those from the Fama-French dataset, we find a clear distinction between the performance of the "1/N outperformers". When Fama-French portfolios are considered, these sophisticated mean-variance strategies outperform the naive 1/N rule in both statistically and economically significant fashion. However, the outperformance is subject to the selection bias of the dataset. Specifically, the Fama-French portfolios have a large expected return dispersion, and they also possess smaller volatilities compared to individual stocks. These properties of the dataset potentially provide the mean-variance strategies an unfair advantage over the naive benchmark, and thus, result in a sample selection bias. To reduce the selection bias, empirical comparisons are conducted upon sampled datasets of stock and portfolios. The results from these sampled datasets, however, suggest the contrary. A majority of the 13 "1/N outperformers" investigated in fact underperform the naive diversification strategy significantly. Although there may exist 1 or 2 strategies comparable to naive benchmark in certain scenarios, there is no outperformance found. The overall picture of this empirical comparison suggests that there is a strong dataset selection bias in claiming outperformance over the 1/N rule for a number of advanced mean-variance strategies, and that the naive diversification strategy remain hard to beat when constructing stock portfolios.

D. Robustness Checks

In this section, I discuss the robustness of the sample selection bias. The robustness check is conducted in several aspects, including different risk aversion, portfolio constraints, expanding versus rolling estimation window, alternative length and sampling scheme of the datasets.

D.1. Different Risk Aversion

It is shown in Proposition 4 that the loss of the 1/N rule with respect to the optimal mean-variance portfolio depends on risk aversion. Thus, risk aversion coefficient may affect the results. I change the risk aversion to 1 and 5, and re-test each portfolio against the 1/N rule by using the 6000 datasets. Results show that sophisticated strategies tend to improve performance with respect to the 1/N rule when γ is larger, especially for the unconstrained combination strategies. This is because with a larger risk aversion, a mean-variance rule is calculated with less extreme portfolio weights, and thus, reduces estimation error and improves performance. However, none of the sophisticated portfolio can consistently outperform the naive benchmark, and significant outperformance is achieved by no more than 10 % for each advanced strategy.

D.2. Portfolio Constraint

According to Jagannathan and Ma (2003), shortsale constraints have a shrinkage effect on portfolio weights, and thus, may improve portfolio performance by reducing estimation error. For timing strategies, they involve non-negative portfolio weights by construction, so only the optimal risky portfolio and the combination strategies are imposed with shortsale constraints and tested against the naive rule. Results show that shortsale constraints indeed improve portfolio performance. For instance, with the constraint, each combination strategy is able to deliver higher CEQ return than the 1/N rule, but they on average significantly underperform the naive benchmark in terms of Sharpe ratio. In addition, none of the constrained portfolio can consistently outperform the naive benchmark, and significant outperformance is achieved by no more than 10 % for each constrained strategy.

D.3. Expanding versus Rolling Estimation Window

DeMiguel et al. (2009) show that a larger sample size stabilizes portfolio weights and improves portfolio performance. Therefore, an expanding rather than rolling estimation window is applied for each portfolio strategy when compete with the 1/N rule. The result is qualitatively the same as the one with rolling window: (1) none of the constrained portfolio can consistently outperform the naive benchmark, (2) significant outperformance is achieved by no more than 10 % for each constrained strategy, and (3) a majority of the strategies on average underperform the naive 1/N rule.

D.4. Alternative Length of the Datasets

The baseline datasets each has a length of 20 years, including 10 years for estimation and 10 years for evaluation. I also check whether a different length of the datasets may affect the results. To do this, I generate datasets with lengths of 15 years, 30 years and 40 years, with each category consisting of 6000 datasets as in the baseline case. Extensive empirical comparison between the 1/N rule and sophisticated strategies is conducted based on datasets with different length. Results show that the empirical evidence of the sample selection bias is robust to a different length of the datasets.

D.5. Different Sampling Scheme of the Datasets

The datasets applied in the empirical comparison use randomly selected samples. To check whether the result is robust to alternative sampling scheme, I sample the datasets with the assets that capture popular anomalies (characteristics), such as size, value and momentum. I sort CRSP stocks into deciles according to a certain characteristic, and randomly select the ones in the bottom (top) decile for small (value or momentum) stocks. Each dataset has a length of 20 years, including 10 years for estimation and 10 years for evaluation. If a stock falls out of the target decile, it is replaced by another satisfying stock. For each anomaly, I construct 1000 stock datasets for each case of 10, 25 and 50 stocks, and repeat the portfolio comparison. Results show that although some mean-variance rules (such as BRT1 and BRT2) may outperform the 1/N rule on average basis in cases of value and momentum stocks, a majority of the 13 Markowitz rules underperform the naive benchmark. The sample selection bias is robust to different sampling scheme of the datasets.

Overall, the sample selection bias of the outperformance over the 1/N rule for various mean-variance strategies is robust in several aspects. These include different risk aversion, portfolio constraint, expanding versus rolling estimation window, and alternative length and sampling scheme of the datasets.

V. Asset Characteristics and Portfolio Performance

In this section, I study whether investors can exploit the information on assets' return moments when selecting between optimal and naive diversification. The empirical results from the previous section suggest that the naive 1/N rule is still hard to be defeated when stocks or stock portfolios are used to form the optimal portfolio. However, sophisticated strategies are able to outperform the naive rule in some scenarios. If investors can identify these scenarios ex ante, it is likely to improve their returns on a risk adjusted basis. Motivated by the theoretical properties of the mean-variance portfolio relative to the 1/N rule derived earlier, I develop four hypotheses that are used to predict the performance of a mean-variance strategy relative to the naive rule. Next, predictive results are presented based on the estimates from the 6000 sampled datasets used in the empirical analysis. Finally, I will carry out a portfolio switching exercise to demonstrate how investors could use this predictability to improve portfolio performance.

A. Predictive Regression

I consider all the 13 "1/N outperformers" examined in this study, and try to predict the out-of-sample CEQ return difference between a sophisticated strategy and the naive 1/N benchmark. For the predictors, I select a number of in-sample assets' return moments. These return moments include the average asset return ($\bar{\mu}$), return dispersion ($\Delta\mu$), average volatility ($\bar{\sigma}$), volatility dispersion ($\Delta\sigma$), average pair-wise correlation ($\bar{\rho}$), correlation dispersion ($\Delta\rho$). The interaction terms for each return moment ($\bar{\mu} \times \Delta\mu$, $\bar{\sigma} \times \Delta\sigma$ and $\bar{\rho} \times \Delta\rho$) are of particular importance, because these quantities theoretically affect the performance of the 1/N rule, which is proven in Section II. DeMiguel et al. (2009) shows that estimation error in mean-variance portfolios increases with the number of assets. Therefore, the number of assets (N) is also considered as one of the predictors for portfolio performance. Brown et al. (2013) find that the performance of naive diversification relative to optimal diversification represents a compensation for the increase in tail risk and the reduced upside potential associated with the concave payoff. For this reason, I also look at a number of tail risk measures as control variables, including the average skewness ($\bar{S}$), skewness dispersion (ΔS), average kurtosis ($\bar{K}$), and kurtosis dispersion (ΔK).

Define ΔCEQ_i as the out-of-sample annualized CEQ return difference in dataset i between a sophisticated portfolio and the 1/N portfolio. The following regression is estimated across M ($M = 6000$ in this case) datasets.

$$\begin{aligned} \Delta CEQ_i =& \alpha + \beta_1 N_i + \beta_2 \bar{\mu}_i + \beta_3 \Delta\mu_i + \beta_4 \Delta\sigma_i + \beta_5 \bar{\sigma}_i + \beta_6 \Delta\rho_i + \beta_7 \bar{\rho}_i \\ & + \gamma_1 \bar{\mu}_i \times \Delta\mu_i + \gamma_2 \bar{\sigma}_i \times \Delta\sigma_i + \gamma_3 \bar{\rho}_i \times \Delta\rho_i \\ & + Controls + e_i (i = 1, ..., M) \end{aligned} \tag{14}$$

where $\Delta\mu_i$, $\Delta\sigma_i$, $\Delta\rho_i$ are the dispersions in mean return, volatility and correlation, respectively, for the annualized excess returns in dataset i; $\bar{\mu}_i$, $\bar{\sigma}_i$, $\bar{\rho}_i$, are the average levels of expected return, volatility and correlation, , respectively; and N_i refers the number of assets in dataset i. Control variables are tail risk measures including $\bar{S}_i$ (average skewness), $\bar{K}_i$ (average kurtosis), ΔS_i (skewness dispersion) and ΔK_i (kurtosis dispersion). All of these return moments are computed based on only the in-sample information in each dataset.

B. Hypotheses

Based on the predictive regression and the properties of naive diversification derived earlier in this paper, I proposed several hypotheses regarding the predictive directions for a number of variables.

H1 - Expected return effect: Consistent with Proposition 1, a larger average expected return and a wider expected return dispersion collectively predict better performance of a mean-variance strategy with respect to the naive 1/N rule ($\gamma_1 > 0$).

H2 - Volatility effect: Consistent with Proposition 2, a smaller average volatility and a narrower volatility dispersion collectively predict better performance of a mean-variance strategy with respect to the naive 1/N rule ($\gamma_2 < 0$).

H3 - Diversification effect: Consistent with Proposition 3, a smaller average correlation and a narrower correlation dispersion collectively predict better performance of a mean-variance strategy with respect to the naive 1/N rule ($\gamma_3 < 0$).

H4 - Heterogeneity effect: Consistent with Proposition 3, a larger average correlation and a wider correlation dispersion collectively predict better performance of a mean-variance strategy with respect to the naive 1/N rule ($\gamma_3 > 0$).

C. Predictive Results

The Equation 14 is estimated based on the 6000 sampled datasets used in this paper, and Table IX reports the predictive results. In general, in-sample return moments have a strong predictive power for the out-of-sample portfolio performance. The predictive regression explains more than 5% of the out-of-sample performance for all the 13 strategies reflected by the adjusted R-square. However, sophisticated strategies show a large variation in predictability. For instance, the regression explains on average 32% to 33% of the performance for the reward-to-risk timing strategies (RRT1, RRT2, RRT4) relative to the 1/N benchmark, while the R-square is about 5% to 6% for several beta-to-risk timing strategies (BRT1, BRT2, BRT4). Different strategies have various sensitivities to the return moments of assets used for portfolio construction.

[Insert Table IX here]

Individual predictors deliver predictive directions consistent with the hypotheses above. For the expected return effect, 11 out of the 13 strategies are associated with better performance when the average expected return and expected return dispersion are larger. Although the remaining two strategies have negative slopes on $\bar{\mu} \times \Delta\mu$, they are statistically insignificant. Take the QL strategy for example to illustrate the economic significance. If the average expected return and expected return dispersion both increase by 10%, the strategy improves its CEQ return by 26 basis points relative to the naive 1/N rule.

The volatility effect is much stronger than the expected return effect. The slope coefficients on $\bar{\sigma} \times \Delta\sigma$ are significantly negative at 1% level for all the 13 strategies considered. The magnitude of the coefficients also have economic significance. For instance, if the average volatility and volatility dispersion both increase by 10%, the CPJ and CKZ strategy improve its CEQ return by about 45 basis point relative to the naive 1/N rule.

In terms of impact of correlations, the heterogeneity effect dominates the diversification effect. The diversification effect suggests that a mean-variance strategy provides more diversification benefit relative to the 1/N rule when average correlation and correlation dispersion are small. The heterogeneity effect, on the hand, suggests that when average correlation is large and diversification is hard, a mean-variance strategy improves its performance relative to the naive rule by taking advantage of heterogeneity of assets reflected by a large dispersion in correlations. The results show that the slope on $\bar{\rho} \times \Delta\rho$ are significantly positive at 1% level for all the 13 strategies, which supports the heterogeneity effect in contrast to the diversification effect.

