

# Being Naive about Naive Diversification: Can Investment Theory Be Consistently Useful?

Jun Tu and Guofu Zhou\*

*JEL classification:* G11; G12; C11

*Keywords:* Portfolio choice; parameter uncertainty; shrinkage; admissibility

First version: August, 2007

Current version: October, 2008

\*Jun Tu is from Singapore Management University, and Guofu Zhou from Washington University in St. Louis. We are grateful to Yacine Aït-Sahalia, Doron Avramov, Anil Bera, Henry Cao, Winghong Chan (the AFA-NFA discussant), Arnaud de Servigny, Victor DeMiguel, Frans de Roon (the EFA discussant), David Disatnik, Lorenzo Garlappi, Eric Ghysels, William Goetzmann, Bruce Hansen, Harrison Hong, Yongmiao Hong, Jing-zhi Huang (the SMUFSC discussant), Ravi Jagannathan, Raymond Kan, Hong Liu, Andrew Lo, Luboš Pástor, Eduardo Schwartz, Chu Zhang (the CICF discussant), seminar participants at Tsinghua University and Washington University, and participants at 2008 China International Conference in Finance, 2008 AsianFA-NFA International Conference, 2008 Singapore Management University Finance Summer Camp, 2008 European Finance Association Meetings, and Workshop on Advances in Portfolio Optimization at London Imperial College Business School for many helpful comments. We also thank Lynnea Brumbaugh-Walter for many helpful editorial suggestions. Tu acknowledges financial support for this project from Singapore Management University Research Grant C207/MSS6B006.

Corresponding Author: Guofu Zhou, Olin School of Business, Washington University, St. Louis, MO 63130. Phone: (314) 935-6384 and e-mail: zhou@wustl.edu

Being Naive about Naive Diversification:  
Can Investment Theory Be Consistently Useful?

The modern portfolio theory pioneered by Markowitz (1952) is widely used in practice and taught in MBA texts. DeMiguel, Garlappi and Uppal (2007), however, show that, due to estimation errors, existing theory-based portfolio strategies are not as good as we once thought, and the estimation window needed for them to outperform the naive  $1/N$  rule (that invests equally across  $N$  risky assets) is “around 3000 months for a portfolio with 25 assets and about 6000 months for a portfolio with 50 assets.” In this paper, based on an optimal combination of the  $1/N$  rule with the three-fund rule of Kan and Zhou (2007), we provide the first theory-based portfolio strategy that performs consistently well across various simulated data sets and real data sets, with estimation window as small as 120 months, while all others cannot do so and can even lose money on a risk-adjusted basis. Our results suggest that investment theory can be consistently useful in practice if properly applied.

ALTHOUGH MORE THAN HALF A CENTURY has passed since Markowitz's (1952) seminal work, the mean-variance framework is still the major model used in practice today in asset allocation and active portfolio management despite many sophisticated models developed by academics.<sup>1</sup> One of the main reasons is that many real-world issues, such as factor exposures and trading constraints, can be accommodated easily within this framework with analytical insights and fast numerical solutions. Another reason is that the intertemporal hedging demand is typically small. However, as is the case with any model, the true parameters in the mean-variance setup are unknown and have to be estimated from data, which introduces a parameter uncertainty problem since the estimated optimal portfolio rules are subject to random errors and can be substantially different from the true optimal rule. Brown (1976), Bawa, Brown, and Klein (1979), and Jorion (1986) are examples of earlier work that provide portfolio rules accounting for parameter uncertainty. Recently, Kan and Zhou (2007) compare the performances of various strategies including their three-fund rule that uses a third portfolio to hedge the estimation risk in the usual sample-based two-fund strategy.<sup>2</sup>

DeMiguel, Garlappi, and Uppal (2007), in their thought-provoking paper, find, however, that the parameter uncertainty problem can be so severe that existing sophisticated and estimated portfolio rules cannot even outperform the naive diversification strategy – the  $1/N$  rule that invests equally across  $N$  risky assets, even when the sample size is unrealistically large. In particular, they state in their paper that “Based on parameters calibrated to the U.S. equity market, our analytical results and simulations show that the estimation window needed for the sample-based mean-variance strategy and its extensions to outperform the  $1/N$  benchmark is around 3000 months for a portfolio with 25 assets and about 6000 months for a portfolio with 50 assets. This suggests that there are still many ‘miles to go’ before the gains promised by optimal portfolio choice can actually be realized out of sample.” Their finding challenges researchers to develop new methods for overcoming the estimation problem.<sup>3</sup>

---

<sup>1</sup>See Grinold and Kahn (1999), Litterman (2003) and Meucci (2005) for practical applications of the mean-variance framework; and see Brandt (2004) for an excellent survey of the academic literature.

<sup>2</sup>Pástor (2000), Pástor and Stambaugh (2000), Harvey, Liechty, Liechty, and Müller (2004), and Tu and Zhou (2004) are examples of recent Bayesian studies on the parameter uncertainty problem. We focus here on the classical framework and leave the search for suitable priors that work for general situations elsewhere.

<sup>3</sup>The finding also challenges the wisdom of the recently fast-growing 130-30 strategy (see, e.g., Lo and Patel (2008)), which, involving trillions of dollars, is one of Wall Street quantitative equity strategies based

Before addressing this challenge, we should point out first that it is inconsequential if the sample-based mean-variance strategy and other theory-based ones cannot outperform the  $1/N$  only in some special cases. This is because the  $1/N$  rule is the best one when the true optimal portfolio happens to be equal to it. In this case, it has a zero error from the optimal portfolio and cannot be improved any further, while *any* estimated rule must be subject to random errors with a positive variance, and therefore must perform worse than the  $1/N$ . Hence, in cases when the  $1/N$  is close to the true optimal portfolio as is the case in the exact one-factor model of DeMiguel, Garlappi, and Uppal (2007), it is expected that the estimated strategies will underperform the  $1/N$ . Thus, a random rule that only underperforms the  $1/N$  when the  $1/N$  is good by design is not sufficient to say that the rule is bad.

However, if the theory-based strategies are of value consistently across models and data sets, we would also expect that their performances should be close to that of the  $1/N$  when the  $1/N$  is set to be good, and better when the  $1/N$  is set to be poor. Unfortunately, this is not the case for the real data sets examined by DeMiguel, Garlappi, and Uppal (2007) as well as for our simulated data sets from various models. For example, for some of the data sets, *all* of the existing theory-based strategies (under our consideration) not only underperform the  $1/N$ , but also have negative risk-adjusted returns!<sup>4</sup> (The  $1/N$  fails to produce positive risk-adjusted returns sometimes too). Moreover, in a three-factor model even when the  $1/N$  is significantly different from the true optimal portfolio, we find that the theory-based strategies can still underperform the  $1/N$  substantially. That is, investors can be worse off by following the theory-based strategies than by simply putting 100% of the money into the riskless asset, due to presence of estimation errors. This raises the serious need for proposing new theory-based strategies that can perform well consistently across models and data sets.

To address this need, we, in this paper, propose a number of new theory-based portfolio strategies based on various assumptions on the data-generating process. All of the new rules have simple analytical expressions and can be estimated easily in practice. While it is likely that one strategy may be the best in some scenarios but not so in others, we do find that

---

almost entirely on the mean-variance portfolio theory (see, e.g., Chincarini and Kim (2006), and Qian, Hua, and Sorensen (2007)).

<sup>4</sup>In particular, the rules do not even work well for a few assets, raising questions for using the standard rule to allocate funds to a few asset classes by pensions and endowments.

an optimal combination of the  $1/N$  rule with the three-fund rule of Kan and Zhou (2007) performs consistently well across models and data sets. It performs as well as or better than all other theory-based rules on a consistent basis. It also outperforms substantially the  $1/N$  across almost all models: in a one-factor model with mispricing, in multiple factor models with and without mispricing, and in models calibrated from real data without any factor structures, even when the estimation window (sample size)  $T$  is as small as 120. For example, in a one-factor model with 25 assets and with pricing error alphas ranging from  $-5\%$  to  $5\%$  per year, it achieves average expected utility  $5.81\%$ ,  $7.44\%$ ,  $10.02\%$ , and  $12.99\%$  per year, respectively, while the  $1/N$  rule has a constant level of  $3.89\%$  per year, as  $T$  goes up from 120 months to 240, 480, and 960 months. In a model calibrated with Fama and French's (1993) 25 assets without any factor structures, its utility values are  $12.99\%$ ,  $21.53\%$ ,  $30.74\%$ , and  $37.49\%$  per year, in contrast to a much smaller value of  $4.28\%$  per year for the  $1/N$ . Moreover, it is the only rule that never loses money (on a risk-adjusted basis) across models and data sets.

Why does the combination rule perform so well? First, combination always helps. To see the intuition, imagine that, for some data sets, one rule is good and the other rule is bad, and for other data sets, the reverse happens. Then, the combination obtains the average performance. For a concave utility, the average performance is preferred to both good and bad. This also parallels to diversification with two risky assets. A suitable combination or an optimal portfolio of them is always preferred to holding either one alone. Second, the combination rule here takes advantage of the good properties of both the  $1/N$  and the three-fund rules. Economically, in the CAPM world with the sum of the betas being equal to one, the  $1/N$  rule holds the market portfolio plus an average noise across assets. Statistically, the  $1/N$  is an excellent shrinkage point for improving the estimation of the mean of a multivariate distribution. On the other hand, the  $1/N$  is biased since it will never converge to the true optimal rule unless it happens to be equal to it. But it is a rule with zero variance, and hence it contributes the utility loss from using it only by its bias. The three-fund rule of Kan and Zhou (2007) is designed to diversify the estimation risk with the use of two sample frontier portfolios. It is asymptotically unbiased, but can have sizable variance in small samples. Therefore, the combination of the  $1/N$  rule with the three-fund

rule is to make the optimal tradeoff between adding bias and reducing variance. When the sample size is small, the variance of the three-fund rule is large. Increasing the weight on the  $1/N$  in the combination will increase the bias, but decrease the variance. Thus, the optimal sample-dependent weight should make the combination better than using either the  $1/N$  or the three-fund rule alone. Clearly, though, as the sample size goes up, more weight will be placed on the three-fund rule. With an infinite amount of data, the weight will eventually go to one, and the combination rule will converge to the true optimal portfolio.

The central question of this paper is whether investment theory can be consistently useful.<sup>5</sup> Our proposed optimal combination rule is theory-based and performs consistently well across all models and real data sets under our study, with sample sizes of only 120 and 240 months, far less than the incredible sample sizes of “around 3000 months for a portfolio with 25 assets and about 6000 months for a portfolio with 50 assets.” Our results therefore support firmly the proposition of our paper that investment theory can be consistently useful for practical sample sizes, despite of parameter uncertainty, as long as it is applied in the way like the new strategy.

The remainder of the paper is organized as follows. Section I provides the various new estimators of the true but unknown optimal portfolio rule. Section II compares the performance of the  $1/N$  with rules proposed here and some of the existing ones. Section III discusses directions for future research. Section IV concludes.

## I. Portfolio Strategies Under Parameter Uncertainty

In this section, we review first the mean-variance framework, then introduce the combination or shrinkage rules and a rule based on the assumption of factor model structure, and finally present two new three- and four-fund strategies.

### *A. The Portfolio Choice Problem*

Consider the standard portfolio choice problem in which an investor chooses his optimal

---

<sup>5</sup>This question is related but different from the question whether investment theory can outperform the  $1/N$ , which, as explained earlier, is impossible in some specific scenarios. As a matter of fact, this is true for any fixed constant rule that is independent of data.

portfolio among  $N$  risky assets and a riskless asset. Let  $r_{ft}$  and  $r_t$  be, respectively, the rates of returns on the riskless asset and the  $N$  risky assets at time  $t$ . We define  $R_t \equiv r_t - r_{ft}1_N$  as the excess returns, i.e., the returns in excess of the riskless asset, where  $1_N$  is an  $N$ -vector of ones. Note that allowing for the riskless asset is not only practical in asset allocation problems, but also meaningful to fund managers. If a fund is restricted to equity only, the returns on utility companies should be a close proxy of the riskless asset. Since the performances of most institutional managers are benchmarked by an index, say the S&P500, the S&P500 index portfolio is the riskless asset and the returns in excess of it are what matter in their investment decisions. In this case, mathematically, the return on the S&P500 plays the role of  $r_{ft}$  below, and the framework developed here applies without any problems.<sup>6</sup>

For the probability distribution of  $R_t$ , we make the common assumption that  $R_t$  is independent and identically distributed over time, and has a multivariate normal distribution with mean  $\mu$  and covariance matrix  $\Sigma$ . To obtain analytical solutions, we focus our analysis on the standard mean-variance framework. In this framework, the investor at time  $T$  chooses his portfolio weights  $w$  so as to maximize the quadratic objective function

$$U(w) = E[R_p] - \frac{\gamma}{2} \text{Var}[R_p] = w'\mu - \frac{\gamma}{2}w'\Sigma w, \quad (1)$$

where  $R_p = w'R_{T+1}$  is the future uncertain portfolio return and  $\gamma$  is the coefficient of relative risk aversion. It is well-known that, when both  $\mu$  and  $\Sigma$  are assumed known, the portfolio weights are

$$w^* = \frac{1}{\gamma}\Sigma^{-1}\mu, \quad (2)$$

and the maximized expected utility is

$$U(w^*) = \frac{1}{2\gamma}\mu'\Sigma^{-1}\mu = \frac{\theta^2}{2\gamma}, \quad (3)$$

where  $\theta^2 = \mu'\Sigma^{-1}\mu$  is the squared Sharpe ratio of the *ex ante* tangency portfolio of the risky assets.

However,  $w^*$  is not computable in practice because  $\mu$  and  $\Sigma$  are unknown. To implement the above mean-variance theory of Markowitz (1952), the optimal portfolio weights are usually estimated by using a two-step procedure. First, the mean and covariance matrix of the

---

<sup>6</sup>See, e.g., Grinold and Kahn (1999) for active portfolio management with benchmarks.

asset returns are estimated based on the observed data. The standard estimates are

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T R_t, \quad (4)$$

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T (R_t - \hat{\mu})(R_t - \hat{\mu})', \quad (5)$$

which are the maximum likelihood (ML) estimator. Second, these sample estimates are then treated as if they were the true parameters, and are simply plugged into (2) to compute the popular ML estimator of the optimal portfolio weights,

$$\hat{w}^{\text{ML}} = \frac{1}{\gamma} \hat{\Sigma}^{-1} \hat{\mu}. \quad (6)$$

Since  $\hat{w}^{\text{ML}}$  is a random variable that is distributed around  $w^*$  at most, this gives rise to a parameter uncertainty problem because the utility associated with using  $\hat{w}^{\text{ML}}$  is different from  $U(w^*)$  due to using the estimated rule rather than the true one. As shown by Brown (1976), Bawa, Brown, and Klein (1979), Jorion (1986) and Kan and Zhou (2007), the difference can be quite substantial in realistic applications.

Since the true portfolio weights  $w^*$  are unknown, the task is how to best estimate them based on available observations  $R_1, \dots, R_T$ . Any estimator must be a function of the data; say  $\tilde{w} = \tilde{w}(R_1, \dots, R_T)$  is such an estimator. The classical criterion for its performance is the *expected loss function*

$$L(w^*, \tilde{w}) = U(w^*) - E[\tilde{U}(\tilde{w})], \quad \tilde{U}(\tilde{w}) \equiv \tilde{w}'\mu - \frac{\gamma}{2}\tilde{w}'\Sigma\tilde{w}, \quad (7)$$

where  $U(w^*)$  is the expected utility of knowing the true parameters, and  $E[\tilde{U}(\tilde{w})]$ , as nicely put by DeMiguel, Garlappi, and Uppal (2007), is the average utility realized by an investor who plays the estimated strategy  $\tilde{w}$  infinitely many times. One can also imagine that playing the same strategy works in different markets, such as the US and other countries. Brown (1976), Jorion (1986), Frost and Savarino (1986), Stambaugh (1997), TerHorst, DeRoos, and Werkerzx (2002), Kan and Zhou (2007), and DeMiguel, Garlappi, and Uppal (2007) are examples of using  $L(w^*, \tilde{w})$  to evaluate portfolio rules. In practice, even though there is a long time series of data in the US equity market, the utilities from simulated data based on similar sample lengths can still be substantially smaller than the true hypothetical utility



(see, e.g., Section II). Hence, parameter uncertainty is an important issue in practice (see, e.g., Meucci, 2005).

For any portfolio rule, we note first that the loss can be written as

$$\begin{aligned}
L(w^*, \tilde{w}) &= \frac{\gamma}{2} \left[ \frac{1}{\gamma^2} \mu' \Sigma^{-1} \mu - \frac{2}{\gamma} \mu' [E(\tilde{w})] + E[\tilde{w}' \Sigma \tilde{w}] \right] \\
&= \frac{\gamma}{2} E \left[ \left( \frac{1}{\gamma} \Sigma^{-1} \mu - \tilde{w} \right)' \Sigma \left( \frac{1}{\gamma} \Sigma^{-1} \mu - \tilde{w} \right)' \right] \\
&= \frac{\gamma}{2} E [(\tilde{w} - w^*)' \Sigma (\tilde{w} - w^*)], \tag{8}
\end{aligned}$$

i.e., a quadratic function of the errors in estimating  $w^*$ . In contrasting this with the usual statistical optimal estimation, there are two differences. First, it is a function of the primitive parameters of the data-generating process that is of interest, not the parameters themselves. Second, the weighting matrix,  $\Sigma$ , is unknown. These differences make a simple and analytical solution to the best possible estimator of  $w$  impossible, as will be clear from the analysis below.

