1-2008

Markowitz Meets Talmud: A Combination of Sophisticated and Naive Diversification Strategies

Jun TU
Singapore Management University, tujun@smu.edu.sg

Guofu Zhou
Washington University in St Louis

Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies

Jun Tu and Guofu Zhou*

JEL classification: G11; G12

Keywords: Portfolio choice; mean-variance analysis; parameter uncertainty

First version: August, 2007
Current version: March, 2010

*We are grateful to Yacine Aït-Sahalia, Doron Avramov, Anil Bera, Henry Cao, Winghong Chan (the AFA-NFA discussant), Frans de Roon (the EFA discussant), Arnaud de Servigny, Victor DeMiguel, David Disatnik, Lorenzo Garlappi, Eric Ghysels, William Goetzmann, Yufeng Han, Bruce Hansen, Harrison Hong, Yongmiao Hong, Jing-zhi Huang (the SMUFSC discussant), Ravi Jagannathan, Raymond Kan, Hong Liu, Andrew Lo, Todd Milbourn, Ľuboš Pástor, Eduardo Schwartz, G. William Schwert (the managing editor), Paolo Zaffaroni, 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, 2008 Workshop on Advances in Portfolio Optimization at London Imperial College Business School, and especially to an anonymous referee for insightful and detailed comments that have substantially improved the paper. 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

Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies

The modern portfolio theory pioneered by Markowitz (1952) is widely used in practice and extensively taught to MBAs. However, the estimated Markowitz's portfolio rule and most of its extensions not only underperform the naive $1/N$ rule (that invests equally across N assets) in simulations, but also lose money on a risk-adjusted basis in many real data sets. In this paper, we propose an optimal combination of the naive $1/N$ rule with one of the four sophisticated strategies— the Markowitz rule, the Jorion (1986) rule, the MacKinlay and Pástor (2000) rule, and the Kan and Zhou (2007) rule— as a way to improve performance. We find that the combined rules not only have a significant impact in improving the sophisticated strategies, but also outperform the $1/N$ rule in most scenarios. Since the combinations are theory-based, our study may be interpreted as reaffirming the usefulness of the Markowitz theory in practice.

1. Introduction

Although more than half a century has passed since Markowitz's (1952) seminal paper, the mean-variance framework is still the major model used in practice today in asset allocation and active portfolio management despite many other models developed by academics.[1] One main reason 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 intertemporal hedging demand is typically found to be small. However, as is the case with any model, the true model parameters are unknown and have to be estimated from the data, resulting in the well-known parameter uncertainty or estimation error problem – the estimated optimal portfolio rule is subject to random errors and can thus be substantially different from the true optimal rule. Brown (1976), Bawa and Klein (1976), Bawa, Brown, and Klein (1979), and Jorion (1986) are examples of earlier work that provide sophisticated portfolio rules accounting for parameter uncertainty. Recently, MacKinlay and Pástor (2000) and Kan and Zhou (2007) provide more such rules.[2]

In contrast to the above approaches, the naive $1/N$ diversification rule, which invests equally across N assets of interest, relies on neither any theory nor any data. The rule, attributed to Talmud by Duchin and Levy (2009), has been known for about 1500 years, and corresponds to the equal weight portfolio in practice. Brown (1976) seems the first academic study of this rule. Due to estimation errors, Jobson and Korkie (1980) state that "naive formation rules such as the equal weight rule can outperform the Markowitz rule." Michaud (2008) further notes that "an equally weighted portfolio may often be substantially closer to the true MV optimality than an optimized portfolio." With similar conclusions, Duchin and Levy (2009) provide an up-to-date comparison of the $1/N$ rule with the Markowitz rule, and DeMiguel, Garlappi, and Uppal (2009) compare the $1/N$ rule further with almost all sophisticated extensions of the Markowitz rule. Not only that the naive $1/N$ investment strategy can perform better than those sophisticated rules recommended from investment

[1]See Grinold and Kahn (1999), Litterman (2003), and Meucci (2005) for practical applications of the mean-variance framework; and see Brandt (2009) for an excellent survey of the academic literature.

[2]Pástor (2000), Pástor and Stambaugh (2000), Harvey, Liechty, Liechty, and Müller (2004), Tu and Zhou (2004), and Wang (2005) are examples of recent Bayesian studies on the parameter uncertainty problem. We focus here on the classical framework.

theory, but also, as shown elsewhere and below, most of the Markowitz type rules do not perform well in real data sets and can even lose money on a risk-adjusted basis in many cases. These findings raise a serious doubt on the usefulness of the investment theory.