In summary, the predictive results suggest: (1) the predictability of performance based on return moments is strong for various mean-variance strategies; (2) the expected return effect is valid for most of the strategies (with the exceptions of CPJ and CKZ); (2) the volatility effect is very strong for all the strategies; and (4) the heterogeneity effect dominates the diversification effect in all the 13 cases.

D. A Portfolio Switching Strategy

To examine whether the predictability of portfolio performance based on return moments can be translated into better out-of-sample portfolio benefits, I propose a portfolio switching strategy based on forecasts from the cross-sectional predictive regression. The switching strategy is a binary choice between a sophisticated mean-variance portfolio and the 1/N portfolio depending on the predictive values of CEQ return difference between the two portfolios. If the predicted CEQ return difference is positive (negative), the switching strategy applies the sophisticated portfolio (1/N rule) for the entire out-of-sample period. Mathematically, the switching strategy takes the following form:

$$w^{Switch} = I_{\{\Delta C\hat{E}Q>0\}} w^* + (1 - I_{\{\Delta C\hat{E}Q>0\}}) w_{1/N} \tag{15}$$

where w^* is one of the 13 mean-variance rules examined in this study, and $w_{1/N}$ is the 1/N rule. $I_{\{\Delta C\hat{E}Q>0\}}$ is an indicator function which takes the value of one (zero) if the predicted CEQ return difference between the two rules is positive (negative).

I demonstrate the benefit of the switching strategy in contrast to the raw mean-variance rules by using the 6000 sampled datasets. Recall that the 6000 datasets used in this paper consist of 6 groups, datasets of 10, 25 and 50 stocks and datasets of 10, 25 and 50 portfolios. The predictive ability of return moments may vary across groups, and thus, I apply the predictive regression for each sophisticated strategy within each group. Specifically, for each group of size 1000, I use 999 observations to estimate predictive regressions, and leave out one observation for out-of-sample validation. The cross-sectional predictive regression (Equation 14 without the variable N_i which does not change within each group) is applied for the 999 in-sample observations, and a forecast is generated based on the estimated regression and the in-sample return moments for the out-of-sample observation. For each sophisticated portfolio, the switching strategy is then applied based on the predicted out-of-sample performance, and is exercised across the 6000 datasets.

[**Insert Table X here**]

Table X summarizes the performance of the switching strategy for each of the 13 "1/N outperformers", and compares the strategy with the raw mean-variance rule and the 1/N rule. Based on the aggregate performance of the 6000 datasets, the 1/N rule on average generates an annualized CEQ return of 7.57%. Consistent with the empirical results shown earlier, no mean-variance rules on average deliver significantly higher CEQ returns than the 1/N rule, and most of the strategies underperform the naive rule by economically significant margin. However, when each Markowitz rule is combined with the 1/N rule via the switching strategy, the portfolio performance improves significantly. For instance, the CML rule with the switching option delivers a mean CEQ return of

7.93% on an annual basis, while the CEQ return of the raw strategy is significantly smaller (5.75%). Not only do all of the 13 switching strategies outperform their raw counterparts significantly, they also outperform the 1/N rule in a statistically significant manner. Moreover, some switching strategies, such as CPJ and CKZ, even deliver economically significant portfolio improvement over the 1/N rule.

Overall, the study on predictability of out-of-sample portfolio performance based on in-sample return moments reveals several interesting findings. First, return moments show strong power in predicting the performance of mean-variance rules relative to the naive 1/N rule. However, predictability varies across strategies. Second, motivated by the theoretical performance of optimal versus naive diversification, three predictive effects are identified empirically, namely the expected return effect, volatility effect and the heterogeneity effect. Third, the predictability can be used to improve portfolio performance by a statistically and economically significant margin. The switching strategy is proposed here to show the existence of a portfolio benefit, and there are potentially other advanced methods that can be used to predict and further improve portfolio performance. An investigation of such alternative methods is beyond the scope of this paper.

VI. Conclusions

In this paper, I document a sample selection bias in a novel context – portfolio horse race. It refers to the bias that a number of mean-variance strategies outperform the naive 1/N rule in only a limited number of selected datasets. I analyze the bias theoretically by proving that the performance of the naive rule with respect to the mean-variance portfolio is a function of a number of return moments. I quantify the bias empirically by comparing the 1/N rule with thirteen mean-variance strategies that are claimed to outperform the naive benchmark based on a large number of datasets in an sampled setting. Results show that none of these "1/N outperformers" consistently outperform the 1/N rule. In addition, a majority of these sophisticated strategies on average underperform the naive benchmark significantly. More surprisingly, all the mean-variance rules considered cannot significantly outperform the 1/N rule with more than 10% chance in terms of Sharpe ratio, or with more than 5% chance in terms of CEQ return.

The study also provides important implications for investment decisions. First of all, the naive

1/N portfolio rule is still hard to be defeated by sophisticated mean-variance strategies consistently. However, the return moments of assets used to construct portfolios convey useful information in selecting between optimal versus naive diversification. Theoretically and empirically, I identify three effects that predict the out-of-sample performance of a mean-variance strategy relative to the naive 1/N rule based on the in-sample return moments, namely the expected return effect, the volatility effect and the heterogeneity effect. More importantly, the predictability improves portfolio performance by a statistically and economically significant margin. Specifically, a portfolio switching strategy is proposed which selects between a sophisticated portfolio and the 1/N rule based on the forecast from a predictive model of performance on return moments. Empirical results show that the switching strategy outperforms not only its raw sophisticated counterparts but also the naive 1/N benchmark significantly.

Last but not least, this paper can be regarded as a lesson for academic studies involving portfolio horse races. Recent literature experiences a huge trend of studies developing advanced mean-variance strategies to challenge the naive 1/N rule. The empirical results in this paper suggest that these empirical successes are biased to some extent because they only exist in limited scenarios. Newly developed mean-variance strategies that seek to outperform the 1/N portfolio should not merely show their outperformance in a small number of datasets that consist of primarily characteristic-based portfolios. To mitigate the dataset selection bias, they should be instead tested across a large of number datasets.

Appendix A. Proof of Propositions

Proof of Proposition 1:

Denote the expected return vector of the assets as $\mu = \phi\tilde{\mu}$ ($\phi > 0$), where $\tilde{\mu}$ ($\tilde{\mu} \neq 0 \times \mathbf{1}_N$) is a N by 1 benchmark mean vector for asset returns The first order derivative of the loss function of the 1/N rule with respect to ϕ is derived below.

$$
\begin{aligned}
\frac{\partial L(x^*, x^{ew})}{\partial \phi} &= \frac{\partial L}{\partial \mu}\frac{\partial \mu}{\partial \phi} \\
&= \frac{\partial}{\partial \mu}\left[\frac{1}{2\gamma}\left(\mu^T\Sigma^{-1}\mu - \frac{(\mathbf{1}_N^T\mu)^2}{\mathbf{1}_N^T\Sigma\mathbf{1}_N}\right)\right]\frac{\partial \mu}{\partial \phi} \\
&= \frac{\partial}{\partial \mu}\left[\frac{1}{2\gamma}\left(\mu^T\Sigma^{-1}\mu - \frac{(N\bar{\mu})^2}{\mathbf{1}_N^T\Sigma\mathbf{1}_N}\right)\right]\frac{\partial \mu}{\partial \phi} \\
&= \frac{\mu^T\Sigma^{-1}}{\gamma}\tilde{\mu} \\
&= \frac{\phi}{\gamma}\tilde{\mu}^T\Sigma^{-1}\tilde{\mu}
\end{aligned}
$$

$\tilde{\mu}^T\Sigma^{-1}\tilde{\mu}$ is the square of the Sharpe ratio for the optimal mean-variance portfolio when the mean vector is $\tilde{\mu}$, so $\tilde{\mu}^T\Sigma^{-1}\tilde{\mu} > 0$. For an risk averse investor, $\gamma > 0$. Therefore, $\frac{\partial L(x^*,x^{ew})}{\partial \phi} = \frac{\phi}{\gamma}\tilde{\mu}^T\Sigma^{-1}\tilde{\mu} > 0$

Since $\tilde{\mu}^T\Sigma^{-1}\tilde{\mu}$ is independent of ϕ, $\frac{\partial}{\phi}\left(\frac{\partial L}{\partial \phi}\right) = \frac{\partial}{\phi}\left(\frac{\phi}{\gamma}\tilde{\mu}^T\Sigma^{-1}\tilde{\mu}\right) = \frac{1}{\gamma}\tilde{\mu}^T\Sigma^{-1}\tilde{\mu} > 0$

Proof of Proposition 2:

Denote the variance-covariance matrix of the asset returns as $\Sigma = \lambda^2\tilde{\Sigma}$, where $\lambda > 0$ controls for the average of the volatilities and $\tilde{\Sigma}$ is a benchmark variance-covariance matrix of the asset returns. The first order derivative of the loss function of the 1/N rule with respect to λ is derived below.

$$
\begin{aligned}
\frac{\partial L(x^*, x^{ew})}{\partial \lambda} =& \frac{\partial}{\partial \lambda}\left[\frac{1}{2\gamma}\left(\mu^T\Sigma^{-1}\mu - \frac{(N\bar{\mu})^2}{\mathbf{1}_N^T\Sigma\mathbf{1}_N}\right)\right] \\
=& \frac{1}{2\gamma}\left[\mu^T\frac{d\Sigma^{-1}}{d\lambda}\mu + \frac{(\mathbf{1}_N^T\mu)^2}{(\mathbf{1}_N^T\Sigma\mathbf{1}_N)^2}\mathbf{1}_N^T\frac{d\Sigma}{d\lambda}\mathbf{1}_N\right] \\
=& \frac{1}{2\gamma}\left[-\mu^T\Sigma^{-1}\frac{d\Sigma}{d\lambda}\Sigma^{-1}\mu + \left(\frac{\mathbf{1}_N^T\mu}{\mathbf{1}_N^T\Sigma\mathbf{1}_N}\right)^2\mathbf{1}_N^T\frac{d\Sigma}{d\lambda}\mathbf{1}_N\right] \\
=& \frac{1}{2\gamma}\left[-\mu^T(\lambda^{-2}\tilde{\Sigma}^{-1})(2\lambda\tilde{\Sigma})(\lambda^{-2}\tilde{\Sigma}^{-1})\mu + \left(\frac{\mathbf{1}_N^T\mu}{\mathbf{1}_N^T\lambda^2\tilde{\Sigma}\mathbf{1}_N}\right)^2\mathbf{1}_N^T(2\lambda\tilde{\Sigma})\mathbf{1}_N\right] \\
=& \frac{1}{\gamma\lambda^3}\left[\frac{(\mathbf{1}_N^T\mu)^2}{\mathbf{1}_N^T\tilde{\Sigma}\mathbf{1}_N} - \mu^T\tilde{\Sigma}^{-1}\mu\right] \\
=& \frac{1}{\gamma\lambda}\left[\frac{(\mathbf{1}_N^T\mu)^2}{\mathbf{1}_N^T\Sigma\mathbf{1}_N} - \mu^T\Sigma^{-1}\mu\right] \\
=& \frac{1}{\gamma\lambda}\left(S_{ew}^2 - S_*^2\right) \\
=& -\frac{2L(x^*, x^{ew})}{\lambda}
\end{aligned}
$$

Unless the 1/N portfolio coincides with the optimal mean-variance portfolio, the loss function of the 1/N rule $L(x^*, x^{ew})$ is positive. Therefore, $\frac{\partial L(x^*,x^{ew})}{\partial \lambda} < 0$.