### B. Optimal Combinations

The naive  $1/N$  rule is a special estimator of  $w^*$  that ignores all data information, and can be expressed as

$$w_e \equiv c_e \mathbf{1}_N, \tag{9}$$

where  $c_e$  is a scalar determining the total allocation to risky assets per dollar. The simple naive diversification  $1/N$  rule takes  $c_e = 1/N$ , and so  $w_e = \mathbf{1}_N/N$ , which invests  $1/N$  of each dollar into each of the  $N$  risky assets. In general,  $w_e$  allocates funds equally among the  $N$  risky assets with the total allocation equal to  $Nc_e$ , and it allocates the rest,  $1 - Nc_e$ , into the riskless asset. Since DeMiguel, Garlappi, and Uppal (2007) focus their studies on the naive  $1/N$  rule, we will also do so in what follows.

Although the naive  $1/N$  rule is quite simple, DeMiguel, Garlappi, and Uppal (2007) show that it can perform remarkably well under certain conditions. Indeed, when the assets returns have equal means and variances and when they are independent,  $1/N$  is the best one with suitable risk aversion adjustment. As is well known in statistics (see, e.g., Lehmann, E. L., and George Casella, 1998),  $1/N$  is the common choice of a good shrinkage point for

improving the estimation of the mean of a multivariate distribution.

Another economic reason, which seems not recognized in the literature, is that it is proportional to the market portfolio in a one-factor model with the market as the factor. To see this, consider the standard market model regression,

$$R_{jt} = \alpha_j + \beta_j R_{mt} + \epsilon_{jt}, \quad j = 1, 2, \dots, N, \quad (10)$$

where  $R_{jt}$  is the excess return on the  $j$ -th asset,  $R_{mt}$  is the market excess return, and  $\epsilon_{jt}$ 's are the regression residuals with a covariance matrix  $\Sigma_\epsilon$ . Averaging the model over  $j$ , we obtain the following portfolio of the  $1/N$  rule,

$$w'_e R_t = \bar{\alpha} + \bar{\beta} R_{mt} + \bar{\epsilon}, \quad (11)$$

where  $\bar{\alpha}$ ,  $\bar{\beta}$  and  $\bar{\epsilon}$  are the average alpha, beta and residuals, respectively. When the CAPM is true,  $\bar{\alpha}$  must be zero. In this case,  $w'_e R_t$  is proportional to the market portfolio except the average residual term which can be much smaller than the individual ones. If  $\bar{\beta}$  is also one,  $w'_e R_t$  deviates from the market portfolio only by  $\bar{\epsilon}$ . Empirically, the CAPM is clearly a starting point for understanding the data.<sup>7</sup> Given that the market portfolio is not easy to beat, the  $1/N$  will be a good starting portfolio as well.

However, there is an important problem with the  $1/N$  rule. It makes no use of sample information, and will always fail to converge to the true optimal rule when it does not happen to be equal to it. If it has a large difference from the true optimal rule, especially when the data-generating process is complex and when the market portfolio is far from efficiency, its performance must be poor.

To improve the  $1/N$  rule with sample information, consider the following combination of the  $1/N$  with an unbiased ML estimator of  $w^*$ ,

$$\hat{w}_s = (1 - \delta)w_e + \delta\bar{w}, \quad (12)$$

where

$$\bar{w} = \frac{1}{\gamma} \tilde{\Sigma}^{-1} \hat{\mu} \quad (13)$$

---

<sup>7</sup>Black and Litterman's (1992) popular asset allocation model starts from the CAPM and updates it with an investor's views. But their analysis is ad hoc Bayesian without using the data-generating process.

is simply a scale adjustment of  $\hat{w}^{\text{ML}}$  so that it is unbiased to satisfy  $E\bar{w} = w^*$ , and  $\tilde{\Sigma} = T\hat{\Sigma}/(T - N - 2)$ .  $\delta$  is the combination parameter,  $0 \leq \delta \leq 1$ . Intuitively, an optimal combination of  $w_e$  and  $\bar{w}$  should be at least as good as any of them used alone. Since  $w_e$  is constant, its loss will remain the same even if we have an infinite amount of data. On the other hand,  $\bar{w}$  performs well when the available sample size is large enough. Hence, a combination of  $w_e$  with  $\bar{w}$  can make use of the sample information to pin down where the true rule is, and in so doing improves the  $1/N$ . The combination is also known as a shrinkage estimator in statistics, which shrinks the  $1/N$  rule toward the true rule.

Formally, because of (8) and  $E\bar{w} = w^*$ , the expected loss associated with using  $\hat{w}_s$  is

$$\begin{aligned} L(w^*, \hat{w}_s) &= \frac{\gamma}{2} [(1 - \delta)^2 (w_e - w^*)' \Sigma (w_e - w^*) + \delta^2 E((\bar{w} - w^*)' \Sigma (\bar{w} - w^*))] \\ &= \frac{\gamma}{2} [(1 - \delta)^2 \pi_1 + \delta^2 \pi_2], \end{aligned} \quad (14)$$

where

$$\pi_1 = w_e' \Sigma w_e - \frac{2}{\gamma} w_e' \mu + \frac{1}{\gamma^2} \theta^2, \quad (15)$$

$$\pi_2 = \frac{1}{\gamma^2} (c_1 - 1) \theta^2 + \frac{c_1 N}{\gamma^2 T} \quad (16)$$

with  $\theta^2 = \mu' \Sigma^{-1} \mu$  and

$$c_1 = \frac{(T - 2)(T - N - 2)}{(T - N - 1)(T - N - 4)}. \quad (17)$$

Equation (15) is trivial, and equation (16) follows from both equation (30) of Kan and Zhou (2007) and equation (14) here. Equation (14) is quite intuitive. The  $1/N$  rule is an estimator of  $w$  with bias  $\pi_1$ , but zero variance, while  $\bar{w}$  is unbiased, but with nonzero variance  $\pi_2$ . Therefore, the loss depends on  $\delta$ , which determines the tradeoff between bias and variance. If the bias is large, the  $1/N$  should be weighted less and vice versa.

Interestingly, as long as  $w_e$  is not exactly equal to  $w^*$ ,  $\delta$  can always be chosen to be a number small enough to make the loss of the combination rule smaller than it would be using the  $1/N$  rule alone. Summarizing this, we have

**Proposition 1:** *If  $0 < \delta < 2\pi_1/(\pi_1 + \pi_2)$ , then the combination estimator  $\hat{w}_s$  has a strictly smaller loss than the  $1/N$  rule.*

Proposition 1 (proofs of all propositions are given in the appendix) says that the  $1/N$  can be dominated by the combination estimator as long as the true  $w^*$  lies outside any given neighborhood of  $1/N$ . For example, in an application in which we are confident that  $w_e$  must have at least some bias so that  $\pi_1 > a_1$ , a given positive constant, and if the weighted variance of  $\bar{w}$ , as measured by  $\pi_2$ , is less than  $a_2$ , another given positive constant, then any positive  $\delta$  that is less than  $2a_1/(a_1 + a_2)$  will always make the combination estimator to have a smaller loss than the  $1/N$  rule.<sup>8</sup>

However, improving upon  $1/N$  is not the goal. What we need is a good rule that can perform well across models. For this purpose, we optimize  $\delta$  in equation (14) to get a new rule. It is clear that the optimal choice of  $\delta$  is

$$\delta^* = \frac{\pi_1}{\pi_1 + \pi_2}, \quad (18)$$

the midpoint of the bound given by Proposition 1. But this value is unknown, and has to be estimated. Given the data,  $\pi_1$  and  $\pi_2$  can be estimated by

$$\hat{\pi}_1 = w_e' \hat{\Sigma} w_e - \frac{2}{\gamma} w_e' \hat{\mu} + \frac{1}{\gamma^2} \tilde{\theta}^2, \quad (19)$$

$$\hat{\pi}_2 = \frac{1}{\gamma^2} (c_1 - 1) \tilde{\theta}^2 + \frac{c_1}{\gamma^2} \frac{N}{T}, \quad (20)$$

where  $\tilde{\theta}^2$  is an accurate estimator  $\theta^2$ , proposed by Kan and Zhou (2007) and given by

$$\tilde{\theta}^2 = \frac{(T - N - 2)\hat{\theta}^2 - N}{T} + \frac{2(\hat{\theta}^2)^{\frac{N}{2}}(1 + \hat{\theta}^2)^{-\frac{T-2}{2}}}{TB_{\hat{\theta}^2/(1+\hat{\theta}^2)}(N/2, (T-N)/2)}, \quad (21)$$

where  $\hat{\theta}^2 = \hat{\mu}' \hat{\Sigma}^{-1} \hat{\mu}$  and

$$B_x(a, b) = \int_0^x y^{a-1} (1-y)^{b-1} dy \quad (22)$$

is the incomplete beta function. Then, we obtain  $\hat{\delta}$ , an estimator of  $\delta^*$ , by plugging  $\hat{\pi}_1$  and  $\hat{\pi}_2$  into (18). This will give us a completely new rule. We summarize the result as

**Proposition 2:** *Among the combination rules  $\hat{w}_s = (1 - \delta)w_e + \delta\bar{w}$ , the estimated optimal one is*

$$\hat{w}^{\text{CML}} = (1 - \hat{\delta})w_e + \hat{\delta}\bar{w}, \quad (23)$$

---

<sup>8</sup>Proposition 1 can be extended to any fixed constant rule.

where the combination coefficient  $\hat{\delta} = \hat{\pi}_1/(\hat{\pi}_1 + \hat{\pi}_2)$  with  $\hat{\pi}_1$  and  $\hat{\pi}_2$  given by (19) and (20).

Proposition 2 provides a simple and practical way to combine the  $1/N$  with the unbiased ML estimator  $\bar{w}$ . Theoretically, if  $\delta^*$  were known, the combination rule must dominate  $1/N$  unless  $w^* = 1/N$ . But  $\delta^*$  is unknown and has to be estimated in practice. This will introduce a loss in the expected utility due to errors in estimating  $\delta^*$ , making it uncertain whether  $\hat{w}^{\text{CML}}$  can still outperform  $1/N$ . Although the magnitude of the estimation error varies over empirical applications,  $\hat{w}^{\text{CML}}$  does outperform  $1/N$  with  $T$  as small as 120 in most scenarios of later simulations, and has close performances in other scenarios. Clearly, as  $T$  goes to infinity,  $\hat{w}^{\text{CML}}$  converges to the true optimal portfolio.

As alternatives, we can also consider an optimal combination of the  $1/N$  with either the three-fund rule of Jorion (1986) or the three-fund rule of Kan and Zhou (2007). Since the latter two rules are better than the unbiased ML one, the new combinations are likely to be even better. However, terms like  $1'_N \hat{\Sigma}^{-1} 1_N$  and  $\hat{\mu} \hat{\Sigma}^{-1} \hat{\mu}$  enter Jorion's estimator nonlinearly in both numerators and denominators of the function of interest (see, e.g., (A34) in the Appendix). As a result, analytical expressions for the combination coefficients are not feasible, and hence we will derive here only the combination of the  $1/N$  with Kan and Zhou's three-fund rule,

$$w_s = (1 - \delta_k)w_e + \delta_k \hat{w}^{\text{KZ}}, \quad (24)$$

where  $\hat{w}^{\text{KZ}}$  denotes the three-fund rule of Kan and Zhou (2007), which plays the role of the earlier unbiased ML rule in equation (12).  $\hat{w}^{\text{KZ}}$  is motivated by adding the global minimum variance portfolio into the usual ML estimator to hedge the estimation risk, and can be analytically written as

$$\hat{w}^{\text{KZ}} = \frac{T - N - 2}{\gamma c_1 T} \left[ \hat{\eta} \hat{\Sigma}^{-1} \hat{\mu} + (1 - \hat{\eta}) \hat{\mu}_g \hat{\Sigma}^{-1} 1_N \right], \quad (25)$$

where

$$\hat{\eta} = \hat{\psi}^2 / (\hat{\psi}^2 + N/T), \quad \hat{\psi}^2 = (\hat{\mu} - \hat{\mu}_g 1_N)' \hat{\Sigma}^{-1} (\hat{\mu} - \hat{\mu}_g 1_N) \quad (26)$$

and  $\hat{\mu}_g = \hat{\mu}' \hat{\Sigma}^{-1} 1_N / 1'_N \hat{\Sigma}^{-1} 1_N$ .

The optimal combination parameter  $\delta_k$  can be analytically solved. For practical applications, however, what we care the most is its estimator. To estimate it, we introduce the

following auxiliary parameter estimators (whose meaning is evident from the proof in the Appendix),

$$\hat{\pi}_{13} = \frac{1}{\gamma^2} \tilde{\theta}^2 - \frac{1}{\gamma} w'_e \hat{\mu} + \frac{1}{\gamma c_1} \left( [\hat{\eta} w'_e \hat{\mu} + (1 - \hat{\eta}) \hat{\mu}_g w'_e \mathbf{1}_N] - \frac{1}{\gamma} [\hat{\eta} \hat{\mu}' \tilde{\Sigma}^{-1} \hat{\mu} + (1 - \hat{\eta}) \hat{\mu}_g \hat{\mu}' \tilde{\Sigma}^{-1} \mathbf{1}_N] \right), \quad (27)$$

$$\hat{\pi}_3 = \frac{1}{\gamma^2} \tilde{\theta}^2 - \frac{1}{\gamma^2 c_1} \left( \tilde{\theta}^2 - \frac{N}{T} \hat{\eta} \right). \quad (28)$$

With these preparations, we are ready to summarize the result as

**Proposition 3:** *Among the combination rules  $\hat{w}_s = (1 - \delta_k)w_e + \delta_k \hat{w}^{\text{KZ}}$  of the  $1/N$  with  $\hat{w}^{\text{KZ}}$ , the estimated optimal one is*

$$\hat{w}^{\text{CKZ}} = (1 - \hat{\delta}_k)w_e + \hat{\delta}_k \hat{w}^{\text{KZ}}, \quad (29)$$

where the combination coefficient  $\hat{\delta}_k = (\hat{\pi}_1 - \hat{\pi}_{13})/(\hat{\pi}_1 - 2\hat{\pi}_{13} + \hat{\pi}_3)$  with  $\hat{\pi}_1$ ,  $\hat{\pi}_{13}$  and  $\hat{\pi}_3$  given by (19), (27) and (28), respectively.

Proposition 3 provides the estimated optimal combination rule that combines the  $1/N$  optimally with  $\hat{w}^{\text{KZ}}$ . By design, it should be better than the  $1/N$  if the errors in estimating  $\delta_k$  are small and if the  $1/N$  is not exactly identical to the optimal rule. This is indeed often the case in our later simulations. Overall, in fact,  $\hat{w}^{\text{CKZ}}$  will emerge as the best rule that performs well consistently across models and data sets.

### C. Rules Based on Factor Models

Consider a general  $K$ -factor model,

$$R_{tq} = \alpha + \beta F_t + \epsilon_t, \quad t = 1, 2, \dots, T, \quad (30)$$

where  $F_t$  is a  $K$ -vector of excess returns on  $K$  investable factors,  $R_{tq}$  is an  $(N - K)$ -vector of excess returns on non-factor risky assets, and  $\epsilon_t$  are the residuals with diagonal covariance matrix  $\Sigma_\epsilon$ . Putting the  $K$  factor returns in the first component, then we have the mean and covariance of the  $N$  risky assets,

$$\mu = \begin{pmatrix} \mu_F \\ \mu_R \end{pmatrix} = \begin{pmatrix} 0_K \\ \alpha \end{pmatrix} + \begin{pmatrix} \mu_F \\ \beta \mu_F \end{pmatrix} \quad (31)$$

and

$$\Sigma = \begin{pmatrix} \Sigma_F & \Sigma_F \beta' \\ \beta \Sigma_F & \beta \Sigma_F \beta' + \Sigma_\epsilon \end{pmatrix}, \quad (32)$$

where  $\mu_F$  and  $\Sigma_F$  are the mean and covariance matrix of  $F_t$ , and  $\mu_R$  is the mean of  $R_{tq}$ .

The question here is that, given the K-factor model for the return generating process, how one can make use of this information in forming the optimal portfolio in the presence of parameter uncertainty? Let  $\hat{\mu}_F$  and  $\hat{\Sigma}_F$  be the sample mean and covariance matrix of the factors, and  $\hat{\alpha}$ ,  $\hat{\beta}$  and  $\hat{\Sigma}_\epsilon$  be the standard ML estimator of the parameters. Then, it is easy to write out the ML estimator of the optimal rule in terms of these sample statistics. While the K-factor model is likely to improve the estimate accuracy on  $\Sigma$ , it does little in estimating the asset means. To provide a better estimator for the means which are related to the pricing errors, we use a James-Stein estimator for  $\alpha$ ,

$$\hat{\alpha}^{\text{JS}} = \left[ 1 - \frac{(N-3)(1 + \hat{\mu}'_F \hat{\Sigma}_F^{-1} \hat{\mu}_F)}{T \hat{\alpha}' \hat{\Sigma}_\epsilon^{-1} \hat{\alpha}} \right]^+ \hat{\alpha}. \quad (33)$$

With the above preparations, we can summarize our K-factor model based rule as:

**Proposition 4:** *Given the K-factor model, the ML rule that uses both the factor structure and the James-Stein estimator for the alphas is*

$$\hat{w}^{\text{FAC}} = \frac{1}{\gamma} \begin{pmatrix} \hat{\Sigma}_F^{-1} \hat{\mu}_F - \hat{\beta}' \hat{\Sigma}_\epsilon^{-1} \hat{\alpha}^{\text{JS}} \\ \hat{\Sigma}_\epsilon^{-1} \hat{\alpha}^{\text{JS}} \end{pmatrix}, \quad (34)$$

where  $\hat{\alpha}^{\text{JS}}$  is the James-Stein estimator given by (33).

McKinlay and Pastor (2000) propose a similar rule for factor models. They assume a latent factor structure that can be more reasonable in practice. In contrast,  $\hat{w}^{\text{FAC}}$  assumes a factor model not only with a known number of factors, but also known factor observations. If the factors are misidentified in an application, it is unlikely to perform well, as shown later. Hence,  $\hat{w}^{\text{FAC}}$  will be useful only in comparisons for knowing how much the factor structure can help. It should be used with caution unless one is sure of the known factor models.

#### D. Optimal Three- and Four-fund Rules

Consider first  $\hat{w}^{\text{CML}}$ , the combination rule that combines  $1/N$  with the unbiased estimator. This rule is a particular two-fund rule that invests money into funds determined by

weights  $w_e$  and  $\bar{w}$ . As an extension, we consider

$$\hat{w} = \phi_1 \bar{w}_e + \phi_2 \bar{w}, \quad (35)$$

where  $\phi_1$  and  $\phi_2$  are free parameters. If we restrict  $\phi_2 \geq 0$  and  $\phi_1 = 1 - \phi_2 \geq 0$ , we obtain  $\hat{w}^{\text{CML}}$ . Without that restriction, we invest per dollar an amount of  $\phi_1$  in fund  $\bar{w}_e$ ,  $\phi_2$  in fund  $v$ , and the difference  $1 - \phi_1 - \phi_2$  (positive or negative) in the riskfree asset. Hence,  $\hat{w}$  is a three-fund rule, of  $\bar{w}$ ,  $\hat{w}$  and the riskfree asset.

The utility maximizing optimal choice of  $\phi_1$  and  $\phi_2$  is given by

**Proposition 5:** *Among the rules given by (A7), the optimal combination parameters,  $\phi_1$  and  $\phi_2$ , that maximize the expected utility, are*

$$\begin{pmatrix} \phi_1^* \\ \phi_2^* \end{pmatrix} = \begin{pmatrix} 1'_N \Sigma 1_N & c_2 1'_N \mu \\ c_2 1'_N \mu & c_3 (\theta^2 + N/T) \end{pmatrix}^{-1} \begin{pmatrix} \frac{N}{\gamma} 1'_N \mu \\ c_2^2 \theta^2 \end{pmatrix}, \quad (36)$$

where  $c_1$  is given by (17),  $c_2 = T/(T - N - 2)$  and  $c_3 = \frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)}$ .

Proposition 5 provides the optimal three-fund rule that allocates money into the three funds. Although  $\phi_1^*$  and  $\phi_2^*$  depend on unknown parameters, they can be estimated from data. We will refer to the estimated optimal three-fund rule as

$$\hat{w}^{3F} = \hat{\phi}_1 w_e + \hat{\phi}_2 \bar{w}, \quad (37)$$

where  $\hat{\phi}_1$  and  $\hat{\phi}_2$  are the sample analogues of  $\phi_1^*$  and  $\phi_2^*$  by replacing the unknown parameters with their sample estimates.

Analogously, we can also consider an extension of  $\hat{w}^{\text{CKZ}}$ . This will be a four fund rule because there is an addition fund determined by the global minimum variance portfolio. That is, we examine

$$\hat{w} = \phi_{1k} w_e + \phi_{2k} \bar{w} + \phi_{3k} \bar{w}_g, \quad (38)$$

where  $\bar{w}_g = \hat{\Sigma}^{-1} 1_N$  is proportional to the estimated global minimum mean-variance portfolio, and the  $\phi$ 's are constant parameters. This rule contains all previous rules as special cases.

For the optimal choice of the  $\phi$ 's, we summarize the result as



**Proposition 6:** *Among the rules given by (A13), the optimal coefficients  $\phi_{1k}$ ,  $\phi_{2k}$ , and  $\phi_{3k}$ , that maximize the expected utility, are given by*

$$\begin{pmatrix} \phi_{1k}^* \\ \phi_{2k}^* \\ \phi_{3k}^* \end{pmatrix} = \begin{pmatrix} 1'_N \Sigma 1_N & c_2 1'_N \mu & c_2 N \\ c_2 1'_N \mu & c_3(\theta^2 + N/T) & c_3 1'_N \Sigma^{-1} \mu \\ c_2 N & c_3 1'_N \Sigma^{-1} \mu & c_3 1'_N \Sigma^{-1} 1_N \end{pmatrix}^{-1} \begin{pmatrix} \frac{N}{\gamma} 1'_N \mu \\ c_2^2 \theta^2 \\ \gamma c_1 c_2^2 1'_N w^* \end{pmatrix}. \quad (39)$$

Proposition 6 provides the optimal allocation among the four funds: cash,  $w_e$ ,  $\bar{w}$  and  $\bar{w}_g$ , with the cash position of  $1 - \phi_{1k}^* 1'_N w_e - \phi_{2k}^* 1'_N \bar{w} - \phi_{3k}^* 1'_N \bar{w}_g$ . As before, although  $\phi_{1k}$ ,  $\phi_{2k}$ , and  $\phi_{3k}$  depend on unknown parameters, they can be estimated from data. We will refer to the estimated optimal four-fund rule as

$$\hat{w}^{4F} = \hat{\phi}_{1k} 1_N + \hat{\phi}_{2k} \bar{w} + \hat{\phi}_{3k} \bar{w}_g, \quad (40)$$

where  $\hat{\phi}_{1k}$ ,  $\hat{\phi}_{2k}$ , and  $\hat{\phi}_{3k}$  are the sample analogues of  $\phi_{1k}$ ,  $\phi_{2k}$ , and  $\phi_{3k}$ . Theoretically, if the optimal  $\phi$ 's are known, the four-fund rule must outperform all of the other three combination rules. However, the four-fund rule must be estimated, and it has one or two more parameters to estimate than the others. Hence, empirically, whether it outperforms the others depends on the estimation errors in obtaining the  $\phi^*$ 's. This is an issue addressed in the next section.

## II. Performance Evaluation

In this section, we evaluate the performances of various rules (the  $1/N$ , some of the best existing rules and those proposed here) with data simulated from a range of possible models of the asset returns as well as with real data sets.

### A. Comparison in A One-factor Model

DeMiguel, Garlappi, and Uppal (2007) simulated data from a one-factor model only. Their approach is similar to that of MacKinlay and Pastor (2000). In their simulations, they assume that the factor,  $f_t = F_t$  in equation (30) with  $K = 1$ , has an annual excess return of 8% and an annual standard deviation of 16%. The mispricing  $\alpha$  is set to zero, and the factor loadings,  $\beta$ , are evenly spread between 0.5 and 1.5. Finally, the variance-covariance matrix of noise,  $\Sigma_\epsilon$ , is assumed to be diagonal, with elements drawn from a uniform distribution

with support  $[0.10, 0.30]$  so that the cross-sectional average annual idiosyncratic volatility is 20%. We follow their procedure exactly in what follows with two extensions. The first is that we examine not only a case of risk-aversion  $\gamma = 3$ , but also a case of  $\gamma = 1$ . The second is that we allow the case of nonzero alphas as well to assess the impact of mispricing on the results. The latter seems of practical interest because no known one-factor or K-factor models hold exactly in the real world.

Table I provides the average expected utilities of various rules in the one-factor model without mispricing and with  $N = 25$  assets. The results both here and later are all based on 10,000 simulated data sets. Panel A of the table corresponds to the case studied earlier by DeMiguel, Garlappi, and Uppal (2007) with  $\gamma = 3$ . The true expected utility is 4.17, while the  $1/N$  rule achieves a close value of 3.89 (all utilities are annualized and in percentage points). In contrast, the combination rules,  $\hat{w}^{\text{CML}}$  and  $\hat{w}^{\text{CKZ}}$ , have utility values of only 1.68 and 3.71, respectively, when  $T = 120$ . Although the values from  $\hat{w}^{\text{CKZ}}$  are close to those of the  $1/N$ , they are smaller until  $T$  reaches 3000. Theoretically, if the true combination coefficient were known,  $\hat{w}^{\text{CKZ}}$  must outperform the  $1/N$ . But the coefficient is unknown and has to be estimated from data. As a result, the estimation errors make  $\hat{w}^{\text{CKZ}}$  underperform. However, the differences are small and negligible. It should be noted that the underperformances occur only in this special simulation setup, as will be clear later.

Why does the  $1/N$  perform so well in the above simulation? This is because that the  $1/N$  rule is equivalent roughly to a 100% investment in holding the factor portfolio in the assumed factor model. To see why, we note first that the betas are evenly spread between 0.5 and 1.5, and so the  $1/N$ , an equal-weighted portfolio of the risky assets, should be close to the factor portfolio. Second, under the assumption of no mispricing, the factor portfolio is on the efficient frontier, and hence the optimal portfolio must be proportional to it. The proportion depends on  $\gamma$ . The optimal weights on the factor portfolio are

$$w_m^* = \frac{1}{\gamma} \frac{\mu_f}{\sigma_f^2}, \quad (41)$$

where  $\mu_f$  and  $\sigma_f^2$  are the factor excess return and variance, respectively. When  $\mu_f = 8\%$  and  $\sigma_f = 16\%$ , and when  $\gamma = 3$ ,  $w_m^* \approx 0.33 \times 0.08/0.16^2 = 1.03$ . This means that with  $\gamma = 3$  the optimal portfolio is 103% of the factor portfolio. Hence, the  $1/N$  portfolio should be

close to the optimal one. This is also evident by its utility value of 3.89. Since this value is close to the maximum possible, it is therefore true that the  $1/N$  performs well, and it will be difficult for any other rules that are estimated from the data to outperform it.

Theoretically,  $\hat{w}^{3F}$  and  $\hat{w}^{4F}$  should dominate  $\hat{w}^{\text{CML}}$  and  $\hat{w}^{\text{CKZ}}$ , respectively, if the combination coefficients were known. But the combination coefficients have to be estimated, and there is one more parameter to be estimated in comparison with the earlier two rules. As a result, the performances of  $\hat{w}^{3F}$  and  $\hat{w}^{4F}$  depend on the tradeoff between the gains from using one additional parameter with the losses from the estimation errors in estimating the additional parameter. The underperformance results of Panel A simply say that the estimation errors in this case are more important than the gains. However, it will not always be the case, as soon be shown in Panel B of Table I.

Of the existing estimated rules examined by Kan and Zhou (2007) and DeMiguel, Garlappi, and Uppal (2007), we study only four of them here. The first three are the better ones: MacKinlay and Pastor's (2000) rule, Jorion's (1986) three-fund rule (see the Appendix for the details of these two rules) and Kan and Zhou's (2007)  $\hat{w}^{\text{KZ}}$ . The fourth one is the popular ML estimator,  $\hat{w}^{\text{ML}}$ . Results on these four rules as well as  $\hat{w}^{\text{FAC}}$  (denoted as Factor ML), are reported in the last five column of Panel A in Table I.

Among the five rules, both the MacKinlay and Pastor (2000) and  $\hat{w}^{\text{FAC}}$  perform very well. It seems that the factor structure information is valuable if the data are indeed drawn from a factor model. For example, when  $T = 120$ , due to the estimation errors, both Jorion's and Kan-Zhou's rule have negative utilities of  $-12.85$  and  $-2.15$ , and the standard ML is the worst with a utility value of  $-85.72$ . This means that the three rules lose money on a risk-adjusted basis, and they make an investor worse off than putting money in the riskfree asset. In contrast, the MacKinlay and Pastor (2000) rule and  $\hat{w}^{\text{FAC}}$  have positive utilities of 2.11 and 2.29. As  $T$  increases, the five rules perform better. However, consistent with DeMiguel, Garlappi, and Uppal's (2007) finding, they except  $\hat{w}^{\text{FAC}}$  still underperform the  $1/N$  even when the sample size is as large as 6000. Overall, when  $T \leq 480$ , the  $1/N$  rule performs the best among all the 10 rules, the  $1/N$  and the nine estimated rules. But this will not be the case in other models, as pointed out earlier.

Equation (41) reveals also that, when  $\gamma = 1$ , the  $1/N$  rule will not be close to the optimal one. This is also evident from Panel B of Table I. In this case, the optimal investment is more aggressive and uses leverage. The expected utility is 12.50 from holding the true optimal portfolio. In contrast, if the  $1/N$  rule is followed, the expected utility is much lower: 6.63. Note that, although the  $1/N$  is not optimal, it still outperforms other rules with the exception of  $\hat{w}^{\text{FAC}}$  when  $T = 120$ . The reason is that it holds correctly the right efficient portfolio, though the proportion is incorrect. In contrast, the other rules must hold a portfolio based on estimated weights, which approximate the efficient portfolio weights with potentially large estimation errors. However, when  $T \geq 240$ ,  $\hat{w}^{\text{CKZ}}$  along with three other estimated rules outperforms the  $1/N$ . Nevertheless, the utility from  $\hat{w}^{\text{CKZ}}$  has a very close value of 6.36 when  $T = 120$ , and it outperforms the  $1/N$  when  $T \geq 240$ . Overall,  $\hat{w}^{\text{CKZ}}$  continues to perform well consistently in all the cases. Although not reported here, the results are qualitatively similar when  $\gamma$  is set to 6. Therefore, we find that, even without mispricing, the  $1/N$  can perform poorly for certain risk aversion parameters. After understanding the sensitivity of the  $1/N$  to  $\gamma$ , we assume  $\gamma = 3$  in what follows.

When there is mispricing, the  $1/N$  rule will get the composition of the optimal portfolio incorrect as well, since the factor portfolio will no longer be on the efficient frontier. In this case, the expected utility of the  $1/N$  rule can be far away from the true expected utility. Table II reports the results for two cases of the pricing errors in which the annualized alphas are evenly spread over  $-2\%$  and  $2\%$ , and over  $-5\%$  and  $5\%$ , respectively. In the first case (Panel A), the  $1/N$  rule has an expected utility of 3.89, about 40% less than 6.50, the true expected utility. Now even when  $T = 120$ ,  $\hat{w}^{\text{CKZ}}$  has an almost identical value as the  $1/N$ . As  $T$  increases, it outperforms the  $1/N$  easily. In the second case (Panel B), as the pricing errors become larger, the  $1/N$  rule has still an expected utility of 3.89, which becomes about 80% less than 18.73, the true expected utility. In this case, both  $\hat{w}^{\text{CML}}$  and  $\hat{w}^{\text{CKZ}}$  outperform the  $1/N$  substantially, even when  $T = 120$ , and much more so as  $T$  increases. Moreover, when  $T = 480$ , all the other rules, including the standard ML estimator, outperform the  $1/N$ . The concern of DeMiguel, Garlappi, and Uppal (2007) in the need of more than 3000 samples vanishes completely in this larger pricing errors case.

Overall, among all the four scenarios examined thus far, the combination rule  $\hat{w}^{\text{CKZ}}$

performs as well as the  $1/N$  in special cases, and much better in general. This suggests that there is indeed value-added when using portfolio theory to guide portfolio choice over the use of the naive  $1/N$  diversification. In addition, when  $T$  is less than or equal to 240,  $\hat{w}^{\text{CKZ}}$ , though occasionally outperformed by others slightly, is the best among all the rules across all scenarios and sample sizes. The above conclusion is also true when the number of assets is 50, as shown in Table III.

Following DeMiguel, Garlappi, and Uppal (2007), we also compare the performances of different rules in terms of Sharpe ratios. Table IV provides the results in the one-factor model. Panel A of the table corresponds to the case studied earlier by DeMiguel, Garlappi, and Uppal (2007). The  $1/N$  rule achieves a value of 13.95, which is close to the true Sharpe ratio of 14.43 (all Sharpe ratios are monthly and in percentage points, following the practice in the literature). In contrast,  $\hat{w}^{\text{CML}}$  and  $\hat{w}^{\text{CKZ}}$  have values of only 12.04 and 13.70, respectively, when  $T = 120$ . Although these two rules have close Sharpe ratios that of the  $1/N$ , they and other rules, with the exception of  $\hat{w}^{\text{FAC}}$ , have smaller values until  $T$  reaches 3000. Similar to the case of utility comparison, the results are driven by the fact that the  $1/N$  portfolio was set roughly equal to the true optimal one.