The second order derivative of the loss function of the 1/N rule with respect to λ is:

$$
\begin{aligned}
\frac{\partial^2 L(x^*, x^{ew})}{\partial \lambda^2} =& \frac{\partial}{\partial \lambda}\left(\frac{\partial L}{\partial \lambda}\right) \\
=& \frac{1}{\gamma\lambda^3}\left(\frac{(\mathbf{1}_N^T\mu)^2}{\mathbf{1}_N^T\tilde{\Sigma}\mathbf{1}_N} - \mu^T\tilde{\Sigma}^{-1}\mu\right) \\
=& \frac{\partial}{\partial \lambda}\left[\frac{1}{\gamma\lambda^3}\left(\frac{(\mathbf{1}_N^T\mu)^2}{\mathbf{1}_N^T\tilde{\Sigma}\mathbf{1}_N} - \mu^T\tilde{\Sigma}^{-1}\mu\right)\right] \\
=& -\frac{3}{\gamma\lambda^4}\left(\frac{(\mathbf{1}_N^T\mu)^2}{\mathbf{1}_N^T\tilde{\Sigma}\mathbf{1}_N} - \mu^T\tilde{\Sigma}^{-1}\mu\right) \\
=& \frac{6}{\lambda^4}\times\frac{1}{2\gamma}\left(\mu^T\tilde{\Sigma}^{-1}\mu - \frac{(\mathbf{1}_N^T\mu)^2}{\mathbf{1}_N^T\tilde{\Sigma}\mathbf{1}_N}\right)
\end{aligned}
$$

$\frac{1}{2\gamma}\left(\mu^T\tilde{\Sigma}^{-1}\mu - \frac{(\mathbf{1}_N^T\mu)^2}{\mathbf{1}_N^T\tilde{\Sigma}\mathbf{1}_N}\right)$ can be regard as the loss function of the 1/N portfolio when the variance-covariance matrix of assets returns is $\tilde{\Sigma}$. Unless the 1/N portfolio coincides with the optimal mean-variance portfolio with covariance matrix of $\tilde{\Sigma}$, $\frac{1}{2\gamma}\left(\mu^T\tilde{\Sigma}^{-1}\mu - \frac{(\mathbf{1}_N^T\mu)^2}{\mathbf{1}_N^T\tilde{\Sigma}\mathbf{1}_N}\right)$ is positive. Therefore,

$\frac{\partial^2 L(x^*,x^{ew})}{\partial \lambda^2} = \frac{6}{\lambda^4} \times \frac{1}{2\gamma} \left(\mu^T \tilde{\Sigma}^{-1} \mu - \frac{(\mathbf{1}_N^T \mu)^2}{\mathbf{1}_N^T \tilde{\Sigma} \mathbf{1}_N} \right) > 0.$

Proof of Proposition 3:

Denote the variance-covariance matrix of the asset returns as $\Sigma = \delta\tilde{\Sigma} + (1-\delta)\tilde{\Sigma}_d$, where $0 < \delta \leq 1$ controls for the level of the correlations, $\tilde{\Sigma}$ is a benchmark variance-covariance matrix of the assets, and $\tilde{\Sigma}_d$ is a N by N diagonal matrix whose elements come from $\tilde{\Sigma}$. This specification allows δ to adjust the correlation level while keep the volatility level of asset return unchanged. This setup can help us study the correlation effect on the performance of 1/N per se. Larger δ is, higher average correlation between asset return. To see how does the mean-variance utility loss of the 1/N rule changes according to the change in correlations, we can study the sign of $\frac{\partial L(x^*,x^{ew})}{\partial \delta}$.

$$
\begin{aligned}
\frac{\partial L(x^*, x^{ew})}{\partial \delta} =& \frac{\partial}{\partial \delta} \left[\frac{1}{2\gamma} \left(\mu^T \Sigma^{-1} \mu - \frac{(N\bar{\mu})^2}{\mathbf{1}_N^T \Sigma \mathbf{1}_N} \right) \right] \\
=& \frac{1}{2\gamma} \left[\mu^T \frac{d\Sigma^{-1}}{d\delta} \mu + \frac{(\mathbf{1}_N^T \mu)^2}{(\mathbf{1}_N^T \Sigma \mathbf{1}_N)^2} \mathbf{1}_N^T \frac{d\Sigma}{d\delta} \mathbf{1}_N \right] \\
=& \frac{1}{2\gamma} \left[-\mu^T \Sigma^{-1} \frac{d\Sigma}{d\delta} \Sigma^{-1} \mu + \left(\frac{\mathbf{1}_N^T \mu}{\mathbf{1}_N^T \Sigma \mathbf{1}_N} \right)^2 \mathbf{1}_N^T \frac{d\Sigma}{d\delta} \mathbf{1}_N \right] \\
=& \frac{1}{2\gamma} \left[-\mu^T \Sigma^{-1} (\tilde{\Sigma} - \tilde{\Sigma}_d) \Sigma^{-1} \mu + \left(\frac{\mathbf{1}_N^T \mu}{\mathbf{1}_N^T \Sigma \mathbf{1}_N} \right)^2 \mathbf{1}_N^T (\tilde{\Sigma} - \tilde{\Sigma}_d) \mathbf{1}_N \right] \\
=& \frac{\gamma}{2} \left[-\left(\frac{1}{\gamma} \Sigma^{-1} \mu \right)^T (\tilde{\Sigma} - \tilde{\Sigma}_d) \left(\frac{1}{\gamma} \Sigma^{-1} \mu \right) + \left(\frac{\mathbf{1}_N^T \mu \mathbf{1}_N}{\gamma \mathbf{1}_N^T \Sigma \mathbf{1}_N} \right)^T (\tilde{\Sigma} - \tilde{\Sigma}_d) \left(\frac{\mathbf{1}_N^T \mu \mathbf{1}_N}{\gamma \mathbf{1}_N^T \Sigma \mathbf{1}_N} \right) \right] \\
=& \frac{\gamma}{2} \left[-(x^*)^T (\tilde{\Sigma} - \tilde{\Sigma}_d) x^* + (x^{ew})^T (\tilde{\Sigma} - \tilde{\Sigma}_d) x^{ew} \right] \\
=& \frac{\gamma}{2\delta} \left[-(x^*)^T (\delta\tilde{\Sigma} - \delta\tilde{\Sigma}_d) x^* + (x^{ew})^T (\delta\tilde{\Sigma} - \delta\tilde{\Sigma}_d) x^{ew} \right] \\
=& \frac{\gamma}{2\delta} \left[-(x^*)^T (\Sigma - \tilde{\Sigma}_d) x^* + (x^{ew})^T (\Sigma - \tilde{\Sigma}_d) x^{ew} \right] \\
=& \frac{\gamma}{2\delta} \left[-(x^*)^T \Sigma x^* + (x^*)^T \tilde{\Sigma}_d x^* + (x^{ew})^T \Sigma x^{ew} - (x^{ew})^T \tilde{\Sigma}_d x^{ew} \right] \\
=& \frac{\gamma}{2\delta} \left[-\left(\sigma_*^2 - \sum_{i=1}^N \sigma_i^2 x_i^{*2} \right) + \left(\sigma_{ew}^2 - \sum_{i=1}^N \sigma_i^2 x_i^{ew2} \right) \right] \\
=& \frac{\gamma}{2\delta} \left[-\sum_{i \neq j} \sigma_{ij} x_i^* x_j^* + \sum_{i \neq j} \sigma_{ij} x_i^{ew} x_j^{ew} \right]
\end{aligned}
$$

where σ_*^2 and σ_{ew}^2 are variances of the mean-variance portfolio and the 1/N rule, respectively. σ_i^2

is the variance of the asset i, and σ_{ij} is the covariance between asset i and asset j. The portfolio weight in asset i for the mean-variance portfolio is denoted as x_i^*, and that for the 1/N rule is denoted as x_i^{ew}.

The mean-variance portfolio variance attributed to the covariance between asset returns is $\sum_{i\neq j} \sigma_{ij} x_i^* x_j^*$, and the 1/N rule variance attributed to the covariances is $\sum_{i\neq j} \sigma_{ij} x_i^{ew} x_j^{ew}$. If the variance attributed to the covariances between asset returns for mean-variance portfolio is larger (not larger) than that for the 1/N rule, then the change in mean-variance utility loss of the 1/N rule is a decreasing (non-decreasing) function in the proportional change in correlations.

We can further derive the second order derivative of the 1/N loss with respect to δ. Applying the 1/N rule in practice involves no consideration of expected return or covariance, so $x_i^{ew} = \frac{1}{N}$ which is independent of δ. Therefore, we have:

$$
\begin{aligned}
\frac{\partial^2 L(x^*, x^{ew})}{\partial \delta^2} =& \frac{\partial}{\partial \delta}\left[\frac{\partial L(x^*, x^{ew})}{\partial \delta}\right] \\
=& \frac{\partial}{\partial x^*}\left(\frac{\partial L}{\partial \delta}\right)\frac{\partial x^*}{\partial \delta} \\
=& \frac{\gamma}{2}\left[-2(x^*)^T(\tilde{\Sigma}-\tilde{\Sigma}_d)\right]\frac{\partial x^*}{\partial \delta} \\
=& \gamma\left[-(x^*)^T(\tilde{\Sigma}-\tilde{\Sigma}_d)\right]\left[-\frac{1}{\gamma}\Sigma^{-1}(\tilde{\Sigma}-\tilde{\Sigma}_d)\Sigma^{-1}\mu\right] \\
=& \gamma\left[-\frac{1}{\gamma}\mu^T\Sigma^{-1}(\tilde{\Sigma}-\tilde{\Sigma}_d)\right]\left[-\frac{1}{\gamma}\Sigma^{-1}(\tilde{\Sigma}-\tilde{\Sigma}_d)\Sigma^{-1}\mu\right] \\
=& \frac{1}{\gamma}\left[\mu^T\Sigma^{-1}(\tilde{\Sigma}-\tilde{\Sigma}_d)\Sigma^{-1}(\tilde{\Sigma}-\tilde{\Sigma}_d)\Sigma^{-1}\mu\right]
\end{aligned}
$$

As Σ^{-1} is a positive semidefinite matrix, for any 1 by N vector z, $z(\tilde{\Sigma}-\tilde{\Sigma}_d)z^T$ is non-negative. Let $z = \mu^T\Sigma^{-1}(\tilde{\Sigma}-\tilde{\Sigma}_d)$, so $\frac{\partial^2 L(x^*,x^{ew})}{\partial\delta^2} = z(\tilde{\Sigma}-\tilde{\Sigma}_d)z^T \geq 0$. Therefore, for a small δ, $\frac{\partial L(x^*,x^{ew})}{\partial\delta}$ is likely to be negative.

Proof of Proposition 4:

We have $L(x^*, x^{ew}) = \frac{1}{2\gamma}\left(S_*^2 - S_{ew}^2\right)$. If we assume the expected returns and the variance-covariance matrix of asset returns are fixed, then the Sharpe ratios of the mean-variance portfolio and the 1/N rule are independent of γ. Hence,

$$
\begin{aligned}
\frac{\partial L(x^*, x^{ew})}{\partial \gamma} =& \frac{\partial}{\partial \gamma}\left[\frac{1}{2\gamma}\left(\mu^T \Sigma^{-1} \mu - \frac{(N\bar{\mu})^2}{\mathbf{1}_N^T \Sigma \mathbf{1}_N}\right)\right] \\
=& \frac{\partial}{\partial \gamma}\left[\frac{1}{2\gamma}\left(S_*^2 - S_{ew}^2\right)\right] \\
=& -\frac{1}{2\gamma^2}\left(S_*^2 - S_{ew}^2\right) \\
=& -\frac{1}{\gamma} L(x^*, x^{ew})
\end{aligned}
$$

Unless the 1/N portfolio coincides with the mean-variance portfolio, the loss function of the 1/N is positive. Therefore, $\frac{\partial L(x^*, x^{ew})}{\partial \gamma} < 0$.