There are two surprising facts about the performances in terms of Sharpe ratios. In the absence of parameter uncertainty, the optimal portfolio that maximizes the expected utility must also maximize simultaneously the Sharpe ratio. But, in the presence of parameter uncertainty, this is no longer the case. For example, Kan and Zhou (2007) show that an optimal scaling of the covariance matrix can be applied to improve the ML rule to obtain higher expected utility because the scaling affects the mean linearly, but the variance nonlinearly. However, any such scaling is irrelevant here since the same Sharpe ratio will be retained. Because of this, it is surprising that the estimated rules that are designed to maximize the expected utility also have good Sharpe ratios. Second, the usual ML estimator of the true portfolio rule has close Sharpe ratios to the  $1/N$  when  $T = 960$ , a much better performance than the case in terms of the utilities.

When there is mispricing, for brevity, we consider only the case in which the annualized pricing errors ( $\alpha$ 's) are evenly spread over  $-2\%$  and  $2\%$ . Panel B of Table IV reports the results. Now the  $1/N$  rule has an average Sharpe ratio of 13.95, about 22% less than 18.02,

the true Sharpe ratio. In contrast, even when  $T = 120$ ,  $\hat{w}^{\text{CKZ}}$  has a higher value than the  $1/N$ . As  $T$  increases, it outperforms the  $1/N$  even more. In general, other rules perform well too. Table V provides similar results when  $N = 50$ . Hence, in terms of Sharpe ratios, the use of portfolio theory over the naive  $1/N$  diversification rule becomes even more attractive.

### *B. Comparison in A Three-factor Model*

Let us see now how the rules perform in a three-factor model. We use the same assumptions as before, except now we have three factors, which are the market portfolio plus the Fama-French's size and book-to-market portfolios. In the simulation, the means and covariance matrix of factors are calibrated from the monthly data from July 1963 to August 2007. The factor loadings of the non-benchmark risky assets are randomly paired and evenly spread between 0.9 and 1.2 for the market  $\beta$ 's, -0.3 and 1.4 for the size portfolio  $\beta$ 's, and -0.5 and 0.9 for the book-to-market portfolio  $\beta$ 's.<sup>9</sup>

In the three-factor model, the  $1/N$  rule is no longer close to the optimal portfolio. This is evident from Table VI, which reports the results for the two cases of the pricing errors, with the annualized  $\alpha$ 's at zero and evenly spread over  $-2\%$  and  $2\%$ , respectively. In the first case, the  $1/N$  rule has an expected utility of 3.85, about 70% less than 12.97, the true expected utility. Now even when  $T = 120$ ,  $\hat{w}^{\text{CKZ}}$  has a higher expected utility, 5.03, than the  $1/N$ . As  $T$  increases, both  $\hat{w}^{\text{CML}}$  and  $\hat{w}^{\text{CKZ}}$  outperform the  $1/N$  substantially. In the second case, when there are some pricing errors, the  $1/N$  rule still has an expected utility of 3.85, which becomes about 75% less than 14.60, the true expected utility. In this case, both  $\hat{w}^{\text{CML}}$  and  $\hat{w}^{\text{CKZ}}$  outperform the  $1/N$  by a much greater amount when  $T = 240$  and beyond. Moreover, when  $T = 960$ , and both with and without mispricing, all the other rules except MacKinlay and Pastor's rule, outperform the  $1/N$ . Similar results are found in Table VII when  $N = 50$ .

Table VIII reports the Sharpe ratios in the three-factor model when  $N = 25$ . Now the  $1/N$  has a Sharpe ratio about half of the true one. In contrast, most of the rules outperform it substantially even when  $T = 120$ . This is consistent with our earlier observation that

---

<sup>9</sup>These three ranges for the factor loadings are based on the ranges of the sample factor loadings of Fama-French's 25 size and book-to-market assets for the monthly data from July 1963 to August 2007.

outperforming the  $1/N$  is easier in terms of Sharpe ratios than in terms of utilities. When  $N = 50$ , Table IX provides similar results. Overall, in the three-factor model, we find even stronger evidence for outperforming the  $1/N$  than in the one-factor model. The reason is that the  $1/N$  portfolio deviates more from the optimal portfolio in the three-factor model than in the one-factor one. As the case with utilities,  $\hat{w}^{\text{CKZ}}$  performs well consistently.

### *C. Comparison with Calibrated Parameters*

The comparison so far assumes a factor model structure for the return-generating process. In general, investors may have doubts about the validity of any given factor models since no such models can capture fully the dynamics of the returns. It is therefore of interest to compare the performance without imposing any factor model structures. To do so, we consider two cases of using real data to calibrate the parameters. The first case is to use the monthly excess returns of the Fama-French 25 portfolios sorted on size and book-to-market ratio from July 1963 to August 2007, and the second is to use the 49 industry portfolios from July 1969 to August 2007 provided on French's web site. The sample means and covariance matrix are treated as the true parameters in the calibration, and then 10,000 data sets are simulated from the normal distribution with the calibrated parameters.

Table X reports the results for both of the cases. In the first case when  $N = 25$ , the  $1/N$  rule has an expected utility of 4.28, about 90% less than 44.96, the true expected utility. Now even when  $T = 120$ ,  $\hat{w}^{\text{CML}}$  and  $\hat{w}^{\text{CKZ}}$  have utilities of 17.40 and 12.99, more than three times larger than  $1/N$ . In addition, except the McKinlay and Pastor (2000) rule and especially the factor ML and ML rules, all the others outperform the  $1/N$  significantly. When  $T = 960$ , their utilities are quickly approaching 44.96. Since now there are no factor structures, this is why the McKinlay and Pastor (2000) rule and  $\hat{w}^{\text{FAC}}$  do not perform as well as before. A similar conclusion also holds for the second case when  $N = 49$ . However, when  $T = 120$ ,  $\hat{w}^{\text{CML}}$  and  $\hat{w}^{\text{CKZ}}$  do not outperform the  $1/N$  as greatly as before. This is because as  $N$  increases, their estimation errors are larger for a given  $T$ . Nevertheless, as  $T$  increases, they perform much better.

In terms of Sharpe ratios, Table XI reports the results. The Sharpe ratios are about twice or more as that of the  $1/N$  for most of the other rules. Now the ML rule has an impressive

performance given that how bad it was in terms of utilities. Overall, in comparison with the factor models, the performance of the  $1/N$  rule worsens greatly in the calibrated models. Therefore, there is an unambiguous evidence for the use of the proposed portfolio rules over the naive  $1/N$  one, and, again,  $\hat{w}^{\text{CKZ}}$  performs well consistently.

#### D. Comparison with Real Data

The results so far are based on simulated data sets. As emphasized by DeMiguel, Garlappi, and Uppal (2007), the advantage of using simulated data is to insulate the comparison results from the small-firm effect, calendar effects, momentum, mean-reversion, fat tails, or other anomalies that have been documented in the literature. In other words, because of the anomalies, results from real data do not constitute a proof that one rule is theoretically better than another. Nevertheless, due to the inclusion of real data in other studies, we in this subsection examine how the rules perform relative to one another with real data. The real data sets used in our analysis below are those used by DeMiguel, Garlappi, and Uppal (2007),<sup>10</sup> as well as the earlier Fama-French 25 portfolios with the three-factors, and the 49 industry portfolios plus the three factors.<sup>11</sup>

Following DeMiguel, Garlappi and Uppal (2007), we use a “rolling-sample” approach in the estimation. Given a  $T$ -month-long dataset of asset returns, we choose an estimation window of length  $M = 120$  and  $240$  months. In each month  $t$ , starting from  $t = M$ , we use the data in the most recent  $M$  months up to month  $t$  to compute the various portfolio rules, and apply them to determine the investments in the next month. For instance, let  $w_{z,t}$  be the estimated optimal portfolio rule in month  $t$  for a given rule ‘ $z$ ’, and let  $r_{t+1}$  be the excess return on the risky assets realized in month  $t + 1$ . The realized excess return on the portfolio is  $r_{z,t+1} = w'_{z,t}r_{t+1}$ . We then compute the Sharpe ratio associated with  $z$  by dividing the average value of the  $T - M$  realized returns,  $\hat{\mu}_z$ , by the standard deviation,  $\hat{\sigma}_z$ ; and calculate the *certainty-equivalent* return as

$$CER_z = \hat{\mu}_z - \frac{\gamma}{2}\hat{\sigma}_z^2,$$

---

<sup>10</sup>We thank Victor DeMiguel for the data, a detailed description of which can be found in DeMiguel, Garlappi, and Uppal (2007).

<sup>11</sup>Following Wang (2005), one can exclude the five largest of the Fama-French portfolios to make their linear combinations are not so close to the factors. But doing so has little impact on the results below.



which can be interpreted as the risk-free rate that an investor is willing to accept in stead of adopting a given risky portfolio rule  $z$ . Clearly the higher the CER, the better the rule. As before, we set the risk aversion coefficient  $\gamma$  be 3.

With the real data, the true optimal rule is unknown, but can be approximated by using the ML estimator based on the entire example. This will be referred as the in-sample ML rule. Although this rule is not implementable in practice, it is the rule that one would have obtained based on the ML estimator had he known all the data. Its performance serves as a useful benchmark to see how the estimation errors affect the out-of-sample results. Table XII reports the results for the five data sets used by DeMiguel, Garlappi, and Uppal (2007) in their Table 3, and the two additional data sets mentioned earlier.<sup>12</sup> Indeed, due to the limited sample size used in their estimation, all rules have CERs (annualized as before) less than half of those from the in-sample ML rule in most cases.

The first data set, the 10 industry returns plus the market, is a good example that highlights the problem of existing estimated rules. When  $M = 120$ , the in-sample ML has a CER of 8.42, the  $1/N$  rule has a decent value of 3.66, and  $\hat{w}^{\text{CKZ}}$  has 3.02. But the others have negative CERs, ranging from -38.18 to -0.76. For the international portfolios, the  $1/N$  remains hard to beat. Unlike other estimated optimal rules, the CER of  $\hat{w}^{\text{CKZ}}$  is significantly positive, but its difference with the  $1/N$  widens. However, for all the remaining five data sets,  $\hat{w}^{\text{CKZ}}$ , always performs the best among estimated rules, and outperforms the  $1/N$  by a large margin, with CERs about twice or much more. In contrast, the other estimated rules have varying performances, and lose money at least for one of the five data sets. This is really a serious problem with existing rules that have to be estimated from data.

The  $1/N$  rule is not immune either. When the data set is FF-4-factor (the twenty size- and book-to-market portfolios and the MKT, SMB, HML, and UMD factors), the  $1/N$  performs so poorly to have a negative return the first time. Interestingly, in this case, all estimated optimal rules except the ML, have significantly positive CERs, and  $\hat{w}^{\text{CKZ}}$  even has an CER of 25.40. This is an example where  $1/N$  should not be used, while the estimated rules have

---

<sup>12</sup>Note that, in comaprison with DeMiguel, Garlappi, and Uppal's (2007) Table 3, there is one missing column of results on the S&P sector data set, which is proprietary and not available here. In addition, the estimated rules are not normalized here, i.e., the weights on risky assets will not necessarily equal to one. This is desired from the way they are derived.

positive economic values. Once again,  $\hat{w}^{\text{CKZ}}$  is the best among all estimated rules, and is the only one that never loses money.

When the sample size increases to  $M = 240$ , the performances of all the estimated rules become better in many cases. Note that, unlike in simulation models, the  $1/N$  rule now has different values. Theoretically, the performance of the  $1/N$  rule should be invariant to  $M$ . However, when we increase  $M$  from 120 to 240, we have to drop 120 observations to make a fair comparison with other rules, resulting a new CER value for the  $1/N$ . Nevertheless,  $\hat{w}^{\text{CKZ}}$  remains the best among all the estimated rules, and it has a close performance in one case with the  $1/N$  and outperforms it in all other cases.

A related question is whether any of the portfolio strategies can beat the market out-of-sample. Suppose that one uses the standard ML rule to allocate his wealth among cash and the market index portfolio. The out-of-sample CERs are -0.88 and 2.40 when  $M = 120$  and 240, respectively. This has two implications. First, the standard ML rule requires  $M > 120$  to be meaningful even with the market as the single risky asset. Second, when  $M = 240$ , most of the estimated rules are better than investing into the market alone. It suggests that there are potential gains in devising rules that account for parameter uncertainty to beat the market.<sup>13</sup>

Similar to the simulation case, Table XIII shows that the estimated rules perform much better in terms of Sharpe ratios than in terms of the CERs. For example, most of them have close values to the  $1/N$  for the last five data sets even when  $M = 120$ . Again, the usual ML rule has remarkable performance and sometimes becomes close to the best when  $M = 240$ . In short, conclusions about Sharpe ratios from the simulations largely carry through to the real data case.

### *E. Further Analysis*

In this subsection, we analyze the strategies further in two ways. First, we provide the standard errors for both the utilities and Sharpe ratios of all the strategies. Second, we report the average estimated combination parameters for all of the four strategies that contain the

---

<sup>13</sup>While beyond the scope of this paper, it will be of interest to adapt the strategies here to form optimal 130-30 funds.

$1/N$  as part of the components.

So far,  $\hat{w}^{\text{CKZ}}$  emerges as the best rule that performs well consistently across simulation models. Hypothetically, this might happen with high standard errors in the utilities across data sets. To address this issue, Table XIV reports the standard errors of all the strategies when the data are drawn from a three-factor model with mispricing between  $-2\%$  to  $2\%$ , the case corresponding to Panel B of Table VI.<sup>14</sup> Both the true and the  $1/N$  rules are data-independent, and so their expected utilities are the same regardless of what data sets are used. For the estimated rules, their expected utilities are data-dependent, and their standard errors across data sets range from  $0.29\%$  to  $12.37\%$ , when  $T = 120$ . Interestingly,  $\hat{w}^{\text{CKZ}}$  has the smallest standard error while the ML has the largest. However, when  $T \geq 480$ , due to the factor structure, the McKinlay and Pastor (2000) and  $w^{\text{FAC}}$  have the best standard errors, while the ML still has the worst. Similar results are also true for the errors of the Sharpe ratios, as reported in Panel B of Table XIV.

Finally, it will be of interest to see how the  $1/N$  contributes to the other strategies when they are optimally combined. Table XV reports both the true and the average estimated combination parameters, in the same simulation model as in Table XIV, for all of the four strategies that contain the  $1/N$  as part of the components. Let us consider first  $\hat{w}^{\text{CML}}$  and  $\hat{w}^{\text{CKZ}}$ . When  $T = 120$ , the true optimal  $\delta$  for  $\hat{w}^{\text{CML}}$ , denoted simply by  $\delta$  in the table, is  $15.74\%$ , and the average estimated one  $20.56\%$ , biased upward. So it uses  $79.44\%$  ( $= 1 - 20.56\%$ ) of the  $1/N$  rule. In contrast, the  $\delta$  for  $\hat{w}^{\text{CKZ}}$  is much larger,  $53.78\%$ , and the average estimated value is  $56.18\%$ , slightly biased upward with much less usage of the  $1/N$ . The standard error of the  $\delta$  estimate is also much smaller for  $\hat{w}^{\text{CKZ}}$ . This might also help to explain why  $\hat{w}^{\text{CKZ}}$  performs well consistently. As  $T$  increases, the  $\delta$ 's are increasing as expected. It is of interest to note that the  $1/N$  remains to possess a few percentage points in the weighting even when the sample size is 6000.

Consider now  $\hat{w}^{3\text{F}}$  and  $\hat{w}^{4\text{F}}$ . In contrast with  $\hat{w}^{\text{CML}}$ ,  $\hat{w}^{3\text{F}}$  relaxes the constraint that the sum of the two  $\phi$ 's is one. Interestingly, the sums of both the optimal and estimated  $\phi$ 's are less than one, respectively. Without the constraint, we have to estimate both  $\phi$ 's, and the standard errors are much larger than the case with  $\hat{w}^{\text{CML}}$ . For the same reason,  $\hat{w}^{4\text{F}}$  also has

---

<sup>14</sup>Results in other simulation models are similar, and are omitted for brevity.

large standard errors with its estimates on  $\phi_{1k}$  and  $\phi_{2k}$ . However, the standard errors for both  $\phi_2$  and  $\phi_{2k}$  are relatively much smaller than  $\phi_1$  and  $\phi_{1k}$ , suggesting that the weighting on the  $1/N$  is more difficult to determine by the data than on the other funds.

### III. Future Research

In this section, we explore two directions for future research. The first is to obtain in some sense the best possible rule. The second is to find the optimal number of assets for asset allocation given a finite sample size.

In statistical decision theory (see Berger, 1985, or Lehmann and Casella, 1998), one way for judging an estimator is its admissibility. An estimator portfolio  $\hat{w}$  of the true optimal one is *admissible* if there is no other estimator  $\tilde{w}$  such that

$$L(w^*, \tilde{w}) \leq L(w^*, \hat{w}) \quad (42)$$

and if the inequality holds strictly for some true parameter values. Hence, if an estimator is admissible, one cannot find another estimator that is better sometimes and never worse. The ML rule estimator is an example of an inadmissible estimator, since, as shown by Kan and Zhou (2007), for all possible unknown parameters,

$$L(w^*, \tilde{w}) < L(w^*, \hat{w}_m) \quad (43)$$

where  $\tilde{w} = c_m \hat{w}^{\text{ML}}$ , a scaling adjustment of the ML rule with  $c_m$  as the scalar. However, whether  $\tilde{w}$  is admissible or not is still an open question.

The common tool for proving admissibility of an estimator is to relate it to a generalized Bayes estimator (GBE), which is defined as the estimator that minimizes the expected loss:

$$\min_{\hat{w}_b} E[L(w^*, \hat{w}_b)] = \frac{\gamma}{2} \int \int p(\mu, \Sigma) [(\hat{w}_b - w^*)' \Sigma (\hat{w}_b - w^*)] d\mu d\Sigma, \quad (44)$$

where  $p(\mu, \Sigma)$  is a prior density on  $\mu$  and  $\Sigma$ . Theoretically, if the prior is proper, and if there is a unique GBE, then the GBE must be admissible. It follows that any constant rule estimator, including the  $1/N$  rule, is admissible. This is because any other estimator must have a nonzero error when the true and unknown rule happens to be equal to the constant,

and hence it cannot dominate the constant estimator always. The constant estimators are known as trivial admissible estimators, which are often discarded in the statistical literature because they can be arbitrarily poor if the true true lies far away from it. This is the inconsistency problem: they do not converge to the true parameter even if there are infinite samples. Hence, in a statistical sense, a good estimator of the rule should be both admissible and consistent.

Although the two combination rules and the three- and four-fund rules are excellent investment strategies and do converge to the true optimal rule as the sample size increases to infinity, it is an open question whether or not they are admissible. In fact, it is unclear at all how a nontrivial admissible rule can be obtained in the context of mean-variance utility maximization. To see the difficulty, consider an estimator of the following type,

$$\hat{w}_a = \frac{1}{\gamma} \hat{\Sigma}_a^{-1} \hat{\mu}_a, \quad (45)$$

where  $\hat{\mu}_a$  and  $\hat{\Sigma}_a$  are GBEs of  $\mu$  and  $\Sigma$  to be determined below. Under any proper Bayes prior  $p(\mu, \Sigma)$ , the associated GBE for  $\mu$  can be solved,

$$\hat{\mu}_a = [E(\hat{\Sigma}_a^{-1} \Sigma \hat{\Sigma}_a^{-1})]^{-1} E(\hat{\Sigma}_a^{-1} \mu), \quad (46)$$

where the expectation is taken under  $p(\mu, \Sigma)$  and  $\hat{\Sigma}_a$  is not unique, and can in fact be arbitrary. Hence, the usual theory about the GBE does not apply.

To obtain an approximate admissible rule estimator, we assume that  $\Sigma$  is known for a moment. Then, the loss function, by equation (8), can be written as:

$$L(w^*, \hat{w}_a) = \frac{1}{2\gamma} E [(\hat{\mu}_a - \mu)' \Sigma^{-1} (\hat{\mu}_a - \mu)], \quad (47)$$

which is a problem of estimating  $\mu$  with a quadratic loss. Lin and Tsai (1973) provide an admissible estimator for this reduced loss function, even with  $\Sigma$  unknown,

$$\hat{\mu}_a = (1 - c_4/\hat{\theta}^2) \hat{\mu}, \quad (48)$$

where

$$c_4 = \frac{N-2}{T-N+2} - \frac{2}{T-N+2} \left[ \int_0^1 \frac{(1+\hat{\theta}^2)^{T/2}}{(1+\hat{\theta}^2 t)^{(T+2)/2}} t^{(N-4)/2} dt \right]^{-1}. \quad (49)$$

(see Appendix A for a proof) A combination of this mean estimator with an estimator of  $\Sigma$ ,  $\hat{\Sigma}_a$ , obtains an estimated rule  $\hat{w}_a = \hat{\Sigma}_a^{-1} \hat{\mu}_a / \gamma$ . Future research is needed to find an estimator of  $\Sigma$  such that  $\hat{w}_a$  can outperform the rules proposed in this paper.

In the parameter uncertainty literature, given  $N$  and  $T$ , one often solves the optimal investment strategy for investing money into *all*  $N$  risky assets, and this paper is no exception. In practice, though, the sample size may be considered as given, but we can devise strategies for investing into  $L$ ,  $L \leq N$ , assets given  $T$ . Then, it is a matter of how one chooses the optimal  $L$  to invest. The greater the  $L$ , the better the investment opportunity set, but the greater the estimation errors. This is evident not only from the formulas for the rules, but also from Tables I and III. Hence, there must be an optimal tradeoff between  $L$  (the optimally selected number of assets to invest) and the estimation errors. This is another interesting direction for future research.

Broadly speaking, the parameter uncertainty problem appears in almost all financial decision-making problems, and there is no reason to limit its studies to asset allocation, one of the oldest topics in finance. For example, how an investor values and hedges derivatives in the presence of parameter uncertainty is an important problem both in theory and practice, as is the question of how a corporate manager makes optimal investment and capital structure decisions when investors' expectations or the projects' opportunity sets or the macroeconomic determinants are unknown and have to be estimated. In short, a number of topics are related to the parameter uncertainty problem and call for future research.

#### IV. Conclusion

The modern portfolio theory pioneered by Markowitz (1952) is widely used in practice and taught in MBA texts. However, DeMiguel, Garlappi and Uppal (2007) raise serious doubts on its value. They show that the naive  $1/N$  investment strategy performs much better than those recommended from theory, and the estimation window needed for the latter to outperform the  $1/N$  benchmark is “around 3000 months for a portfolio with 25 assets and about 6000 months for a portfolio with 50 assets.” Note that existing theory-based strategies are expected to underperform the  $1/N$  when the latter happens to be close

to the true optimal portfolio, as is the case in the exact one-factor model of DeMiguel, Garlappi, and Uppal (2007), but the problem is that they still underperform when the  $1/N$  is substantially different from the true optimal portfolio. Moreover, they also perform poorly with many real data sets. This raises a serious question on the usefulness of the investment theory.

In this paper, we provide many new theory-based portfolio strategies, one of which, the optimal combination of the  $1/N$  rule with the three-fund rule of Kan and Zhou (2007), can perform well consistently across models and data sets for practical sample sizes of 120 and 240. In particular, this proposed strategy not only performs well relative to the  $1/N$  rule in an exact one-factor model that favors the  $1/N$ , but also outperforms it substantially in a one-factor model with mispricing, in multi-factor models with and without mispricing, in models calibrated from real data without any factor structures, and in applications with an array of real data sets. Overall, in comparison with existing rules, the key point is that the new strategy is the first and only one that performs well consistently, reaffirming the usefulness of the investment theory.

Our results are interesting not only in addressing the theoretical challenge posed by DeMiguel, Garlappi and Uppal (2007), but also in providing potentially useful insights into adapting actual quantitative investing strategies (see, e.g., Grinold and Kahn (1999), Litterman (2003) and Lo and Patel (2008)) to accommodate parameter estimation errors. However, there remain many theoretical issues. Whether or not our new portfolio strategies are the best (admissible) is still an open question, as is the problem of optimally choosing both the number of assets to be invested and the estimation strategy. Moreover, since parameter uncertainty problem appears in almost all financial decision-making problems, it is of interest to apply the ideas and techniques of this paper to a number of areas, such as how to value and hedge derivatives in the presence of parameter uncertainty, and how to make optimal investment and capital structure decisions when investors' expectations or the projects' opportunity sets or the macroeconomic determinants are unknown and have to be estimated. While studies of these questions go beyond the scope of this paper, they comprise important topics for future research.

## Appendix A: Proofs

### A.1. Proof of Proposition 1

Based on (14), we need only to show

$$(1 - \delta)^2 \pi_1 + \delta^2 \pi_2 = \pi_1 - 2\delta \pi_1 + \delta^2 (\pi_1 + \pi_2) < \pi_1 \quad (\text{A1})$$

when  $0 < \delta < 2\pi_1/(\pi_1 + \pi_2)$ . The Proposition then follows. Q.E.D.

### A.2. Proof of Proposition 2

We simply plug the estimates into the formula for the optimal combination coefficient,  $\delta^* = \pi_1/(\pi_1 + \pi_2)$ . Q.E.D.

### A.3. Proof of Proposition 3

Now, we have

$$L(w^*, \tilde{w}_s) = \frac{\gamma}{2} E \left[ \left[ (1 - \delta)(w_e - w^*) + \delta(\tilde{w} - w^*) \right]' \Sigma \left[ (1 - \delta)(w_e - w^*) + \delta(\tilde{w} - w^*) \right] \right],$$

where  $\tilde{w}$  denotes  $\hat{w}^{\text{KZ}}$  for brevity. Letting  $a = w_e - w^*$  and  $b = \tilde{w} - w^*$ , the following identity holds,

$$[(1 - \delta)a + \delta b]' \Sigma [(1 - \delta)a + \delta b] = (1 - \delta)^2 a' \Sigma a + 2\delta(1 - \delta) a' \Sigma b + \delta^2 b' \Sigma b.$$

Taking the first-order derivative of this identity, we get the optimal choice of  $\delta$ ,

$$\delta = \frac{a' \Sigma a - a' \Sigma E[b]}{a' \Sigma a - 2a' \Sigma E[b] + E[b' \Sigma b]}. \quad (\text{A2})$$

It is clear that  $\pi_1 = a' \Sigma a$ . Let  $\pi_{13} = a' \Sigma E[b] = w_e' \Sigma E[\tilde{w}] - w_e' \mu - \mu' E[\tilde{w}] + \mu \Sigma^{-1} \mu$ . Since  $E[\hat{\Sigma}^{-1}] = T \Sigma^{-1} / (T - N - 2)$ , we can estimate  $\pi_{13}$  with  $\hat{\pi}_{13}$  as given by (27). Finally, let  $\pi_3 = E[b' \Sigma b]$ . Using equation (63) of Kan and Zhou (2007), we can estimate  $\pi_3$  with  $\hat{\pi}_3$  as given by (28). Q.E.D.

### A.4. Proof of Proposition 4

The partition matrix  $\Sigma$  as given by (32) can be inverted analytically. Based on this and (31), the optimal weights are

$$w^* = \frac{1}{\gamma} \Sigma^{-1} \mu = \frac{1}{\gamma} \begin{pmatrix} \Sigma_F^{-1} \mu_F - \beta' \Sigma_\epsilon^{-1} \alpha \\ \Sigma_\epsilon^{-1} \alpha \end{pmatrix}. \quad (\text{A3})$$



Let  $\hat{\theta}_f^2 = \hat{\mu}'_F \hat{\Sigma}_F^{-1} \hat{\mu}_F$ . Conditional on  $\hat{\theta}_f^2$ , it is well known that

$$\sqrt{T/(1 + \hat{\theta}_f^2)} \hat{\alpha} \sim N(\sqrt{T/(1 + \hat{\theta}_f^2)} \alpha, \Sigma_\epsilon). \quad (\text{A4})$$

Therefore,

$$X = \Sigma_\epsilon^{-1/2} \sqrt{T/(1 + \hat{\theta}_f^2)} \hat{\alpha} \sim N(\sqrt{T/(1 + \hat{\theta}_f^2)} \Sigma_\epsilon^{-1/2} \alpha, I). \quad (\text{A5})$$

Applying the James-Stein shrinkage estimator to the mean of  $X$ , we have

$$\hat{\mu}_X^{\text{JS}} = \left[ 1 - \frac{N-3}{\|X\|^2} \right]^+ X. \quad (\text{A6})$$

This implies (33). Replacing  $\alpha$  by  $\hat{\alpha}^{\text{JS}}$  and replacing  $\Sigma_\epsilon$ , etc, by their ML estimators, we get (34) from (A3).

#### A.5. Proof of Proposition 5

For notational convenience, we rewrite the rule as

$$\hat{w} = q_1 \mathbf{1}_N + q_2 \bar{w}_p, \quad (\text{A7})$$

where  $q_1 = \phi_1/N$ ,  $q_2 = \phi_2/(\gamma c_2)$ ,  $\bar{w}_p = \hat{\Sigma}^{-1} \hat{\mu}$ . The loss function is then

$$L(w^*, \hat{w}) = \frac{\gamma}{2} E[(q_1 \mathbf{1}_N + q_2 \bar{w}_p - w^*) \Sigma (q_1 \mathbf{1}_N + q_2 \bar{w}_p - w^*)]. \quad (\text{A8})$$

Expanding this out and taking the derivatives with respect to the  $q$ 's, we get the first-order conditions,

$$0 = q_1 \mathbf{1}'_N \Sigma \mathbf{1}_N + q_2 E[\mathbf{1}'_N \Sigma \bar{w}_p] - \mathbf{1}'_N \Sigma w^*, \quad (\text{A9})$$

$$0 = q_2 E[\bar{w}'_p \Sigma \bar{w}_p] + q_1 E[\mathbf{1}'_N \Sigma \bar{w}_p] - E[\bar{w}'_p \Sigma w^*]. \quad (\text{A10})$$

Since  $E[\bar{w}_p] = c_2 \Sigma^{-1} \mu$ , we have  $E[\mathbf{1}'_N \Sigma \bar{w}_p] = c_2 \mathbf{1}'_N \mu$  and  $E[\bar{w}'_p \Sigma w^*] = \frac{1}{\gamma} c_2 \theta^2$ . Using equation (16) and (22) of Kan and Zhou (2007), we obtain

$$E[\bar{w}'_p \Sigma \bar{w}_p] = E[\hat{\mu}' \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \hat{\mu}] \quad (\text{A11})$$

$$= c_3 (\theta^2 + N/T). \quad (\text{A12})$$

The Proposition follows easily from here. Q.E.D.

#### A.6. Proof of Proposition 6

Similar to the proof of Proposition 5, we rewrite the rule in a simpler form,

$$\hat{w} = q_1 1_N + q_2 \bar{w}_p + q_3 \bar{w}_g, \quad (\text{A13})$$

where  $q_{1k} = \phi_{1k}/N$ ,  $q_{2k} = \phi_{2k}/(\gamma c_2)$ , and  $q_{3k} = \phi_{3k}/(\gamma c_1 c_2)$ . The loss function is then

$$L(w^*, \hat{w}) = \frac{\gamma}{2} E [(q_1 1_N + q_2 \bar{w}_p + q_3 \bar{w}_g - w^*) \Sigma (q_1 1_N + q_2 \bar{w}_p + q_3 \bar{w}_g - w^*)]. \quad (\text{A14})$$

Expanding this out and taking the derivatives with respect to the  $q$ 's, we get the first-order conditions,

$$0 = q_1 1'_N \Sigma 1_N + q_2 E[1'_N \Sigma \bar{w}_p] + q_3 E[1'_N \Sigma \bar{w}_g] - 1'_N \Sigma w^*, \quad (\text{A15})$$

$$0 = q_2 E[\bar{w}'_p \Sigma \bar{w}_p] + q_1 E[1'_N \Sigma \bar{w}_p] + q_3 E[\bar{w}'_p \Sigma \bar{w}_g] - E[\bar{w}'_p \Sigma w^*], \quad (\text{A16})$$

$$0 = q_3 E[\bar{w}'_g \Sigma \bar{w}_g] + q_1 E[1'_N \Sigma \bar{w}_g] + q_2 E[\bar{w}'_p \Sigma \bar{w}_g] - E[\bar{w}'_g \Sigma w^*]. \quad (\text{A17})$$

Since  $E[\bar{w}_g] = E[\hat{\Sigma}^{-1}] 1_N$ , we have  $E[1'_N \Sigma \bar{w}_g] = c_2 1'_N 1_N = c_2 N$  and  $E[\bar{w}'_g \Sigma w^*] = \frac{c_2}{\gamma} \mu' \Sigma^{-1} 1_N = c_2 1'_N w^*$ . Using equation (22) of Kan and Zhou (2007), we obtain

$$E[\bar{w}'_g \Sigma \bar{w}_g] = E[1'_N \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} 1_N] \quad (\text{A18})$$

$$= c_3 1'_N \Sigma^{-1} 1_N \quad (\text{A19})$$

and

$$E[\bar{w}'_g \Sigma \hat{w}'_p] = E[\hat{\mu}' \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} 1_N] \quad (\text{A20})$$

$$= c_3 1'_N \Sigma^{-1} \mu. \quad (\text{A21})$$

Then the Proposition follows. Q.E.D.