For the second derivative, $\frac{\partial^2 L(x^*, x^{ew})}{\partial \gamma^2} = \frac{\partial}{\partial \gamma}\left[-\frac{1}{2\gamma^2}\left(S_*^2 - S_{ew}^2\right)\right] = \frac{1}{\gamma^3}\left(S_*^2 - S_{ew}^2\right) > 0$

REFERENCES

Anderson, Evan W, and Ai-Ru Cheng, 2016, Robust bayesian portfolio choices, *Review of Financial Studies* 29, 1330–1375.

Behr, Patrick, Andre Guettler, and Felix Miebs, 2013, On portfolio optimization: Imposing the right constraints, *Journal of Banking & Finance* 37, 1232–1242.

Brown, Stephen J, Inchang Hwang, and Francis In, 2013, Why optimal diversification cannot outperform naive diversification: Evidence from tail risk exposure, *Available at SSRN* 2242694.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *The Journal of Finance* 52, 57–82.

Carhart, Mark M, Jennifer N Carpenter, Anthony W Lynch, and David K Musto, 2002, Mutual fund survivorship, *Review of Financial Studies* 15, 1439–1463.

Chan, Louis KC, Narasimhan Jegadeesh, and Josef Lakonishok, 1995, Evaluating the performance of value versus glamour stocks the impact of selection bias, *Journal of Financial Economics* 38, 269–296.

DeMiguel, Victor, Lorenzo Garlappi, Francisco J. Nogales, and Raman Uppal, 2009., A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, *Management Science* 55, 798–812.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2009, Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, *Review of Financial Studies* 22, 1915–1953.

Duchin, Ran, and Haim Levy, 2009, Markowitz versus the talmudic portfolio diversification strategies, *The Journal of Portfolio Management* 35, 71–74.

Fama, Eugene F, and Kenneth R French, 1992, The cross-section of expected stock returns, *The Journal of Finance* 47, 427–465.

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Frost, Peter A, and James E Savarino, 1986, An empirical bayes approach to efficient portfolio selection, *Journal of Financial and Quantitative Analysis* 21, 293–305.

Jagannathan, Ravi, and Tongshu Ma, 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *The Journal of Finance* 58, 1651–1684.

Jobson, J Dave, and Bob M Korkie, 1981, Performance hypothesis testing with the sharpe and treynor measures, *The Journal of Finance* 36, 889–908.

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

Kan, Raymond, Xiaolu Wang, and Guofu Zhou, 2016, Optimal portfolio selection with and without risk-free asset, *Available at SSRN 2729429* .

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

Kim, Dongcheol, 1997, A reexamination of firm size, book-to-market, and earnings price in the cross-section of expected stock returns, *Journal of Financial and Quantitative Analysis* 32, 463–489.

Kirby, Chris, and Barbara Ostdiek, 2012, Its all in the timing: simple active portfolio strategies that outperform naive diversification, *Journal of Financial and Quantitative Analysis* 47, 437–467.

Kirby, Chris, and Barbara Ostdiek, 2013, Optimizing the performance of sample mean-variance efficient portfolios, in *AFA 2013 San Diego Meetings Paper*.

Kothari, Sagar P, Jay Shanken, and Richard G Sloan, 1995, Another look at the crosssection of expected stock returns, *The Journal of Finance* 50, 185–224.

MacKinlay, A Craig, and Lubos Pastor, 2000, Asset pricing models: implications for expected returns and portfolio selection, *Review of Financial studies* 13, 883–916.

Markowitz, Harry, 1952, Portfolio selection, *The Journal of Finance* 7, 77–91.

Markowitz, Harry, 1959, Portfolio selection: Efficient diversification of investments, new york, john wiely & sons.

Pástor, L'uboš, and Robert F Stambaugh, 2000, Comparing asset pricing models: an investment perspective, *Journal of Financial Economics* 56, 335–381.

Pflug, Georg Ch, Alois Pichler, and David Wozabal, 2012, The 1/n investment strategy is optimal under high model ambiguity, *Journal of Banking & Finance* 36, 410–417.

Plyakha, Yuliya, Raman Uppal, and Grigory Vilkov, 2014, Equal or value weighting? implications for asset-pricing tests, *Available at SSRN* 1787045.

Politis, Dimitris N, and Joseph P Romano, 1994, The stationary bootstrap, *Journal of the American Statistical Association* 89, 1303–1313.

Shumway, Tyler, 1997, The delisting bias in crsp data, *The Journal of Finance* 52, 327–340.

Shumway, Tyler, and Vincent A Warther, 1999, The delisting bias in crsp's nasdaq data and its implications for the size effect, *The Journal of Finance* 54, 2361–2379.

Stambaugh, Robert F, 1997, Analyzing investments whose histories differ in length, *Journal of Financial Economics* 45, 285–331.

Tu, Jun, and Guofu Zhou, 2011, Markowitz meets talmud: A combination of sophisticated and naive diversification strategies, *Journal of Financial Economics* 99, 204–215.

Wang, Zhenyu, 2005, A shrinkage approach to model uncertainty and asset allocation, *Review of Financial Studies* 18, 673–705.

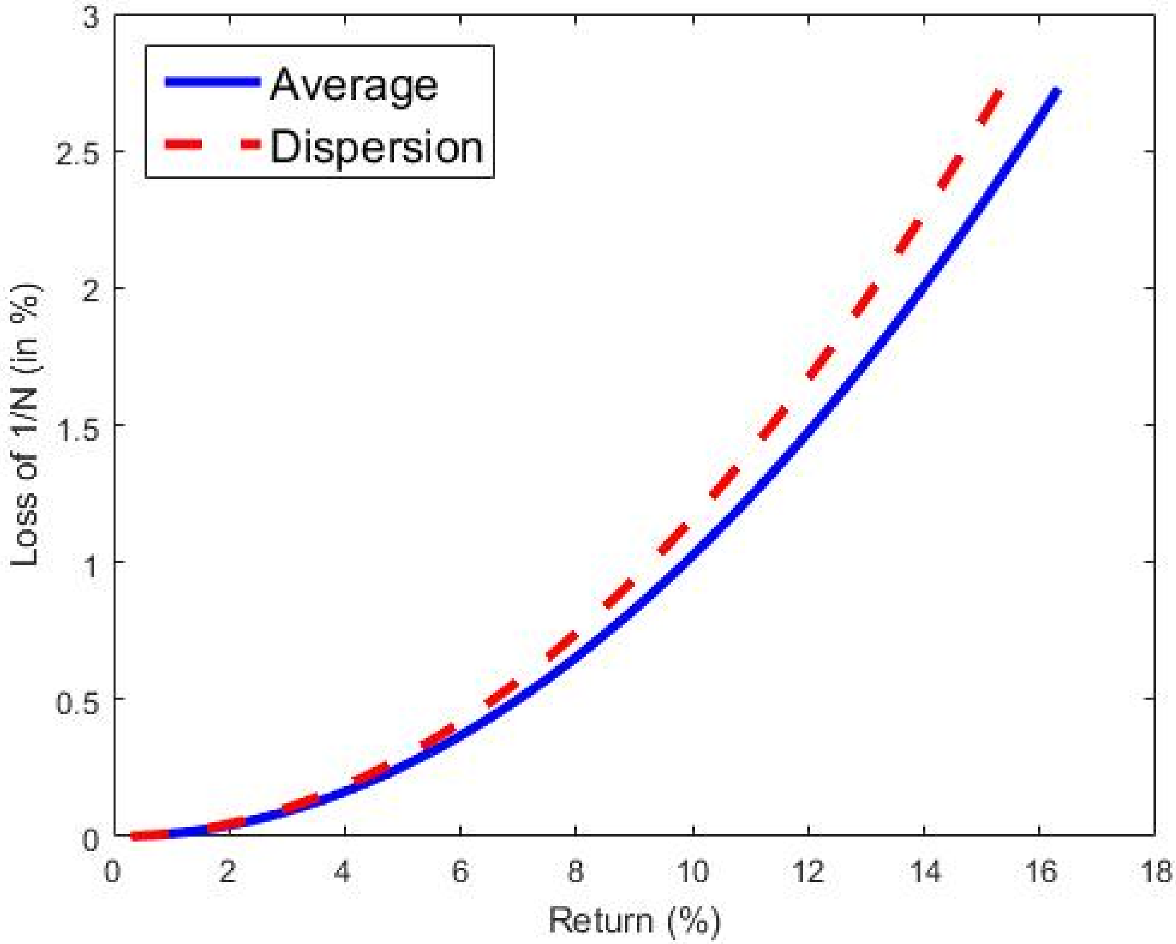

Figure 1. Relation between loss of the 1/N rule and returns. This figure plots the annualized mean-variance utility loss of the 1/N rule (in %) against average expected return and expected return dispersion of assets used to form portfolios. The blue solid line represents the annualized average expected return of constituent assets, while the red dashed line refers to the annualized expected return dispersion of the assets (the largest expected return minus the smallest expected return). The graph is generated by using the simulation setup in DeMiguel et al. (2009). The number of asset is assumed to be 25 and the coefficient of relative risk aversion is set to be one.

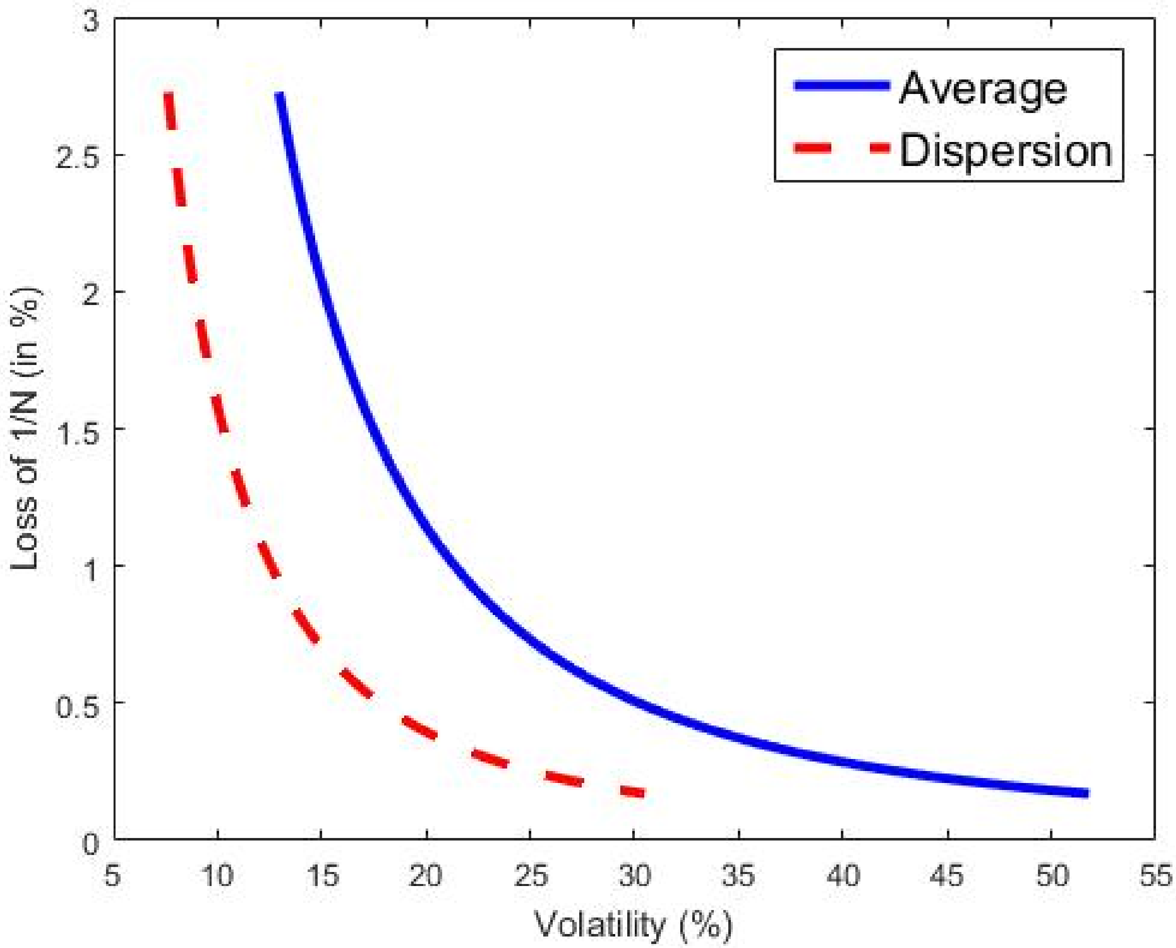

Figure 2. Relation between loss of the 1/N rule and volatilites. This figure plots the annualized mean-variance utility loss of the 1/N rule (in %) against average volatility and volatility dispersion of assets used to form portfolios. The blue solid line represents the annualized average volatility of constituent assets, while the red dashed line refers to the annualized volatility dispersion of the assets (the largest volatility minus the smallest volatility). The graph is generated by using the simulation setup in DeMiguel et al. (2009). The number of asset is assumed to be 25 and the coefficient of relative risk aversion is set to be one.

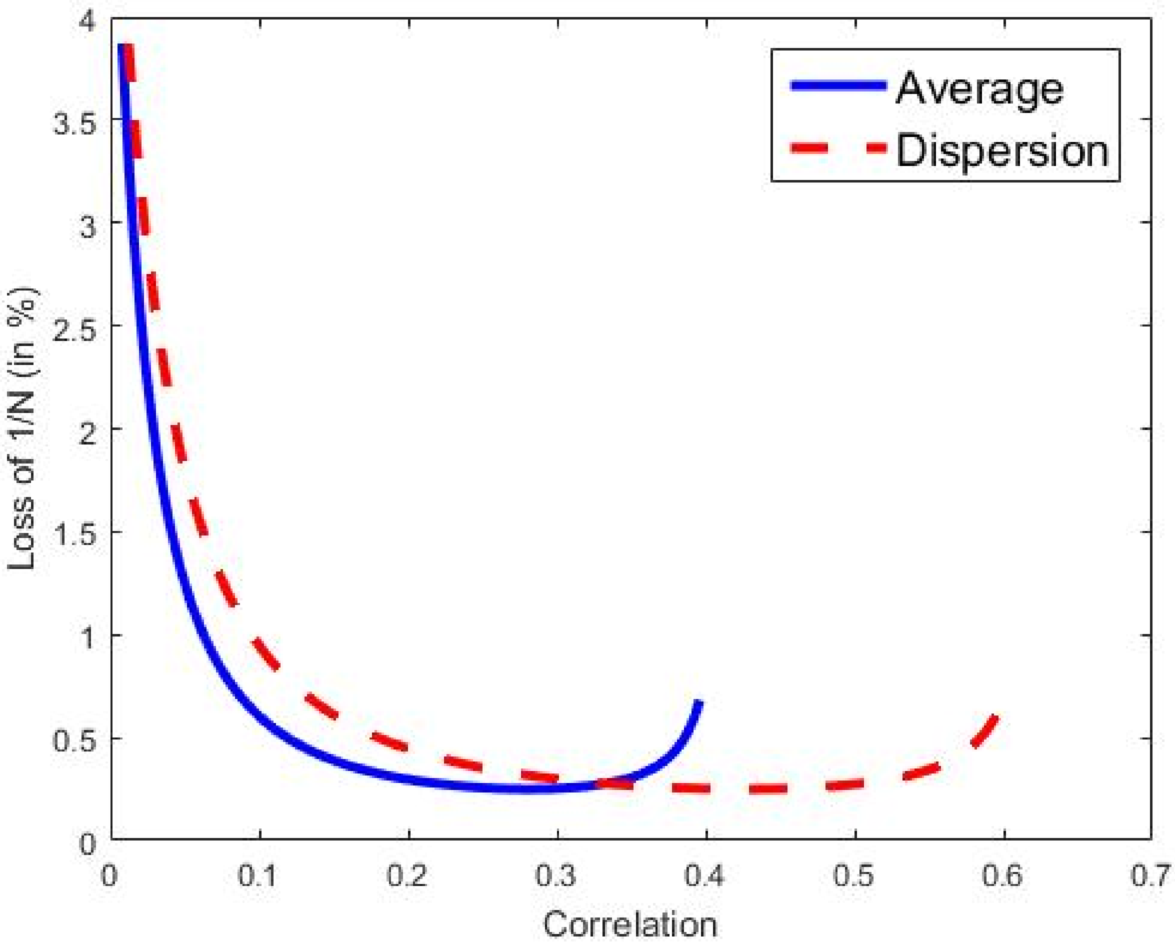

Figure 3. Relation between loss of the 1/N rule and correlations. This figure plots the annualized mean-variance utility loss of the 1/N rule (in %) against average correlation and correlation dispersion of assets used to form portfolios. The blue solid line represents the average correlation of constituent assets, while the red dashed line refers to the correlation dispersion of the assets (the largest correlation minus the smallest correlation). The graph is generated by using the simulation setup in DeMiguel et al. (2009). The number of asset is assumed to be 25 and the coefficient of relative risk aversion is set to be one.

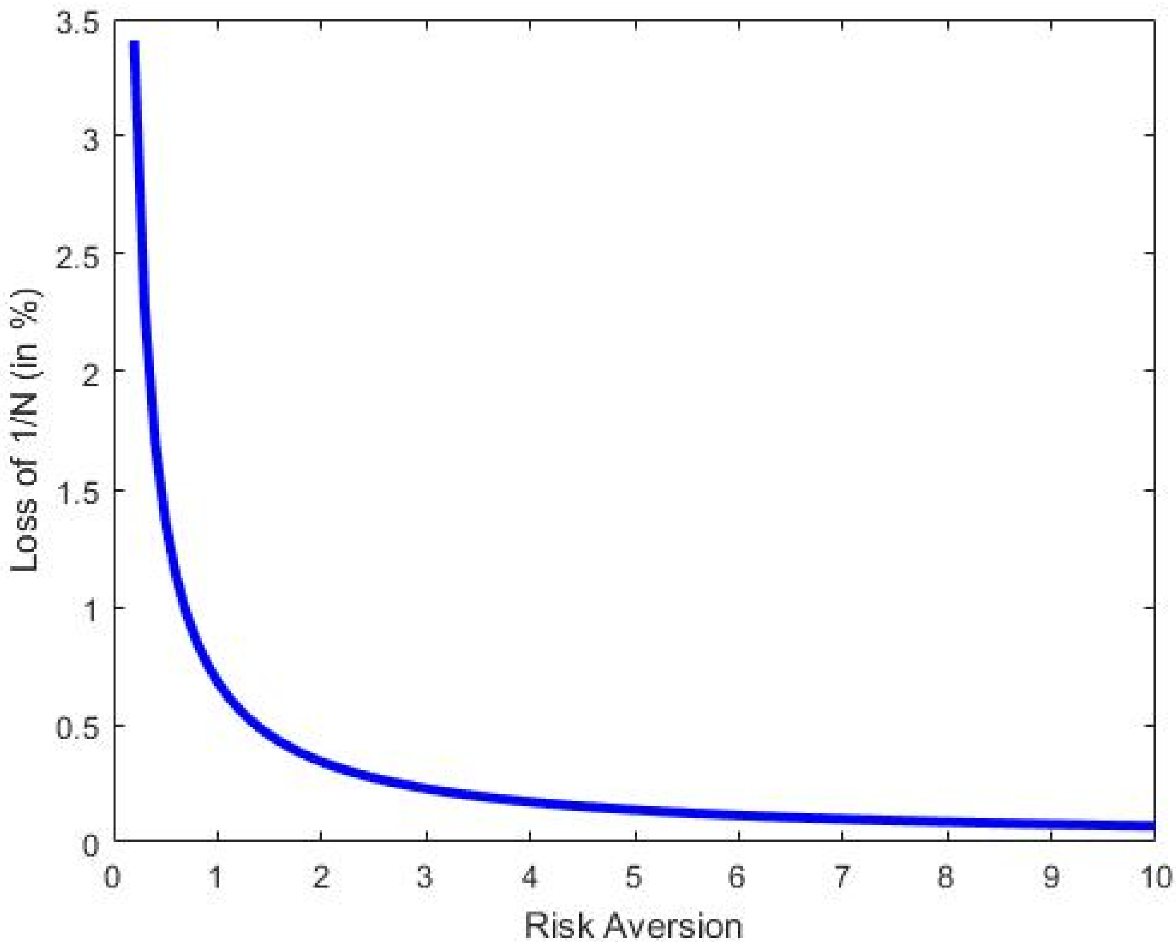

Figure 4. Relation between loss of the 1/N rule and coefficient of risk aversion. This figure plots the annualized mean-variance utility loss of the 1/N rule (in %) against the coefficient of risk aversion. The graph is generated by using the simulation setup in DeMiguel et al. (2009). The number of asset is assumed to be 25.

Table I Portfolio strategies considered

Strategy	Abbreviation
Benchmark Strategy	
Naive diversification	1/N or EW
Traditional Portfolios	
Sample mean-variance portfolio	MV
Tangency portfolio	TP
Minimum-variance portfolio	MIN
Combination Strategies by Tu and Zhou (2011)	
Combination of 1/N and sample mean-variance	CML
Combination of 1/N and Bayes-Stein mean-variance	CPJ
Combination of 1/N and "three-fund" model	CKZ
Optimal Risky Portfolio by Kan et al. (2016)	
Optimal portfolio without risk-free asset	QL
Timing Strategies by Kirby and Ostdiek (2012)	
Volatility timing strategy ($\eta = 1$)	VT1
Volatility timing strategy ($\eta = 2$)	VT2
Volatility timing strategy ($\eta = 4$)	VT4
Return-to-Risk timing strategy ($\eta = 1$)	RRT1
Return-to-Risk timing strategy ($\eta = 2$)	RRT2
Return-to-Risk timing strategy ($\eta = 4$)	RRT4
Beta-to-Risk timing strategy ($\eta = 1$)	BRT1
Beta-to-Risk timing strategy ($\eta = 2$)	BRT2
Beta-to-Risk timing strategy ($\eta = 4$)	BRT4

Table II Out-of-sample portfolio performance in Fama-French Dataset

This table summarizes the annualized out-of-sample portfolio performance using the dataset of Fama-French 25 size/book-to-market portfolios. The dataset consists of monthly portfolio return series spanning from July 1963 to December 2014. Each portfolio strategy is estimated every month based on a rolling estimation window of 120 months. μ and σ, both in %, are the annualized portfolio return and volatility, respectively. SR, U (in %) and T (in %) are the annualized portfolio Sharpe ratio, CEQ return/mean-variance utility ($\gamma = 2$) and turnover, respectively. p-SR (p-U) in parenthesis stands for the p-value for the null hypothesis that a portfolio strategy delivers the same Sharpe ratio (CEQ return) as the 1/N benchmark. The statistical inference is based on the small sample t test by Kirby and Ostdiek (2012) discussed in Section III.D.