#### A.7. MacKinlay and Pastor's (2000) Rule and Its Analytical Solution

MacKinlay and Pástor (2000) impose an exact one-factor structure to provide a more efficient estimator of the expected returns by assuming

$$\Sigma = \sigma^2 I_N + a \mu \mu', \quad (\text{A22})$$

where  $a$  and  $\sigma^2$  are positive scalars. The ML estimator of  $a$ ,  $\sigma^2$  and  $\mu$  are obtained by maximizing the log-likelihood function

$$\ln \mathcal{L} = -\frac{NT}{2} \ln(2\pi) - \frac{T}{2} \ln(|a \mu \mu' + \sigma^2 I_N|) - \frac{1}{2} \sum_{t=1}^T (R_t - \mu)' (a \mu \mu' + \sigma^2 I_N)^{-1} (R_t - \mu). \quad (\text{A23})$$

This is an  $N + 2$  dimensional problem whose numerical solution is difficult.

Since we need to implement the rule hundreds and thousands of times, an analytical solution to the problem is critical.<sup>15</sup> Let  $\hat{U} = \hat{\Sigma} + \hat{\mu}\hat{\mu}'$ . Since

$$\ln(|a\mu\mu' + \sigma^2 I_N|) = (N - 1) \ln(\sigma^2) + \ln(\sigma^2 + a\mu'\mu). \quad (\text{A24})$$

and

$$\begin{aligned} & \sum_{t=1}^T (R_t - \mu)' (a\mu\mu' + \sigma^2 I_N)^{-1} (R_t - \mu) \\ &= \text{tr} \left( (a\mu\mu' + \sigma^2 I_N)^{-1} \sum_{t=1}^T (R_t - \mu)(R_t - \mu)' \right) \\ &= T \left[ \text{tr}((a\mu\mu' + \sigma^2 I_N)^{-1} \hat{\Sigma}) + (\hat{\mu} - \mu)' (a\mu\mu' + \sigma^2 I_N)^{-1} (\hat{\mu} - \mu) \right] \\ &= \frac{T}{\sigma^2} \left[ \text{tr}(\hat{\Sigma}) - \frac{a\mu'\hat{\Sigma}\mu}{\sigma^2 + a\mu'\mu} + (\hat{\mu} - \mu)'(\hat{\mu} - \mu) - \frac{a[(\hat{\mu} - \mu)'\mu]^2}{\sigma^2 + a\mu'\mu} \right] \\ &= \frac{T}{\sigma^2} \left[ \text{tr}(\hat{U}) + \frac{\sigma^2(\mu'\mu - 2\hat{\mu}'\mu) - a\mu'\hat{U}\mu}{\sigma^2 + a\mu'\mu} \right], \end{aligned} \quad (\text{A25})$$

we can minimize

$$f(\mu, a, \sigma^2) = (N - 1) \ln(\sigma^2) + \ln(\sigma^2 + a\mu'\mu) + \frac{1}{\sigma^2} \left[ \text{tr}(\hat{U}) + \frac{\sigma^2(\mu'\mu - 2\hat{\mu}'\mu) - a\mu'\hat{U}\mu}{\sigma^2 + a\mu'\mu} \right] \quad (\text{A26})$$

to obtain the ML estimator.

Let  $\hat{Q}\hat{\Lambda}\hat{Q}'$  be the spectral decomposition of  $\hat{U}$ , where  $\hat{\Lambda} = \text{Diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_N)$  are the eigenvalues in descending order and the columns of  $\hat{Q}$  are the corresponding eigenvectors. Further, let  $\hat{z} = \hat{Q}'\hat{\mu}$ . For any  $c$ ,  $\hat{\lambda}_1 \geq c \geq \hat{\lambda}_N$ , it can be shown that

$$p(\phi) = \sum_{i=1}^N \frac{(\hat{\lambda}_i - c)\hat{z}_i^2}{[1 - \phi(\hat{\lambda}_i - c)]^2} = 0 \quad (\text{A27})$$

has a unique solution, which can be trivially found numerically, in the interval  $(u_N, u_1)$  with  $u_i = 1/(\hat{\lambda}_i - c)$ . Then, the following objective function,

$$g(c) = \ln \left( c - \sum_{i=1}^N \frac{\hat{z}_i^2}{1 - \tilde{\phi}(c)(\hat{\lambda}_i - c)} \right) + (N - 1) \ln \left( \sum_{i=1}^N \hat{\lambda}_i - c \right), \quad (\text{A28})$$

---

<sup>15</sup>We are grateful to Raymond Kan for sharing his analytical solution (that involves only one trivial 1-dimensional optimization) with us.

is well defined, and can be solved easily because it is a one-dimensional problem. Let  $c^*$  be the solution, then the ML estimator of  $\mu$  is given by

$$\tilde{\mu} = \hat{Q}[I_N - \tilde{\phi}(c^*)(\hat{\Lambda} - c^*I_N)]^{-1}\hat{z}, \quad (\text{A29})$$

and hence the ML estimators of  $\sigma^2$  and  $a$  are

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^N \lambda_i - c^*}{N - 1}, \quad (\text{A30})$$

$$\tilde{a} = \frac{c^* - \tilde{\sigma}^2}{\tilde{\mu}'\tilde{\mu}} - 1. \quad (\text{A31})$$

The MacKinlay and Pástor (2000) portfolio rule is thus given by

$$\hat{w}^{\text{MP}} = \frac{\tilde{\mu}}{\gamma(\tilde{\sigma}^2 + \tilde{a}\tilde{\mu}'\tilde{\mu})} = \frac{\tilde{\mu}}{\gamma(c^* - \tilde{\mu}'\tilde{\mu})}. \quad (\text{A32})$$

#### A.8. Jorion (1986) Rule

Jorion (1986) develops a Bayes-Stein estimator of  $\mu$ ,

$$\hat{\mu}^{\text{BS}} = (1 - v)\hat{\mu} + v\hat{\mu}_g\mathbf{1}_N, \quad (\text{A33})$$

where

$$v = \frac{N + 2}{(N + 2) + T(\hat{\mu} - \hat{\mu}_g\mathbf{1}_N)'\tilde{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}_g\mathbf{1}_N)}, \quad \hat{\mu}_g = \frac{\mathbf{1}'_N\tilde{\Sigma}^{-1}\hat{\mu}}{\mathbf{1}'_N\tilde{\Sigma}^{-1}\mathbf{1}_N}. \quad (\text{A34})$$

His rule is then given by

$$\hat{w}^{\text{BS}} = \frac{1}{\gamma}(\hat{\Sigma}^{\text{BS}})^{-1}\hat{\mu}^{\text{BS}}, \quad (\text{A35})$$

where

$$\hat{\Sigma}^{\text{BS}} = \left(1 + \frac{1}{T + \hat{\lambda}}\right)\tilde{\Sigma} + \frac{\hat{\lambda}}{T(T + 1 + \hat{\lambda})}\frac{\mathbf{1}_N\mathbf{1}'_N}{\mathbf{1}'_N\tilde{\Sigma}^{-1}\mathbf{1}_N} \quad (\text{A36})$$

and  $\hat{\lambda} = (N + 2)/[(\hat{\mu} - \hat{\mu}_g\mathbf{1}_N)'\tilde{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}_g\mathbf{1}_N)]$ .

#### A.9. Proof of Equation (48)

The expression is based on Kubokawa (1991, p. 126). Note that  $X$  and  $S$  of that paper are  $\hat{\mu} \sim N(\mu, \Sigma/T)$  and  $\hat{\Sigma} \sim W_N(T - 1, \Sigma/T)$ , respectively. Then the equation follows. Q.E.D.

## REFERENCES

- Bawa, Vijay S., Stephen J. Brown, and Robert W. Klein, 1979, *Estimation Risk and Optimal Portfolio Choice* (North-Holland, Amsterdam).
- Berger, James O., 1985, *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag, New York).
- Black, Fischer, and Robert Litterman, 1992, Global Portfolio Optimization, *Financial Analysts Journal* 48, 28–43.
- Brandt, Michael W., 2004, Portfolio choice problems, in Y. Ait-Sahalia and L P. Hansen, eds.: *Handbook of Financial Econometrics*, forthcoming.
- Brown, Stephen J., 1976, Optimal portfolio choice under uncertainty: A Bayesian approach, Ph.D. dissertation, University of Chicago.
- Chincarini, Ludwig B., and Daehwan Kim, 2006, *Quantitative Equity Portfolio Management* (McGraw-Hill, New York).
- DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2007, Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, forthcoming.
- 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.
- Grinold, Richard C., and Ronald N. Kahn, 1999, *Active Portfolio Management: Quantitative Theory and Applications* (McGraw-Hill, New York).
- Harvey, Campbell R., John Liechty, Merrill W. Liechty, and Peter Müller, 2004, “Portfolio selection with higher moments, Working Paper, Duke University.

- Jorion, Philippe, 1986, Bayes-Stein estimation for portfolio analysis, *Journal of Financial and Quantitative Analysis* 21, 279–292.
- Kan, Raymond, and Guofu Zhou, 2007, Optimal portfolio choice with parameter uncertainty, *Journal of Financial and Quantitative Analysis* 42, 621–656.
- Kubokawa, Tatsuya, 1991, An approach to improving the James-Stein estimator, *Journal of Multivariate Analysis* 36, 121–126.
- Lehmann, E. L., and George Casella, 1998, *Theory of Point Estimation* (Springer-Verlag, New York).
- Lin, Pi-Erh, and Hui-Liang Tsai, 1973, Generalized Bayes minimax estimators of the multivariate normal mean with unknown covariance matrix, *Annals of Statistics* 1, 142–145.
- Litterman, Bob, 2003, *Modern Investment Management: An Equilibrium Approach* (Wiley, New York).
- Lo, Andrew W., and Pankaj N. Patel, 2008, 130/30: The new long-only, *Journal of Portfolio Management* 34, 12–38.
- MacKinlay, A. Craig, and Ľuboš Pástor, 2000, Asset pricing models: Implications for expected returns and portfolio selection, *Review of Financial Studies* 13, 883–916.
- Markowitz, Harry M., 1952, Portfolio selection, *Journal of Finance* 7, 77–91.
- Meucci, Attilio, 2005, *Risk and Asset Allocation* (Springer-Verlag, New York).
- Pástor, Ľuboš, 2000, Portfolio selection and asset pricing models, *Journal of Finance* 55, 179–223.
- Pástor, Ľuboš, and Robert F. Stambaugh, 2000, Comparing asset pricing models: An investment perspective, *Journal of Financial Economics* 56, 335–381.
- Qian, Edward, Ronald Hua, and Eric Sorensen, 2007, *Quantitative Equity Portfolio Management: Modern Techniques and Applications* (Chapman & Hall, New York).

- Stambaugh, Robert F., 1997, Analyzing investments whose histories differ in length, *Journal of Financial Economics* 45, 285–331.
- TerHorst, Jenke, Frans DeRoon, and Bas J. M. Werker, 2002, Incorporating estimation risk in portfolio choice, Working paper, Tilburg University.
- Tu, Jun, and Guofu Zhou, 2004, Data-generating process uncertainty: What difference does it make in portfolio decisions? *Journal of Financial Economics* 72, 385–421.
- Wang, Zhenyu, 2005, A shrinkage approach to model uncertainty and asset allocation, *Review of Financial Studies* 18, 673–705.

**Table I**  
**Utilities in A One-factor Model without Mispricing (N=25)**

This table reports the average utilities of a mean-variance investor under various investment rules: the true optimal one, the  $1/N$ , the two combination rules, the three- and four-funds, McKinlay and Pastor (2000), Jorion (1986), Kan and Zhou (2007), the ML rule with factor structure, and the standard ML estimator, with 10,000 sets of sample size  $T$  simulated data from a one-factor model with zero alphas and with  $N = 25$  assets. Panels A and B assume that the risk aversion  $\gamma$  is 3 and 1, respectively.

Rules	T					
	120	240	480	960	3000	6000
Panel A: $\gamma = 3$						
True	4.17	4.17	4.17	4.17	4.17	4.17
$1/N$	3.89	3.89	3.89	3.89	3.89	3.89
$\hat{w}^{\text{CML}}$	1.68	2.95	3.42	3.60	3.81	3.90
$\hat{w}^{\text{CKZ}}$	3.71	3.77	3.81	3.85	3.91	3.95
$\hat{w}^{3\text{F}}$	0.85	2.41	3.11	3.41	3.73	3.87
$\hat{w}^{4\text{F}}$	-0.33	1.75	2.74	3.19	3.65	3.83
McKinlay-Pastor	2.11	3.00	3.44	3.65	3.79	3.83
Jorion	-12.85	-3.79	-0.18	1.55	2.98	3.47
Kan-Zhou	-2.15	-0.00	1.13	1.90	2.97	3.47
Factor ML	2.29	3.27	3.73	3.95	4.10	4.13
ML	-85.72	-25.81	-8.35	-1.61	2.42	3.30
Panel B: $\gamma = 1$						
True	12.50	12.50	12.50	12.50	12.50	12.50
$1/N$	6.63	6.63	6.63	6.63	6.63	6.63
$\hat{w}^{\text{CML}}$	1.14	4.79	6.39	7.47	9.50	10.62
$\hat{w}^{\text{CKZ}}$	6.36	6.70	6.99	7.41	8.78	9.97
$\hat{w}^{3\text{F}}$	2.55	7.23	9.32	10.23	11.20	11.60
$\hat{w}^{4\text{F}}$	-0.98	5.26	8.21	9.58	10.96	11.49
McKinlay-Pastor	6.33	9.00	10.31	10.94	11.37	11.48
Jorion	-38.55	-11.38	-0.55	4.66	8.95	10.42
Kan-Zhou	-6.44	-0.01	3.38	5.69	8.92	10.40
Factor ML	6.86	9.81	11.18	11.84	12.29	12.39
ML	-257.16	-77.42	-25.05	-4.83	7.25	9.91



**Table II**  
**Utilities in A One-factor Model with Mispricing (N=25)**

This table reports the average utilities of a mean-variance investor under various investment rules: the true optimal one, the  $1/N$ , the two combination rules, the three- and four-funds, McKinlay and Pastor (2000), Jorion (1986), Kan and Zhou (2007), the ML rule with factor structure, and the standard ML estimator, with 10,000 sets of sample size  $T$  simulated data from a one-factor model with  $N = 25$  assets. Panels A and B assume that the mispricing  $\alpha$ 's, spread evenly between -2% to 2% per year and between -5% to 5% per year, respectively. The risk aversion coefficient  $\gamma$  is 3.

Rules	T					
	120	240	480	960	3000	6000
Panel A: $\alpha$ in $[-2\%, 2\%]$						
True	6.50	6.50	6.50	6.50	6.50	6.50
$1/N$	3.89	3.89	3.89	3.89	3.89	3.89
$\hat{w}^{\text{CML}}$	2.02	3.32	3.91	4.43	5.38	5.82
$\hat{w}^{\text{CKZ}}$	3.84	3.95	4.12	4.41	5.14	5.62
$\hat{w}^{3\text{F}}$	1.15	2.74	3.56	4.24	5.33	5.80
$\hat{w}^{4\text{F}}$	0.55	2.84	3.93	4.60	5.50	5.88
McKinlay-Pastor	2.34	3.23	3.67	3.88	4.02	4.06
Jorion	-12.36	-2.99	0.95	3.09	5.06	5.71
Kan-Zhou	-2.35	0.02	1.64	3.14	5.06	5.71
Factor ML	2.32	3.32	3.81	4.16	4.92	5.40
ML	-84.75	-23.84	-6.18	0.65	4.73	5.62
Panel B: $\alpha$ in $[-5\%, 5\%]$						
True	18.73	18.73	18.73	18.73	18.73	18.73
$1/N$	3.89	3.89	3.89	3.89	3.89	3.89
$\hat{w}^{\text{CML}}$	5.13	8.30	11.57	14.34	17.06	17.86
$\hat{w}^{\text{CKZ}}$	5.81	7.44	10.02	12.99	16.62	17.70
$\hat{w}^{3\text{F}}$	4.10	7.66	11.28	14.23	17.05	17.85
$\hat{w}^{4\text{F}}$	6.25	9.97	12.73	14.91	17.16	17.89
McKinlay-Pastor	2.70	3.60	4.04	4.25	4.40	4.43
Jorion	-6.32	5.21	10.76	14.10	17.03	17.85
Kan-Zhou	2.23	6.80	10.90	14.10	17.03	17.85
Factor ML	3.31	5.67	8.88	11.86	14.68	15.49
ML	-79.30	-13.62	5.23	12.52	16.85	17.80

**Table III**  
**Utilities in A One-factor Model (N=50)**

This table reports the average utilities of a mean-variance investor under various investment rules: the true optimal one, the  $1/N$ , the two combination rules, the three- and four-funds, McKinlay and Pastor (2000), Jorion (1986), Kan and Zhou (2007), the ML rule with factor structure, and the standard ML estimator, with 10,000 sets of sample size  $T$  simulated data from a one-factor model with  $N = 50$  assets. Panels A and B assume that the mispricing  $\alpha$ 's are zeros or between -2% to 2% per year, respectively. The risk aversion coefficient  $\gamma$  is 3.