Strategy	μ	σ	SR	p-SR	U	p-U	T
1/N	9.31	17.81	0.523	-	6.14	-	1.80
Traditional Portfolios							
MV	151.87	150.13	1.012	(0.71)	-73.51	(0.01)	6503.95
TP	-44.75	456.82	-0.098	(0.85)	-2131.55	(0.08)	10028.19
MIN	11.53	13.46	0.857	(0.01)	9.72	(0.11)	79.89
Combination Strategies							
CML	80.41	74.87	1.074	(0.37)	24.36	(0.09)	7265.16
CPJ	79.48	72.53	1.096	(0.34)	26.88	(0.05)	1339.44
CKZ	70.32	64.44	1.091	(0.28)	28.80	(0.01)	965.14
Optimal Risky Portfolio							
QL	45.13	51.84	0.871	(0.45)	18.26	(0.14)	805.67
Timing Strategies							
VT1	9.26	16.95	0.546	(0.12)	6.39	(0.36)	1.90
VT2	9.02	16.34	0.552	(0.28)	6.35	(0.67)	2.22
VT4	8.48	15.61	0.543	(0.64)	6.04	(0.89)	3.21
RRT1	10.20	17.57	0.580	(0.00)	7.11	(0.00)	4.07
RRT2	10.36	17.45	0.593	(0.01)	7.31	(0.01)	5.60
RRT4	10.41	17.40	0.598	(0.03)	7.38	(0.04)	8.80
BRT1	9.94	17.47	0.569	(0.00)	6.89	(0.01)	1.89
BRT2	10.26	17.29	0.593	(0.00)	7.27	(0.01)	2.18
BRT4	10.49	17.15	0.612	(0.01)	7.55	(0.01)	3.02

Table III Out-of-sample portfolio performance in sampled datasets of 10 stocks

The table summarizes the annualized out-of-sample portfolio performance based on 1000 datasets of 10 stocks randomly selected from NYSE/AMEX/NASDAQ. Each dataset consists of monthly return series of 20-year period randomly drawn from January 1926 to December 2014. Each portfolio strategy is estimated every month based on a rolling estimation window of 120 months, so the evaluation period for each dataset is 10 years. SR (SRc) is the average annualized Sharpe ratio without (with) transaction cost of 50bp. f-SR (f-SRc) is the frequency (in %) of outperformance over 1/N in terms of SR (SRc). fs-SR (fs-SRc) is the frequency (in %) of significant outperformance (at 10 % level) over 1/N in terms of SR (SRc). U (Uc) refers to the average annualized certainty-equivalent return/mean-variance utility (in %) without (with) transaction cost of 50bp, and the relative risk aversion is set to be two. f-U (f-Uc) is the frequency (in %) of outperformance over 1/N in terms of U (Uc). fs-U (fs-Uc) is the frequency (in %) of significant outperformance (at 10 % level) over 1/N in terms of U (Uc). T refers to the average annualized turnover (in %). In each dataset, statistical inference is based on the t test by Kirby and Ostdiek (2012) discussed in Section III.D, and aggregate performance is tested using the procedure in Section III.E. p-SR (p-SRc) stands for the p-value for the null hypothesis that a portfolio strategy delivers the same Sharpe ratio (Sharpe ratio with transaction cost) as the 1/N benchmark, and p-U (p-Uc) refers to the p-value for the null hypothesis that a portfolio strategy delivers the same CEQ return (CEQ return with transaction cost) as the 1/N benchmark.

Strategy	SR	p-SR	f-SR	fs-SR	U	p-U	f-U	fs-U	T	SRc	p-SRc	f-SRc	fs-SRc	Uc	p-Uc	f-Uc	fs-Uc
1/N	0.582	-	-	-	6.94	-	-	-	74.5	0.562	-	-	-	6.57	-	-	-
MV	0.178	0.00	9	0	-25	0.00	5	0	2271	0.008	0.00	3	0	-46	0.00	2	0
TP	0.173	0.42	11	0	-19920	0.00	8	0	14601	0.036	0.31	6	0	-18411	0.00	6	0
MIN	0.538	0.00	41	3	5.47	0.00	36	2	107.6	0.503	0.00	38	3	4.93	0.00	34	1
CML	0.449	0.00	15	0	4.53	0.00	21	1	460.0	0.366	0.00	7	0	2.13	0.00	11	0
CPJ	0.497	0.00	26	1	6.03	0.00	35	4	447.0	0.422	0.00	16	0	3.59	0.00	24	2
CKZ	0.470	0.00	25	0	5.60	0.00	34	4	425.0	0.394	0.00	15	0	3.42	0.00	22	2
QL	0.524	0.00	36	3	5.46	0.00	31	2	276.8	0.491	0.00	33	2	4.94	0.00	29	2
VT1	0.619	0.00	61	10	6.81	0.06	48	3	65.1	0.599	0.00	61	10	6.48	0.22	49	3
VT2	0.595	0.03	51	7	6.23	0.00	42	1	60.3	0.574	0.04	51	7	5.93	0.00	43	2
VT4	0.521	0.00	39	3	5.31	0.00	35	1	57.6	0.503	0.00	39	3	5.02	0.00	35	1
RRT1	0.535	0.00	37	5	5.48	0.00	27	2	82.6	0.511	0.00	36	5	5.07	0.00	27	2
RRT2	0.502	0.00	35	4	4.76	0.00	27	2	108.0	0.471	0.00	32	3	4.22	0.00	25	1
RRT4	0.446	0.00	29	1	3.74	0.00	25	1	151.5	0.406	0.00	25	1	2.98	0.00	22	1
BRT1	0.599	0.00	59	7	7.05	0.05	54	6	82.3	0.576	0.00	57	6	6.64	0.22	52	5
BRT2	0.594	0.01	56	6	6.90	0.63	51	4	94.4	0.567	0.30	54	5	6.43	0.10	49	4
BRT4	0.556	0.00	47	3	6.30	0.00	45	3	116.9	0.524	0.00	44	3	5.71	0.00	41	3

Table IV Out-of-sample portfolio performance in sampled datasets of 25 stocks

The table summarizes the annualized out-of-sample portfolio performance based on 1000 datasets of 25 stocks randomly selected from NYSE/AMEX/NASDAQ. Each dataset consists of monthly return series of 20-year period randomly drawn from January 1926 to December 2014. Each portfolio strategy is estimated every month based on a rolling estimation window of 120 months, so the evaluation period for each dataset is 10 years. SR (SRc) is the average annualized Sharpe ratio without (with) transaction cost of 50bp. f-SR (f-SRc) is the frequency (in %) of outperformance over 1/N in terms of SR (SRc). fs-SR (fs-SRc) is the frequency (in %) of significant outperformance (at 10 % level) over 1/N in terms of SR (SRc). U (Uc) refers to the average annualized certainty-equivalent return/mean-variance utility (in %) without (with) transaction cost of 50bp, and the relative risk aversion is set to be two. f-U (f-Uc) is the frequency (in %) of outperformance over 1/N in terms of U (Uc). fs-U (fs-Uc) is the frequency (in %) of significant outperformance (at 10 % level) over 1/N in terms of U (Uc). T refers to the average annualized turnover (in %). In each dataset, statistical inference is based on the t test by Kirby and Ostdiek (2012) discussed in Section III.D, and aggregate performance is tested using the procedure in Section III.E. p-SR (p-SRc) stands for the p-value for the null hypothesis that a portfolio strategy delivers the same Sharpe ratio (Sharpe ratio with transaction cost) as the 1/N benchmark, and p-U (p-Uc) refers to the p-value for the null hypothesis that a portfolio strategy delivers the same CEQ return (CEQ return with transaction cost) as the 1/N benchmark.

Strategy	SR	p-SR	f-SR	fs-SR	U	p-U	f-U	fs-U	T	SRc	p-SRc	f-SRc	fs-SRc	Uc	p-Uc	f-Uc	fs-Uc
1/N	0.625	-	-	-	7.32	-	-	-	75.4	0.603	-	-	-	6.94	-	-	-
MV	0.068	0.00	5	0	-90	0.00	1	0	11304	-0.323	0.00	0	0	-908	0.00	0	0
TP	0.072	0.68	6	0	-68956	0.07	7	0	47547	-0.162	0.79	2	0	-320354	0.07	2	0
MIN	0.519	0.00	32	2	4.72	0.00	23	1	227.3	0.435	0.00	25	1	3.58	0.00	18	1
CML	0.487	0.00	15	1	5.47	0.00	24	2	608.5	0.374	0.00	5	0	2.39	0.00	9	1
CPJ	0.493	0.00	21	0	6.12	0.00	31	5	732.1	0.370	0.00	8	0	2.41	0.00	17	2
CKZ	0.447	0.00	18	0	5.25	0.00	28	4	705.4	0.317	0.00	7	0	1.69	0.00	15	2
QL	0.592	0.00	40	6	5.94	0.00	30	2	509.4	0.545	0.00	35	4	5.29	0.00	25	1
VT1	0.659	0.00	61	13	6.88	0.00	39	2	66.3	0.636	0.00	61	13	6.55	0.00	40	2
VT2	0.649	0.00	53	10	6.29	0.00	33	2	63.5	0.624	0.00	53	9	5.97	0.00	34	2
VT4	0.563	0.00	38	3	5.31	0.00	27	1	67.2	0.538	0.00	38	3	4.97	0.00	27	1
RRT1	0.581	0.00	35	7	5.89	0.00	23	1	83.2	0.554	0.00	34	7	5.47	0.00	23	1
RRT2	0.556	0.00	34	6	5.27	0.00	25	1	110.9	0.520	0.00	32	5	4.72	0.00	23	1
RRT4	0.490	0.00	31	3	4.23	0.00	24	1	165.8	0.439	0.00	27	2	3.40	0.00	22	1
BRT1	0.640	0.00	66	8	7.28	0.41	55	5	83.5	0.614	0.00	62	7	6.87	0.06	54	4
BRT2	0.644	0.00	62	7	7.20	0.06	52	3	97.2	0.613	0.01	58	5	6.71	0.00	48	3
BRT4	0.624	0.84	52	5	6.85	0.00	45	2	127.5	0.583	0.00	47	3	6.22	0.00	41	1

Table V Out-of-sample portfolio performance in sampled datasets of 50 stocks

The table summarizes the annualized out-of-sample portfolio performance based on 1000 datasets of 50 stocks randomly selected from NYSE/AMEX/NASDAQ. Each dataset consists of monthly return series of 20-year period randomly drawn from January 1926 to December 2014. Each portfolio strategy is estimated every month based on a rolling estimation window of 120 months, so the evaluation period for each dataset is 10 years. SR (SRc) is the average annualized Sharpe ratio without (with) transaction cost of 50bp. f-SR (f-SRc) is the frequency (in %) of outperformance over 1/N in terms of SR (SRc). fs-SR (fs-SRc) is the frequency (in %) of significant outperformance (at 10 % level) over 1/N in terms of SR (SRc). U (Uc) refers to the average annualized certainty-equivalent return/mean-variance utility (in %) without (with) transaction cost of 50bp, and the relative risk aversion is set to be two. f-U (f-Uc) is the frequency (in %) of outperformance over 1/N in terms of U (Uc). fs-U (fs-Uc) is the frequency (in %) of significant outperformance (at 10 % level) over 1/N in terms of U (Uc). T refers to the average annualized turnover (in %). In each dataset, statistical inference is based on the t test by Kirby and Ostdiek (2012) discussed in Section III.D, and aggregate performance is tested using the procedure in Section III.E. p-SR (p-SRc) stands for the p-value for the null hypothesis that a portfolio strategy delivers the same Sharpe ratio (Sharpe ratio with transaction cost) as the 1/N benchmark, and p-U (p-Uc) refers to the p-value for the null hypothesis that a portfolio strategy delivers the same CEQ return (CEQ return with transaction cost) as the 1/N benchmark.

Strategy	SR	p-SR	f-SR	fs-SR	U	p-U	f-U	fs-U	T	SRc	p-SRc	f-SRc	fs-SRc	Uc	p-Uc	f-Uc	fs-Uc
1/N	0.653	-	-	-	7.66	-	-	-	76.9	0.630	-	-	-	7.28	-	-	-
MV	-0.016	0.00	3	0	-248	0.00	0	0	83870	-0.744	0.74	0	0	-645043	0.20	0	0
TP	0.007	0.88	6	0	-873532	0.09	5	0	129934	-0.360	0.83	1	0	-1652747	0.18	1	0
MIN	0.445	0.00	24	2	3.71	0.00	20	0	494.4	0.264	0.00	13	0	1.23	0.00	10	0
CML	0.531	0.00	15	0	6.44	0.00	25	1	710.0	0.390	0.00	4	0	2.11	0.00	6	0
CPJ	0.487	0.00	17	0	6.10	0.00	26	4	1034.5	0.303	0.00	4	0	0.83	0.00	10	0
CKZ	0.390	0.00	11	0	3.94	0.00	20	3	1107.6	0.176	0.00	2	0	-1.65	0.00	7	0
QL	0.645	0.07	46	9	6.45	0.00	29	1	814.5	0.582	0.00	36	5	5.61	0.00	21	1
VT1	0.691	0.00	60	17	7.14	0.00	35	2	67.2	0.666	0.00	60	17	6.80	0.00	36	2
VT2	0.696	0.00	52	14	6.53	0.00	31	1	64.7	0.670	0.00	52	14	6.20	0.00	31	1
VT4	0.614	0.00	41	5	5.50	0.00	25	0	71.7	0.584	0.00	40	4	5.14	0.00	25	0
RRT1	0.611	0.00	36	10	6.20	0.00	24	1	83.6	0.583	0.00	35	9	5.79	0.00	23	1
RRT2	0.594	0.00	33	10	5.59	0.00	23	1	111.9	0.556	0.00	32	9	5.04	0.00	22	1
RRT4	0.534	0.00	30	6	4.60	0.00	21	1	170.8	0.476	0.00	27	5	3.74	0.00	18	1
BRT1	0.668	0.00	66	10	7.58	0.01	53	5	84.6	0.640	0.00	63	8	7.16	0.00	51	4
BRT2	0.679	0.00	65	9	7.53	0.02	51	4	98.5	0.646	0.00	61	7	7.04	0.00	47	3
BRT4	0.677	0.00	59	7	7.31	0.00	45	3	132.4	0.631	0.91	52	4	6.65	0.00	39	2

Table VI Out-of-sample portfolio performance in sampled datasets of 10 portfolios

The table summarizes the annualized out-of-sample portfolio performance based on 1000 datasets of 10 portfolios randomly formed by using stocks from NYSE/AMEX/NASDAQ. Each dataset consists of monthly return series of 20-year period randomly drawn from January 1926 to December 2014. Each portfolio strategy is estimated every month based on a rolling estimation window of 120 months, so the evaluation period for each dataset is 10 years. SR (SRc) is the average annualized Sharpe ratio without (with) transaction cost of 50bp. f-SR (f-SRc) is the frequency (in %) of outperformance over 1/N in terms of SR (SRc). fs-SR (fs-SRc) is the frequency (in %) of significant outperformance (at 10 % level) over 1/N in terms of SR (SRc). U (Uc) refers to the average annualized certainty-equivalent return/mean-variance utility (in %) without (with) transaction cost of 50bp, and the relative risk aversion is set to be two. f-U (f-Uc) is the frequency (in %) of outperformance over 1/N in terms of U (Uc). fs-U (fs-Uc) is the frequency (in %) of significant outperformance (at 10 % level) over 1/N in terms of U (Uc). T refers to the average annualized turnover (in %). In each dataset, statistical inference is based on the t test by Kirby and Ostdiek (2012) discussed in Section III.D, and aggregate performance is tested using the procedure in Section III.E. p-SR (p-SRc) stands for the p-value for the null hypothesis that a portfolio strategy delivers the same Sharpe ratio (Sharpe ratio with transaction cost) as the 1/N benchmark, and p-U (p-Uc) refers to the p-value for the null hypothesis that a portfolio strategy delivers the same CEQ return (CEQ return with transaction cost) as the 1/N benchmark.

Strategy	SR	p-SR	f-SR	fs-SR	U	p-U	f-U	fs-U	T	SRc	p-SRc	f-SRc	fs-SRc	Uc	p-Uc	f-Uc	fs-Uc
1/N	0.682	-	-	-	7.93	-	-	-	14.1	0.677	-	-	-	7.86	-	-	-
MV	0.227	0.00	4	0	-21	0.00	6	1	11117	-0.610	0.00	0	0	-451	0.15	0	0
TP	0.233	0.21	6	0	-5673	0.01	9	0	54931	-0.400	0.38	0	0	-56246	0.03	0	0
MIN	0.630	0.00	38	2	6.82	0.00	35	2	551.5	0.447	0.00	8	0	4.06	0.00	7	0
CML	0.505	0.00	14	0	5.77	0.00	24	2	2279.1	0.070	0.00	0	0	-5.93	0.00	0	0
CPJ	0.552	0.00	17	0	7.39	0.01	33	4	1872.6	0.232	0.00	0	0	-2.24	0.00	3	0
CKZ	0.529	0.00	16	0	6.96	0.00	32	4	1892.6	0.186	0.00	1	0	-2.73	0.00	2	0
QL	0.678	0.00	40	4	7.74	0.00	34	2	1763.5	0.631	0.00	10	1	7.03	0.00	5	0
VT1	0.683	0.00	67	11	7.93	0.97	50	0	15.2	0.678	0.00	61	7	7.86	0.10	42	0
VT2	0.684	0.00	67	11	7.93	0.84	50	1	17.8	0.679	0.00	56	6	7.84	0.00	39	0
VT4	0.686	0.00	66	10	7.93	0.53	49	3	24.7	0.678	0.01	49	5	7.80	0.00	32	1
RRT1	0.681	0.00	40	6	7.89	0.00	26	0	25.4	0.673	0.00	34	5	7.77	0.00	21	0
RRT2	0.680	0.00	40	6	7.87	0.00	27	0	37.8	0.669	0.00	22	3	7.68	0.00	11	0
RRT4	0.679	0.00	40	6	7.82	0.00	28	1	62.4	0.661	0.00	16	1	7.51	0.00	7	0
BRT1	0.683	0.00	67	11	7.94	0.21	57	0	15.4	0.678	0.00	56	7	7.86	0.59	46	0
BRT2	0.683	0.00	66	11	7.94	0.02	57	0	18.6	0.678	0.40	49	4	7.85	0.00	37	0
BRT4	0.685	0.00	66	11	7.95	0.00	57	2	27.2	0.676	0.00	40	3	7.81	0.00	28	0

Table VII Out-of-sample portfolio performance in sampled datasets of 25 portfolios

The table summarizes the annualized out-of-sample portfolio performance based on 1000 datasets of 25 portfolios randomly formed by using stocks from NYSE/AMEX/NASDAQ. Each dataset consists of monthly return series of 20-year period randomly drawn from January 1926 to December 2014. Each portfolio strategy is estimated every month based on a rolling estimation window of 120 months, so the evaluation period for each dataset is 10 years. SR (SRc) is the average annualized Sharpe ratio without (with) transaction cost of 50bp. f-SR (f-SRc) is the frequency (in %) of outperformance over 1/N in terms of SR (SRc). fs-SR (fs-SRc) is the frequency (in %) of significant outperformance (at 10 % level) over 1/N in terms of SR (SRc). U (Uc) refers to the average annualized certainty-equivalent return/mean-variance utility (in %) without (with) transaction cost of 50bp, and the relative risk aversion is set to be two. f-U (f-Uc) is the frequency (in %) of outperformance over 1/N in terms of U (Uc). fs-U (fs-Uc) is the frequency (in %) of significant outperformance (at 10 % level) over 1/N in terms of U (Uc). T refers to the average annualized turnover (in %). In each dataset, statistical inference is based on the t test by Kirby and Ostdiek (2012) discussed in Section III.D, and aggregate performance is tested using the procedure in Section III.E. p-SR (p-SRc) stands for the p-value for the null hypothesis that a portfolio strategy delivers the same Sharpe ratio (Sharpe ratio with transaction cost) as the 1/N benchmark, and p-U (p-Uc) refers to the p-value for the null hypothesis that a portfolio strategy delivers the same CEQ return (CEQ return with transaction cost) as the 1/N benchmark.

Strategy	SR	p-SR	f-SR	fs-SR	U	p-U	f-U	fs-U	T	SRc	p-SRc	f-SRc	fs-SRc	Uc	p-Uc	f-Uc	fs-Uc
1/N	0.669	-	-	-	7.71	-	-	-	19.8	0.662	-	-	-	7.61	-	-	-
MV	0.055	0.00	3	0	-81	0.00	0	0	38981	-1.038	0.02	0	0	-13087	0.05	0	0
TP	0.066	0.84	4	0	-366757	0.08	6	0	210085	-0.650	0.88	0	0	-2760270	0.02	0	0
MIN	0.580	0.00	34	2	5.87	0.00	28	1	896.7	0.266	0.00	4	0	1.37	0.00	3	0
CML	0.506	0.00	13	0	5.83	0.00	21	1	2153.1	0.091	0.00	0	0	-5.32	0.00	0	0
CPJ	0.507	0.00	18	0	6.51	0.00	29	5	2423.6	0.088	0.00	0	0	-5.95	0.00	1	0
CKZ	0.461	0.00	15	0	5.60	0.00	26	4	2437.7	0.002	0.00	0	0	-6.81	0.00	1	0
QL	0.669	0.54	50	6	7.49	0.00	34	2	1981.0	0.621	0.00	14	0	6.76	0.00	6	0
VT1	0.672	0.00	72	14	7.71	0.88	49	1	21.5	0.665	0.00	68	11	7.60	0.04	43	0
VT2	0.675	0.00	71	15	7.71	0.88	49	2	25.3	0.667	0.00	64	10	7.58	0.00	39	2
VT4	0.680	0.00	70	14	7.70	0.39	48	3	36.0	0.669	0.00	59	8	7.52	0.00	35	1
RRT1	0.666	0.00	41	8	7.62	0.00	21	0	32.7	0.657	0.00	37	8	7.46	0.00	18	0
RRT2	0.666	0.00	41	8	7.57	0.00	22	1	50.5	0.650	0.00	26	5	7.31	0.00	11	1
RRT4	0.665	0.00	41	8	7.47	0.00	22	1	85.9	0.638	0.00	21	4	7.05	0.00	8	0
BRT1	0.670	0.00	77	14	7.72	0.01	59	0	21.9	0.664	0.00	69	9	7.61	0.84	49	0
BRT2	0.672	0.00	77	14	7.73	0.00	59	2	27.1	0.664	0.00	60	6	7.59	0.00	41	1
BRT4	0.676	0.00	75	14	7.74	0.00	58	5	40.7	0.663	0.43	49	4	7.54	0.00	32	1