Rules	T					
	120	240	480	960	3000	6000
Panel A: $\alpha=0$						
True	4.17	4.17	4.17	4.17	4.17	4.17
1/N	4.03	4.03	4.03	4.03	4.03	4.03
$\hat{w}^{\text{CML}}$	1.36	2.87	3.51	3.75	3.91	3.98
$\hat{w}^{\text{CKZ}}$	3.95	3.93	3.94	3.95	3.97	4.00
$\hat{w}^{3\text{F}}$	0.57	2.36	3.21	3.57	3.84	3.94
$\hat{w}^{4\text{F}}$	-0.70	1.69	2.84	3.37	3.77	3.90
McKinlay-Pastor	2.27	3.19	3.60	3.81	3.96	4.00
Jorion	-39.77	-12.03	-3.74	-0.25	2.20	2.96
Kan-Zhou	-3.15	-0.85	0.38	1.16	2.32	2.98
Factor ML	2.34	3.31	3.73	3.95	4.10	4.13
ML	-458.29	-83.43	-25.69	-8.39	0.58	2.42
Panel B: $\alpha$ in $[-2\%, 2\%]$						
True	8.71	8.71	8.71	8.71	8.71	8.71
1/N	4.03	4.03	4.03	4.03	4.03	4.03
$\hat{w}^{\text{CML}}$	1.80	3.46	4.31	5.08	6.62	7.41
$\hat{w}^{\text{CKZ}}$	4.07	4.18	4.38	4.79	6.06	6.98
$\hat{w}^{3\text{F}}$	0.97	2.91	3.96	4.89	6.57	7.39
$\hat{w}^{4\text{F}}$	-0.31	2.21	3.58	4.70	6.53	7.38
McKinlay-Pastor	2.29	3.21	3.62	3.83	3.98	4.01
Jorion	-39.22	-10.05	-1.12	2.90	6.16	7.25
Kan-Zhou	-2.49	0.24	2.02	3.66	6.19	7.25
Factor ML	2.34	3.32	3.75	4.02	4.80	5.60
ML	-467.35	-81.31	-21.88	-4.15	5.04	6.93

**Table IV**  
**Sharpe Ratios in A One-factor Model (N=25)**

This table reports the average Sharpe ratios of a mean-variance investor under various investment rules: the true optimal one, the  $1/N$ , the two combination rules, the three- and four-funds, McKinlay and Pastor (2000), Jorion (1986), Kan and Zhou (2007), the ML rule with factor structure, and the standard ML estimator, with 10,000 sets of sample size  $T$  simulated data from a one-factor model with  $N = 25$  assets. Panels A and B assume that the mispricing  $\alpha$ 's are zeros or between -2% to 2% per year, respectively. The risk aversion coefficient  $\gamma$  is 3.

Rules	T					
	120	240	480	960	3000	6000
Panel A: $\alpha=0$						
True	14.43	14.43	14.43	14.43	14.43	14.43
$1/N$	13.95	13.95	13.95	13.95	13.95	13.95
$\hat{w}^{\text{CML}}$	12.04	12.88	13.34	13.53	13.83	13.98
$\hat{w}^{\text{CKZ}}$	13.70	13.79	13.86	13.91	14.00	14.07
$\hat{w}^{\text{3F}}$	10.21	12.23	13.22	13.49	13.79	13.97
$\hat{w}^{\text{4F}}$	9.20	11.36	12.61	13.13	13.66	13.90
McKinlay-Pastor	12.19	13.51	13.86	13.89	13.89	13.89
Jorion	4.54	6.46	8.40	10.18	12.38	13.24
Kan-Zhou	4.97	7.03	8.80	10.27	12.34	13.24
Factor ML	12.81	14.06	14.39	14.42	14.43	14.43
ML	3.88	5.59	7.54	9.54	12.19	13.18
Panel B: $\alpha$ in $[-2\%, 2\%]$						
True	18.02	18.02	18.02	18.02	18.02	18.02
$1/N$	13.95	13.95	13.95	13.95	13.95	13.95
$\hat{w}^{\text{CML}}$	12.81	13.69	14.30	15.02	16.45	17.09
$\hat{w}^{\text{CKZ}}$	14.02	14.23	14.54	15.04	16.21	16.91
$\hat{w}^{\text{3F}}$	11.13	13.10	14.12	14.93	16.44	17.09
$\hat{w}^{\text{4F}}$	10.95	13.44	14.76	15.53	16.68	17.19
McKinlay-Pastor	12.70	13.98	14.28	14.30	14.31	14.31
Jorion	5.61	8.03	10.69	13.16	16.03	16.95
Kan-Zhou	4.77	7.15	10.09	12.97	16.02	16.95
Factor ML	12.89	14.14	14.54	14.78	15.82	16.53
ML	5.92	8.34	10.94	13.32	16.06	16.97

**Table V**  
**Sharpe Ratios in A One-factor Model (N=50)**

This table reports the average Sharpe ratios of a mean-variance investor under various investment rules: the true optimal one, the  $1/N$ , the two combination rules, the three- and four-funds, McKinlay and Pastor (2000), Jorion (1986), Kan and Zhou (2007), the ML rule with factor structure, and the standard ML estimator, with 10,000 sets of sample size  $T$  simulated data from a one-factor model with  $N = 50$  assets. Panels A and B assume that the mispricing  $\alpha$ 's are zeros or between -2% to 2% per year, respectively. The risk aversion coefficient  $\gamma$  is 3.

Rules	T					
	120	240	480	960	3000	6000
Panel A: $\alpha=0$						
True	14.43	14.43	14.43	14.43	14.43	14.43
1/N	14.19	14.19	14.19	14.19	14.19	14.19
$\hat{w}^{\text{CML}}$	11.90	12.87	13.49	13.77	14.00	14.11
$\hat{w}^{\text{CKZ}}$	14.09	14.07	14.10	14.11	14.14	14.17
$\hat{w}^{3\text{F}}$	9.37	11.95	13.28	13.74	13.97	14.10
$\hat{w}^{4\text{F}}$	8.22	11.00	12.64	13.39	13.85	14.04
McKinlay-Pastor	12.54	13.84	14.17	14.20	14.20	14.20
Jorion	2.68	4.33	6.16	8.08	10.96	12.29
Kan-Zhou	2.91	4.81	6.66	8.34	10.95	12.28
Factor ML	12.77	14.09	14.40	14.43	14.43	14.43
ML	2.39	3.87	5.59	7.53	10.69	12.17
Panel B: $\alpha$ in $[-2\%, 2\%]$						
True	20.87	20.87	20.87	20.87	20.87	20.87
1/N	14.19	14.19	14.19	14.19	14.19	14.19
$\hat{w}^{\text{CML}}$	12.91	14.12	15.01	16.09	18.26	19.29
$\hat{w}^{\text{CKZ}}$	14.31	14.54	14.95	15.72	17.79	19.01
$\hat{w}^{3\text{F}}$	10.68	13.34	14.79	16.00	18.23	19.28
$\hat{w}^{4\text{F}}$	9.45	12.36	14.16	15.70	18.17	19.26
McKinlay-Pastor	12.58	13.89	14.20	14.23	14.23	14.23
Jorion	5.18	8.17	11.18	14.06	17.72	19.10
Kan-Zhou	5.16	8.24	11.16	13.97	17.71	19.10
Factor ML	12.78	14.10	14.44	14.55	15.63	16.87
ML	4.91	7.74	10.75	13.75	17.64	19.07

**Table VI**  
**Utilities in A Three-factor Model (N=25)**

This table reports the average utilities of a mean-variance investor under various investment rules: the true optimal one, the  $1/N$ , the two combination rules, the three- and four-funds, McKinlay and Pastor (2000), Jorion (1986), Kan and Zhou (2007), the ML rule with factor structure, and the standard ML estimator, with 10,000 sets of sample size  $T$  simulated data from a three-factor model with  $N = 25$  assets. Panels A and B assume that the mispricing  $\alpha$ 's are zeros or between  $-2\%$  to  $2\%$  per year, respectively. The risk aversion coefficient  $\gamma$  is 3.

Rules	T					
	120	240	480	960	3000	6000
Panel A: $\alpha=0$						
True	12.97	12.97	12.97	12.97	12.97	12.97
$1/N$	3.85	3.85	3.85	3.85	3.85	3.85
$\hat{w}^{\text{CML}}$	3.41	5.42	7.30	9.21	11.42	12.14
$\hat{w}^{\text{CKZ}}$	5.03	5.84	6.96	8.51	11.01	11.96
$\hat{w}^{3\text{F}}$	2.45	4.78	6.97	9.08	11.40	12.13
$\hat{w}^{4\text{F}}$	2.59	5.35	7.41	9.34	11.46	12.15
McKinlay-Pastor	1.84	2.73	3.16	3.37	3.51	3.54
Jorion	-8.34	2.05	6.55	9.09	11.42	12.14
Kan-Zhou	1.54	4.71	7.02	9.10	11.41	12.13
Factor ML	6.92	10.13	11.59	12.28	12.75	12.86
ML	-81.77	-18.48	-0.13	6.94	11.14	12.06
Panel B: $\alpha$ in $[-2\%, 2\%]$						
True	14.60	14.60	14.60	14.60	14.60	14.60
$1/N$	3.85	3.85	3.85	3.85	3.85	3.85
$\hat{w}^{\text{CML}}$	3.84	6.15	8.44	10.63	13.02	13.76
$\hat{w}^{\text{CKZ}}$	5.09	6.06	7.57	9.59	12.56	13.58
$\hat{w}^{3\text{F}}$	2.87	5.51	8.12	10.51	13.00	13.75
$\hat{w}^{4\text{F}}$	2.53	5.63	8.26	10.61	13.02	13.76
McKinlay-Pastor	1.78	2.66	3.09	3.30	3.44	3.48
Jorion	-7.85	2.84	7.65	10.45	12.99	13.75
Kan-Zhou	1.61	5.12	7.96	10.45	12.99	13.75
Factor ML	6.92	10.13	11.60	12.30	12.92	13.29
ML	-81.09	-17.11	1.39	8.52	12.76	13.69

**Table VII**  
**Utilities in A Three-factor Model (N=50)**

This table reports the average utilities of a mean-variance investor under various investment rules: the true optimal one, the  $1/N$ , the two combination rules, the three- and four-funds, McKinlay and Pastor (2000), Jorion (1986), Kan and Zhou (2007), the ML rule with factor structure, and the standard ML estimator, with 10,000 sets of sample size  $T$  simulated data from a three-factor model with  $N = 50$  assets. Panels A and B assume that the mispricing  $\alpha$ 's are zeros or between -2% to 2% per year, respectively. The risk aversion coefficient  $\gamma$  is 3.

Rules	T					
	120	240	480	960	3000	6000
Panel A: $\alpha=0$						
True	12.97	12.97	12.97	12.97	12.97	12.97
$1/N$	3.87	3.87	3.87	3.87	3.87	3.87
$\hat{w}^{\text{CML}}$	2.15	4.23	5.76	7.55	10.32	11.45
$\hat{w}^{\text{CKZ}}$	4.22	4.71	5.46	6.72	9.52	10.99
$\hat{w}^{3\text{F}}$	1.38	3.68	5.43	7.39	10.28	11.44
$\hat{w}^{4\text{F}}$	0.68	3.73	5.70	7.62	10.36	11.46
McKinlay-Pastor	1.73	2.64	3.06	3.27	3.41	3.44
Jorion	-38.43	-7.80	1.79	6.30	10.14	11.40
Kan-Zhou	-1.45	1.93	4.41	6.84	10.16	11.40
Factor ML	6.97	10.13	11.58	12.28	12.74	12.86
ML	-474.00	-79.17	-18.33	-0.18	9.22	11.14
Panel B: $\alpha$ in $[-2\%, 2\%]$						
True	16.06	16.06	16.06	16.06	16.06	16.06
$1/N$	3.87	3.87	3.87	3.87	3.87	3.87
$\hat{w}^{\text{CML}}$	2.59	5.07	7.26	9.77	13.19	14.46
$\hat{w}^{\text{CKZ}}$	4.38	5.14	6.39	8.41	12.27	14.01
$\hat{w}^{3\text{F}}$	1.79	4.50	6.93	9.62	13.16	14.46
$\hat{w}^{4\text{F}}$	1.24	4.68	7.26	9.86	13.22	14.48
McKinlay-Pastor	1.70	2.61	3.03	3.24	3.38	3.42
Jorion	-37.96	-6.19	3.93	8.84	13.08	14.44
Kan-Zhou	-0.75	3.13	6.23	9.28	13.09	14.44
Factor ML	6.97	10.13	11.58	12.29	12.96	13.64
ML	-480.46	-77.71	-15.72	2.71	12.26	14.21

**Table VIII**  
**Sharpe Ratios in A Three-factor Model (N=25)**

This table reports the average Sharpe ratios of a mean-variance investor under various investment rules: the true optimal one, the  $1/N$ , the two combination rules, the three- and four-funds, McKinlay and Pastor (2000), Jorion (1986), Kan and Zhou (2007), the ML rule with factor structure, and the standard ML estimator, with 10,000 sets of sample size  $T$  simulated data from a three-factor model with  $N = 25$  assets. Panels A and B assume that the mispricing  $\alpha$ 's are zeros or between -2% to 2% per year, respectively. The risk aversion coefficient  $\gamma$  is 3.

Rules	T					
	120	240	480	960	3000	6000
Panel A: $\alpha=0$						
True	25.46	25.46	25.46	25.46	25.46	25.46
1/N	13.89	13.89	13.89	13.89	13.89	13.89
$\hat{w}^{\text{CML}}$	15.35	17.19	19.44	21.66	23.97	24.67
$\hat{w}^{\text{CKZ}}$	16.06	17.49	19.35	21.45	23.91	24.65
$\hat{w}^{3\text{F}}$	14.00	16.72	19.28	21.59	23.96	24.67
$\hat{w}^{4\text{F}}$	15.10	17.75	19.85	21.87	24.02	24.69
McKinlay-Pastor	11.47	12.91	13.33	13.37	13.38	13.38
Jorion	12.59	16.37	19.43	21.69	23.97	24.67
Kan-Zhou	13.35	16.76	19.37	21.62	23.97	24.67
Factor ML	22.07	23.73	24.60	25.02	25.32	25.39
ML	11.09	14.99	18.57	21.32	23.92	24.66
Panel B: $\alpha$ in $[-2\%, 2\%]$						
True	27.02	27.02	27.02	27.02	27.02	27.02
1/N	13.89	13.89	13.89	13.89	13.89	13.89
$\hat{w}^{\text{CML}}$	16.07	18.25	20.89	23.26	25.58	26.26
$\hat{w}^{\text{CKZ}}$	16.15	17.85	20.24	22.81	25.50	26.24
$\hat{w}^{3\text{F}}$	14.79	17.80	20.73	23.20	25.57	26.26
$\hat{w}^{4\text{F}}$	15.10	18.13	20.87	23.28	25.59	26.27
McKinlay-Pastor	11.33	12.77	13.20	13.25	13.25	13.25
Jorion	13.30	17.37	20.72	23.19	25.57	26.26
Kan-Zhou	13.55	17.37	20.55	23.13	25.56	26.26
Factor ML	22.08	23.73	24.62	25.05	25.49	25.82
ML	12.30	16.49	20.22	22.99	25.54	26.25

**Table IX**  
**Sharpe Ratios in A Three-factor Model (N=50)**

This table reports the average Sharpe ratios of a mean-variance investor under various investment rules: the true optimal one, the  $1/N$ , the two combination rules, the three- and four-funds, McKinlay and Pastor (2000), Jorion (1986), Kan and Zhou (2007), the ML rule with factor structure, and the standard ML estimator, with 10,000 sets of sample size  $T$  simulated data from a three-factor model with  $N = 50$  assets. Panels A and B assume the mispricing  $\alpha$ 's, are zeros or between  $-2\%$  to  $2\%$  per year, respectively. The risk aversion coefficient  $\gamma$  is 3.

Rules	T					
	120	240	480	960	3000	6000
Panel A: $\alpha=0$						
True	25.46	25.46	25.46	25.46	25.46	25.46
$1/N$	13.96	13.96	13.96	13.96	13.96	13.96
$\hat{w}^{\text{CML}}$	13.85	15.56	17.36	19.61	22.78	23.96
$\hat{w}^{\text{CKZ}}$	14.55	15.45	16.85	18.97	22.50	23.85
$\hat{w}^{3\text{F}}$	11.79	14.88	17.15	19.54	22.76	23.96
$\hat{w}^{4\text{F}}$	11.63	15.18	17.54	19.83	22.85	23.98
McKinlay-Pastor	11.26	12.72	13.13	13.19	13.19	13.19
Jorion	7.66	11.80	15.62	18.90	22.63	23.92
Kan-Zhou	7.93	12.13	15.65	18.84	22.63	23.92
Factor ML	22.01	23.73	24.59	25.02	25.32	25.39
ML	7.09	11.04	14.96	18.50	22.55	23.89
Panel B: $\alpha$ in $[-2\%, 2\%]$						
True	28.34	28.34	28.34	28.34	28.34	28.34
$1/N$	13.96	13.96	13.96	13.96	13.96	13.96
$\hat{w}^{\text{CML}}$	14.66	16.89	19.43	22.30	25.75	26.93
$\hat{w}^{\text{CKZ}}$	14.82	16.22	18.43	21.49	25.54	26.86
$\hat{w}^{3\text{F}}$	12.80	16.31	19.26	22.23	25.73	26.92
$\hat{w}^{4\text{F}}$	12.81	16.72	19.66	22.49	25.79	26.94
McKinlay-Pastor	11.19	12.67	13.08	13.14	13.14	13.14
Jorion	9.35	14.15	18.40	21.90	25.67	26.91
Kan-Zhou	9.63	14.40	18.37	21.85	25.67	26.91
Factor ML	22.01	23.73	24.59	25.03	25.53	26.16
ML	8.68	13.32	17.75	21.55	25.61	26.89



**Table X**  
**Utilities without Factor Structure**

This table reports the average utilities of a mean-variance investor under various investment rules: the true optimal one, the  $1/N$ , the two combination rules, the three- and four-funds, McKinlay and Pastor (2000), Jorion (1986), Kan and Zhou (2007), the ML rule with factor structure, and the standard ML estimator, with 10,000 sets of sample size  $T$  simulated data without assuming any factor model structure. Panels A and B simulate data sets based on the sample means and covariance matrix calibrated from the monthly excess returns of Fama-French 25 assets sorted on size and book-to-market ratio and Fama-French's 49 industry portfolios, respectively. The risk aversion coefficient  $\gamma$  is 3.