Table VIII Out-of-sample portfolio performance in sampled datasets of 50 portfolios

The table summarizes the annualized out-of-sample portfolio performance based on 1000 datasets of 50 portfolios randomly formed by using stocks from NYSE/AMEX/NASDAQ. Each dataset consists of monthly return series of 20-year period randomly drawn from January 1926 to December 2014. Each portfolio strategy is estimated every month based on a rolling estimation window of 120 months, so the evaluation period for each dataset is 10 years. SR (SRc) is the average annualized Sharpe ratio without (with) transaction cost of 50bp. f-SR (f-SRc) is the frequency (in %) of outperformance over 1/N in terms of SR (SRc). fs-SR (fs-SRc) is the frequency (in %) of significant outperformance (at 10 % level) over 1/N in terms of SR (SRc). U (Uc) refers to the average annualized certainty-equivalent return/mean-variance utility (in %) without (with) transaction cost of 50bp, and the relative risk aversion is set to be two. f-U (f-Uc) is the frequency (in %) of outperformance over 1/N in terms of U (Uc). fs-U (fs-Uc) is the frequency (in %) of significant outperformance (at 10 % level) over 1/N in terms of U (Uc). T refers to the average annualized turnover (in %). In each dataset, statistical inference is based on the t test by Kirby and Ostdiek (2012) discussed in Section III.D, and aggregate performance is tested using the procedure in Section III.E. p-SR (p-SRc) stands for the p-value for the null hypothesis that a portfolio strategy delivers the same Sharpe ratio (Sharpe ratio with transaction cost) as the 1/N benchmark, and p-U (p-Uc) refers to the p-value for the null hypothesis that a portfolio strategy delivers the same CEQ return (CEQ return with transaction cost) as the 1/N benchmark.

Strategy	SR	p-SR	f-SR	fs-SR	U	p-U	f-U	fs-U	T	SRc	p-SRc	f-SRc	fs-SRc	Uc	p-Uc	f-Uc	fs-Uc
1/N	0.679	-	-	-	7.88	-	-	-	26.2	0.671	-	-	-	7.75	-	-	-
MV	-0.081	0.00	2	0	-240	0.00	0	0	240422	-1.081	0.90	0	0	-5264873	0.25	0	0
TP	-0.045	0.88	4	0	-850650	0.15	5	0	927379	-0.835	0.99	0	0	-937333424	0.30	0	0
MIN	0.507	0.00	28	2	4.92	0.00	25	1	1507.2	0.002	0.00	1	0	-2.63	0.00	1	0
CML	0.534	0.00	12	0	6.46	0.00	22	2	1920.5	0.143	0.00	0	0	-3.90	0.00	0	0
CPJ	0.473	0.00	15	0	5.87	0.00	26	3	2775.7	-0.028	0.00	0	0	-8.43	0.00	0	0
CKZ	0.364	0.00	9	0	3.44	0.00	19	2	3058.8	-0.230	0.00	0	0	-12.40	0.00	0	0
QL	0.683	0.00	52	8	7.65	0.00	35	2	2413.4	0.626	0.00	14	0	6.79	0.00	4	0
VT1	0.685	0.00	74	17	7.88	0.57	47	2	28.3	0.676	0.00	71	15	7.74	0.02	44	2
VT2	0.691	0.00	73	17	7.87	0.21	46	3	33.1	0.680	0.00	67	14	7.70	0.00	42	2
VT4	0.701	0.00	70	17	7.82	0.01	45	3	46.8	0.685	0.00	62	12	7.59	0.00	36	1
RRT1	0.674	0.00	34	11	7.72	0.00	19	1	40.1	0.662	0.00	31	11	7.52	0.00	18	1
RRT2	0.673	0.00	34	11	7.61	0.00	20	1	62.4	0.654	0.00	26	8	7.30	0.00	13	1
RRT4	0.671	0.00	37	11	7.43	0.00	22	1	106.5	0.638	0.00	24	6	6.89	0.00	11	0
BRT1	0.682	0.00	83	18	7.90	0.00	63	1	29.2	0.673	0.00	76	10	7.75	0.90	54	0
BRT2	0.685	0.00	82	18	7.91	0.00	62	5	36.4	0.674	0.00	68	7	7.73	0.00	47	2
BRT4	0.691	0.00	78	18	7.93	0.00	59	6	54.7	0.674	0.00	60	5	7.65	0.00	38	1

Table IX Predictability of out-of-sample portfolio performance based on return moments

The table reports the cross-sectional predictive regressions of out-of-sample portfolio performance against in-sample return moments. The dependent variable for each regression is the out-of-sample annualized CEQ return difference ($\gamma = 2$) between a sophisticated portfolio and the 1/N rule. These sophisticated strategies are the 13 "1/N outperformers" studied in this paper. The predictors include the number of assets (N), average expected return ($\bar{\mu}$), expected return dispersion ($\Delta\mu$), average volatility ($\bar{\sigma}$), volatility dispersion ($\Delta\sigma$), average pair-wise correlation ($\bar{\rho}$), correlation dispersion ($\Delta\rho$). Control varaibles are a number of tail risk measures, including average skewness ($\bar{S}$), skewness dispersion (ΔS), average kurtosis ($\bar{K}$), and kurtosis dispersion (ΔK). The sample is based on 6000 datasets used in this paper, and the p-value below each coefficient is calculated based on the White standard error.

Strategy	CML	CPJ	CKZ	QL	VT1	VT2	VT4	RRT1	RRT2	RRT4	BRT1	BRT2	BRT4
Const	-0.1186	-0.1501	-0.1551	-0.0602	0.0066	0.0028	-0.0061	-0.0295	-0.0426	-0.0569	0.0096	0.0084	-0.0028
	0.00	0.00	0.00	0.00	0.00	0.43	0.21	0.00	0.00	0.00	0.00	0.00	0.48
N	-0.0001	-0.0010	-0.0015	-0.0006	-0.0001	0.0000	0.0000	-0.0001	-0.0001	-0.0001	0.0000	0.0000	0.0000
	0.14	0.00	0.00	0.00	0.00	0.09	0.35	0.00	0.00	0.00	0.00	0.04	0.83
$\bar{\mu}$	0.1088	0.3132	0.3897	-0.0024	-0.0097	-0.0078	-0.0015	0.0343	0.0468	0.0657	-0.0132	-0.0149	-0.0128
	0.00	0.00	0.00	0.91	0.08	0.32	0.89	0.00	0.00	0.00	0.00	0.01	0.10
$\Delta\mu$	-0.0840	-0.0618	-0.0552	-0.1111	-0.0009	0.0011	0.0038	-0.0196	-0.0286	-0.0476	-0.0013	-0.0004	0.0010
	0.00	0.00	0.00	0.00	0.86	0.89	0.72	0.00	0.00	0.00	0.76	0.95	0.90
$\bar{\mu} \times \Delta\mu$	**0.1210**	**-0.0360**	**-0.1096**	**0.2586**	**0.0474**	**0.0489**	**0.0305**	**0.1226**	**0.1857**	**0.2794**	**0.0333**	**0.0301**	**0.0273**
	0.03	**0.65**	**0.17**	**0.00**	**0.10**	**0.23**	**0.56**	**0.00**	**0.00**	**0.00**	**0.13**	**0.29**	**0.48**
$\bar{\sigma}$	0.1618	0.1394	0.1032	0.0356	-0.0046	-0.0102	-0.0175	-0.0166	-0.0248	-0.0409	-0.0052	-0.0096	-0.0173
	0.00	0.00	0.00	0.03	0.38	0.21	0.11	0.00	0.00	0.00	0.18	0.09	0.04
$\Delta\sigma$	0.1555	0.2730	0.2714	0.1524	0.0168	0.0316	0.0429	0.0333	0.0595	0.0901	0.0023	0.0067	0.0131
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.36	0.08	0.03
$\bar{\sigma} \times \Delta\sigma$	**-0.2659**	**-0.4578**	**-0.4487**	**-0.2597**	**-0.0465**	**-0.0816**	**-0.1008**	**-0.0630**	**-0.1150**	**-0.1753**	**-0.0121**	**-0.0214**	**-0.0293**
	0.00	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**
$\bar{\rho}$	0.0579	0.0966	0.0970	0.0223	-0.0056	-0.0016	0.0086	0.0262	0.0382	0.0516	-0.0077	-0.0060	0.0064
	0.00	0.00	0.00	0.00	0.01	0.63	0.06	0.00	0.00	0.00	0.00	0.02	0.11
$\Delta\rho$	0.0081	-0.0027	0.0084	-0.0091	-0.0183	-0.0240	-0.0269	-0.0045	-0.0049	-0.0040	-0.0110	-0.0120	-0.0068
	0.32	0.80	0.42	0.33	0.00	0.00	0.00	0.18	0.31	0.55	0.00	0.00	0.19
$\bar{\rho} \times \Delta\rho$	**0.0820**	**0.2100**	**0.2076**	**0.1010**	**0.0530**	**0.0644**	**0.0533**	**0.0564**	**0.0769**	**0.0891**	**0.0322**	**0.0447**	**0.0462**
	0.00	**0.00**	**0.15**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	6000	6000	6000	6000	6000	6000	6000	6000	6000	6000	6000	6000	6000
$\bar{R}^2$	14.22	18.98	23.42	9.47	13.63	16.32	19.11	33.48	32.72	33.51	6.29	5.75	5.51

Table X Performance of the portfolio switching strategy

This table reports the annualized CEQ return (in %) for the portfolio switching strategy discussed in Section V.D. The strategy is a binary choice between a sophisticated portfolio and the 1/N portfolio based on the forecast from a cross-sectional predictive regression. The predictive regression is a "kitchen-sink" regression of the out-of-sample CEQ return difference between the sophisticated portfolio and 1/N against a number of in-sample return moments. The raw strategy is the sophisticated portfolio per se. p-(R-1/N) is the 1-tail p-value on the CEQ return difference between the raw strategy and the 1/N rule; p-(S-1/N) is the 1-tail p-value on the CEQ return difference between the switching strategy and the 1/N rule; and p-(S-R) is the 1-tail p-value on the CEQ return difference between the switching strategy and the raw strategy. P-values are calculated based on the procedure outlined in Section III.E. The sample is based on 6000 datasets described in III.A, and the relative risk aversion is set to be two for portfolio constructions and CEQ return calculations.

Portfolio	Raw Strategy	Switching Strategy	p-(R-1/N)	p-(S-1/N)	p-(S-R)
1/N	7.573	-	-	-	-
CML	5.751	7.932	1.00	0.00	0.00
CPJ	6.337	8.720	1.00	0.00	0.00
CKZ	5.132	8.560	1.00	0.00	0.00
QL	6.788	7.602	1.00	0.00	0.00
VT1	7.390	7.603	1.00	0.00	0.00
VT2	7.092	7.594	1.00	0.02	0.00
VT4	6.594	7.583	1.00	0.06	0.00
RRT1	6.801	7.591	1.00	0.00	0.00
RRT2	6.444	7.606	1.00	0.00	0.00
RRT4	5.881	7.610	1.00	0.00	0.00
BRT1	7.576	7.659	0.40	0.00	0.00
BRT2	7.534	7.669	0.97	0.00	0.00
BRT4	7.345	7.600	1.00	0.03	0.00