Rules	T					
	120	240	480	960	3000	6000
Panel A: $N = 25$ size and book-to-market portfolios						
True	44.96	44.96	44.96	44.96	44.96	44.96
$1/N$	4.28	4.28	4.28	4.28	4.28	4.28
$\hat{w}^{\text{CML}}$	17.40	27.07	34.46	39.17	42.97	43.95
$\hat{w}^{\text{CKZ}}$	12.99	21.53	30.74	37.49	42.69	43.87
$\hat{w}^{3\text{F}}$	16.29	26.53	34.28	39.12	42.97	43.95
$\hat{w}^{4\text{F}}$	16.76	26.88	34.44	39.17	42.98	43.95
McKinlay-Pastor	2.08	3.02	3.48	3.70	3.84	3.88
Jorion	10.28	26.38	34.52	39.22	42.98	43.95
Kan-Zhou	18.26	27.56	34.66	39.24	42.98	43.95
Factor ML	-57.75	-46.72	-42.72	-44.59	-49.08	-51.09
ML	-67.13	8.07	29.68	37.93	42.85	43.92
Panel B: $N = 49$ industry portfolios						
True	27.39	27.39	27.39	27.39	27.39	27.39
$1/N$	2.31	2.31	2.31	2.31	2.31	2.31
$\hat{w}^{\text{CML}}$	3.55	8.64	14.05	18.93	24.06	25.63
$\hat{w}^{\text{CKZ}}$	3.56	5.91	10.23	15.82	23.08	25.29
$\hat{w}^{3\text{F}}$	2.98	8.15	13.81	18.84	24.05	25.63
$\hat{w}^{4\text{F}}$	2.17	7.86	13.74	18.84	24.05	25.63
McKinlay-Pastor	0.42	1.36	1.80	2.02	2.16	2.20
Jorion	-34.14	0.41	12.39	18.67	24.05	25.63
Kan-Zhou	2.05	8.19	13.93	18.91	24.06	25.63
Factor ML	-341.92	-408.68	-507.89	-592.62	-669.56	-691.07
ML	-469.51	-68.70	-5.35	13.66	23.46	25.48

**Table XI**  
**Sharpe Ratios without Factor Structure**

This table reports the average Sharpe ratios of a mean-variance investor under various investment rules: the true optimal one, the  $1/N$ , the two combination rules, the three- and four-funds, McKinlay and Pastor (2000), Jorion (1986), Kan and Zhou (2007), the ML rule with factor structure, and the standard ML estimator, with 10,000 sets of sample size  $T$  simulated data without assuming any factor model structure. Panels A and B simulate data sets based on the sample means and covariance matrix calibrated from the monthly excess returns of Fama-French 25 assets sorted on size and book-to-market ratio and Fama-French 49 industry portfolios, respectively. The risk aversion coefficient  $\gamma$  is 3.

Rules	T					
	120	240	480	960	3000	6000
Panel A: $N = 25$ size and book-to-market portfolios						
True	47.41	47.41	47.41	47.41	47.41	47.41
$1/N$	14.63	14.63	14.63	14.63	14.63	14.63
$\hat{w}^{\text{CML}}$	31.14	37.58	41.88	44.43	46.41	46.90
$\hat{w}^{\text{CKZ}}$	27.00	35.26	41.17	44.26	46.39	46.90
$\hat{w}^{\text{3F}}$	30.52	37.30	41.78	44.40	46.40	46.90
$\hat{w}^{\text{4F}}$	30.85	37.45	41.87	44.43	46.41	46.90
McKinlay-Pastor	12.26	13.55	13.97	13.98	13.98	13.98
Jorion	32.25	38.14	42.06	44.48	46.41	46.91
Kan-Zhou	32.20	37.95	42.00	44.47	46.41	46.91
Factor ML	11.81	14.18	16.62	17.32	16.59	16.33
ML	30.64	37.33	41.77	44.40	46.40	46.90
Panel B: $N = 49$ industry portfolios						
True	37.01	37.01	37.01	37.01	37.01	37.01
$1/N$	11.26	11.26	11.26	11.26	11.26	11.26
$\hat{w}^{\text{CML}}$	16.66	21.70	26.94	30.97	34.75	35.83
$\hat{w}^{\text{CKZ}}$	13.33	17.36	23.71	29.68	34.59	35.79
$\hat{w}^{\text{3F}}$	15.48	21.42	26.86	30.93	34.74	35.83
$\hat{w}^{\text{4F}}$	14.93	21.08	26.76	30.92	34.74	35.83
McKinlay-Pastor	7.82	9.45	10.35	10.55	10.56	10.56
Jorion	15.03	21.69	27.08	31.02	34.75	35.83
Kan-Zhou	15.00	21.57	26.96	30.99	34.75	35.83
Factor ML	14.41	19.15	20.48	20.49	20.49	20.48
ML	14.31	20.96	26.63	30.84	34.73	35.82

**Table XII**  
**Certainty-equivalent Returns Based on Real Data**

This table reports the certainty-equivalent returns of a mean-variance investor under various investment rules: the in-sample ML rule, the  $1/N$ , the two combination rules, the three- and four-funds, McKinlay and Pastor (2000), Jorion (1986), Kan and Zhou (2007), the ML rule with factor structure, and the standard ML estimator. While the in-sample ML rule uses all the data for its estimation, other rules are based on a rolling sample with an estimation window  $M = 120$  or  $240$ , respectively. The real data sets are the five data sets used by DeMiguel, Garlappi, and Uppal (2007), and two additional data sets, the Fama-French 25 size and book-to-market portfolios with the Fama-French three-factors and the Fama-French 49 industry portfolios with the Fama-French three-factors. The risk aversion coefficient  $\gamma$  is 3.

Rules	Industry portfolios N=11	Inter'l portfolios N=9	Mkt/ SMB/HML N=3	FF- 1-factor N=21	FF- 4-factor N=24	FF25 3-factor N=28	Indu49 3-factor N=52
Panel A: M=120							
ML (in-sample)	8.42	7.74	13.61	46.04	54.55	45.24	57.67
$1/N$	3.66	3.26	4.33	5.27	-8.74	5.51	5.14
$\hat{w}^{CML}$	-1.39	-0.34	6.39	22.25	13.17	14.62	-6.40
$\hat{w}^{CKZ}$	3.02	1.79	8.54	28.97	25.40	19.36	8.51
$\hat{w}^{3F}$	-2.77	-2.18	5.63	20.47	23.48	13.05	-8.05
$\hat{w}^{4F}$	-2.51	-3.02	5.62	19.64	20.84	10.36	-14.36
McKinlay-Pastor	-0.76	0.86	-0.20	0.47	2.58	1.02	1.45
Jorion	-9.21	-5.80	9.51	0.82	1.58	-20.72	-152.10
Kan-Zhou	-3.59	-3.42	9.51	20.75	21.84	9.15	-17.77
Factor ML	-1.59	-0.44	5.45	-3.20	9.73	4.57	1.24
ML	-38.18	-18.30	4.90	-100.69	-128.59	-194.33	-1173.78
Panel B: M=240							
ML (in-sample)	8.42	7.74	13.61	46.04	54.55	45.24	57.67
$1/N$	5.04	0.92	3.46	4.44	-16.67	5.09	5.48
$\hat{w}^{CML}$	4.58	0.29	11.96	18.73	1.58	16.70	6.29
$\hat{w}^{CKZ}$	5.40	0.88	11.03	26.84	17.81	20.09	16.28
$\hat{w}^{3F}$	3.71	-0.49	11.84	18.23	11.22	16.34	5.49
$\hat{w}^{4F}$	3.80	0.27	3.83	24.79	11.04	15.03	15.75
McKinlay-Pastor	2.84	-0.02	0.44	2.78	3.24	3.37	4.32
Jorion	-0.76	-1.38	12.40	23.15	9.81	10.44	-18.70
Kan-Zhou	1.89	-0.17	12.21	26.60	18.92	14.08	12.43
Factor ML	1.73	-0.56	9.50	5.37	12.21	9.92	5.01
ML	-14.30	-6.94	12.08	-5.10	-38.63	-20.80	-158.40

**Table XIII**  
**Sharpe Ratios Based on Real Data**

This table reports the Sharpe ratios of a mean-variance investor under various investment rules: the in-sample ML rule, the  $1/N$ , the two combination rules, the three- and four-funds, McKinlay and Pastor (2000), Jorion (1986), Kan and Zhou (2007), the ML rule with factor structure, and the standard ML estimator. While the in-sample ML rule uses all the data for its estimation, other rules are based on a rolling sample with an estimation window  $M = 120$  or  $240$ , respectively. The real data sets are the five data sets used by DeMiguel, Garlappi, and Uppal (2007), and two additional data sets, the Fama-French 25 size and book-to-market portfolios with the Fama-French three-factors and the Fama-French 49 industry portfolios with the Fama-French three-factors. The risk aversion coefficient  $\gamma$  is 3.

Rules	Industry portfolios N=11	Inter'l portfolios N=9	Mkt/ SMB/HML N=3	FF- 1-factor N=21	FF- 4-factor N=24	FF25 3-factor N=28	Indu49 3-factor N=52
Panel A: M=120							
ML (in-sample)	20.52	19.67	26.09	47.98	52.23	47.56	53.70
$1/N$	13.53	12.77	22.40	16.23	23.40	16.87	16.41
$\hat{w}^{CML}$	8.23	6.76	18.09	38.50	41.57	32.32	14.32
$\hat{w}^{CKZ}$	12.37	10.00	21.24	38.56	37.45	32.62	22.45
$\hat{w}^{3F}$	7.43	4.18	17.64	37.25	40.28	31.10	12.18
$\hat{w}^{4F}$	8.58	4.60	21.23	37.37	39.59	30.51	16.54
McKinlay-Pastor	7.65	8.69	5.22	8.38	18.49	9.31	9.93
Jorion	5.64	2.58	22.13	37.98	39.69	29.80	14.74
Kan-Zhou	7.06	3.62	22.07	37.93	38.64	29.99	15.66
Factor ML	6.00	7.25	18.72	8.38	29.82	18.93	16.55
ML	2.74	-0.27	19.66	36.78	40.59	29.39	12.22
Panel B: M=240							
ML (in-sample)	20.52	19.67	26.09	47.98	52.23	47.56	53.70
$1/N$	16.33	8.00	20.48	14.90	20.67	16.39	17.79
$\hat{w}^{CML}$	15.29	6.57	24.56	37.31	40.52	33.72	21.75
$\hat{w}^{CKZ}$	16.80	7.89	24.42	36.72	36.86	31.73	32.09
$\hat{w}^{3F}$	14.03	5.39	24.45	36.81	39.68	33.33	20.94
$\hat{w}^{4F}$	14.62	8.30	17.38	39.22	39.45	32.81	31.47
McKinlay-Pastor	12.04	6.32	5.54	11.86	17.64	12.98	15.18
Jorion	11.48	5.27	24.91	41.19	42.14	34.62	26.24
Kan-Zhou	12.71	6.91	24.71	40.10	41.18	32.54	29.94
Factor ML	9.76	5.13	21.96	16.39	29.75	22.52	18.25
ML	8.31	0.26	24.79	41.15	43.85	37.13	19.78

**Table XIV**  
**Standard Errors of CERs and Sharp Ratios**

This table reports the standard errors (in percentage points) of the utilities (Panel A) and the Sharpe ratios (Panel B) for all the strategies with 10,000 sets of sample size  $T$  simulated data from a three-factor model with  $N = 25$  assets and with mispricing  $\alpha$ 's between -2% to 2% per year. The risk aversion coefficient  $\gamma$  is 3.

Rules	T					
	120	240	480	960	3000	6000
Panel A: Std of CERs						
True	0	0	0	0	0	0
1/N	0	0	0	0	0	0
$\hat{w}^{\text{CML}}$	1.24	0.62	0.47	0.31	0.13	0.07
$\hat{w}^{\text{CKZ}}$	0.29	0.35	0.40	0.36	0.17	0.08
$\hat{w}^{3\text{F}}$	1.25	0.67	0.50	0.32	0.13	0.07
$\hat{w}^{4\text{F}}$	1.44	0.75	0.52	0.32	0.13	0.07
McKinlay-Pastor	0.75	0.36	0.17	0.09	0.03	0.01
Jorion	3.55	1.26	0.61	0.33	0.13	0.07
Kan-Zhou	1.44	0.72	0.50	0.32	0.13	0.07
Factor ML	1.61	0.71	0.34	0.16	0.07	0.06
ML	12.37	3.29	1.24	0.53	0.15	0.08
Panel B: Std of Sharp Ratios						
True	0	0	0	0	0	0
1/N	0	0	0	0	0	0
$\hat{w}^{\text{CML}}$	2.44	2.22	1.85	1.12	0.42	0.22
$\hat{w}^{\text{CKZ}}$	1.66	1.88	1.84	1.30	0.45	0.23
$\hat{w}^{3\text{F}}$	4.23	2.70	1.93	1.15	0.42	0.22
$\hat{w}^{4\text{F}}$	4.43	2.77	2.00	1.15	0.42	0.22
McKinlay-Pastor	6.87	3.54	1.16	0.27	0.04	0.03
Jorion	4.18	2.87	1.84	1.12	0.42	0.22
Kan-Zhou	4.51	2.82	1.90	1.16	0.42	0.22
Factor ML	3.81	1.86	0.89	0.45	0.23	0.21
ML	4.02	3.01	2.00	1.19	0.43	0.22

**Table XV**  
**Combination Parameters**

The table reports both the true, the average estimated combination parameters and their standard errors (in percentage points) for all of the four strategies that uses the  $1/N$ . The data are simulated as in Table XIV.

Parameters	T					
	120	240	480	960	3000	6000
Panel A: $w^{CML}$						
$\delta$	15.74	29.93	47.57	65.12	85.60	92.27
$\hat{\delta}$	20.56 (10.87)	29.38 (13.44)	45.16 (12.49)	63.73 ( 7.61)	85.35 ( 2.05)	92.20 ( 0.80)
Panel B: $w^{CKZ}$						
$\delta_k$	53.78	68.09	79.87	88.35	95.81	97.84
$\hat{\delta}_k$	56.18 ( 6.37)	57.37 ( 7.70)	63.49 ( 7.88)	72.26 ( 6.49)	86.43 ( 3.09)	92.27 ( 1.52)
Panel C: $w^{3F}$						
$\phi_1$	79.60	66.20	49.55	32.97	13.61	7.30
$\hat{\phi}_1$	60.94 (42.81)	61.56 (32.50)	50.18 (20.62)	34.00 (10.87)	13.85 ( 2.74)	7.38 ( 1.04)
$\phi_2$	15.72	29.91	47.54	65.09	85.59	92.27
$\hat{\phi}_2$	18.82 (10.96)	27.73 (13.74)	44.05 (12.86)	63.21 ( 7.78)	85.25 ( 2.07)	92.17 ( 0.80)
Panel D: $w^{4F}$						
$\phi_{1k}$	73.20	61.64	47.56	32.83	14.21	7.75
$\hat{\phi}_{1k}$	57.42 (45.74)	58.99 (35.48)	49.65 (23.25)	34.70 (12.63)	14.64 ( 3.31)	7.88 ( 1.28)
$\phi_{2k}$	12.72	25.11	41.66	59.55	82.43	90.41
$\hat{\phi}_{2k}$	13.25 (12.78)	20.22 (16.26)	35.66 (16.01)	56.23 (10.17)	81.80 ( 2.87)	90.22 ( 1.13)
$\phi_{3k}$	0.09	0.07	0.06	0.04	0.02	0.01
$\hat{\phi}_{3k}$	0.09 ( 0.09)	0.08 ( 0.06)	0.06 ( 0.04)	0.04 ( 0.02)	0.02 ( 0.00)	0.01 ( 0.00)