To address this problem, we examine two related questions. First, we ask whether any of the four sophisticated rules, namely, the Markowitz rule as well as its extensions proposed by Jorion (1986), MacKinlay and Pástor (2000), and Kan and Zhou (2007), can be combined with the naive $1/N$ rule to obtain better portfolio rules that can perform consistently well. Second, whether some or all of the combination rules can be sufficiently better so that they can outperform the $1/N$ rule. Positive answers to these two questions are important, for they will reaffirm the usefulness of the Markowitz theory if the theory-based combination rules can perform consistently well and outperform the non-theory-based $1/N$ rule. The positive answers are also possible based on both economic and statistical intuitions. Economically, a concave utility investor will prefer a suitable average of good and bad performances to either a good or a bad performance randomly, similar to the diversification over two risky assets. Statistically, a combination rule can be interpreted as a shrinkage estimator with the $1/N$ rule as the target. As is known in statistics and in finance (e.g., Jorion, 1986), the shrinkage is a tradeoff between bias and variance. The $1/N$ rule is biased, but has zero variance. In contrast, a sophisticated rule is usually asymptotically unbiased, but can have sizable variance in small samples. When the $1/N$ rule is combined with a sophisticated rule, an increase of the weight on the $1/N$ rule increases the bias, but decreases the variance. The performance of the combination rule depends on the tradeoff between the bias and the variance. Hence the performance of the combination rule can be improved and maximized by choosing an optimal weight.

We find that the four combination rules are substantially better than their sophisticated component rules in almost all scenarios under our study, and some of the combination rules outperform the $1/N$ rule as well, even when the sample size (T) is as small as 120. For example, when $T = 120$, in a three-factor model with 25 assets and with the annualized pricing errors spreading evenly between -2% to 2%, for a mean-variance investor with the risk aversion coefficient $\gamma = 3$, the four sophisticated rules, namely, the Markowitz rule and its extensions of Jorion (1986), MacKinlay and Pástor (2000), and Kan and Zhou (2007), have utilities (or risk-adjusted returns) of −81.09%, −7.85%, 1.78%, and 1.61%, two of them are losing money on a risk-adjusted basis. In contrast, their corresponding combination rules have utilities of 3.84%, 5.79%, 1.86%, and 5.09%. Hence, all the combination rules are better than their uncombined counterparts, and three of them improve greatly.[3] In comparison with the $1/N$ rule, which has a utility of 3.85% and is the best rule before implementing combinations, two of the combination rules have significantly higher utilities. When $T = 240$ or gets larger, while the $1/N$ rule, independent of T, still has the same performance, all the other rules improve and many of them outperform the $1/N$ rule much more significantly.

The methodology of this paper is based on the idea of combining portfolio strategies. Jorion (1986), Kan and Zhou (2007), DeMiguel, Garlappi, and Uppal (2009), and Brandt, Santa-Clara, and Valkanov (2009) have applied similar ideas in various portfolio problems. In contrast to these studies, this paper focuses on the combination of the $1/N$ rule with the aforementioned Markowitz type rules, and on reaffirming the value of the investment theory. In addition, from a Bayesian perspective, the idea of combining portfolio strategies is closely related to the Bayesian model averaging approach on portfolio selection, which Pástor and Stambaugh (2000) apply to compare various asset pricing models and Avramov (2002) applies to analyze return predictability under model uncertainty. This paper shows, along with these studies, that it is important and valuable to combine portfolio strategies in the presence of estimation errors.

The remainder of the paper is organized as follows. Section 2 provides the combination rules. Section 3 compares them with the $1/N$ rule and with their uncombined counterparts based on both simulated and real data sets. Section 4 concludes.

2. Combining portfolio strategies

In this section, we study the combination of the $1/N$ rule with each of the four sophisticated strategies. For easier understanding, we first briefly illustrate the general idea of combining two portfolio rules and then present in details the four combination rules in the order of their

[3]The MacKinlay and Pástor (2000) rule has excellent performance even before implementing any combination. But its combination rule improves little. As discussed later, this is not a problem with the rule itself, but a problem with the lack of a good estimation method for estimating the optimal combination coefficient.

analytical tractability.

2.1. Combination of two rules

We consider the following combination of two portfolio rules:

$$\hat{w}_c = (1-\delta)w_e + \delta\tilde{w}, \tag{1}$$

where $w_e = 1_N/N$ is the constant $1/N$ rule, $\tilde{w}$ is an estimated portfolio rule based on the data, and δ is the combination coefficient, $0 \leq \delta \leq 1$. The $1/N$ rule here is applied to N risky assets of interest.[4] The implied portfolio return of $\hat{w}_c$ at $T+1$ is $R_{pT+1} = r_{fT+1} + \hat{w}_c' R_{T+1}$, where r_{fT+1} is the return on the riskless asset, and R_{T+1} is an N-vector of excess returns on the N risky assets.[5]

Assume that the excess returns of the N risky assets are independent and identically distributed over time, and have a multivariate normal distribution with mean μ and covariance matrix Σ. Then the expected utility of $\hat{w}_c$ is

$$U(\hat{w}_c) = r_{fT+1} + \mu'\hat{w}_c - \frac{\gamma}{2}\hat{w}_c'\Sigma\hat{w}_c, \tag{2}$$

where γ is the mean-variance investor's relative risk aversion coefficient. Our objective is to find an optimal combination coefficient δ so that the following expected loss is minimized:

$$L(w^*, \hat{w}_c) = U(w^*) - E[U(\hat{w}_c)], \tag{3}$$

where $U(w^*)$ is the expected utility of the true optimal portfolio rule $w^* = \Sigma^{-1}\mu/\gamma$. This loss function is standard in the statistical decision theory, and is the criterion that Brown (1976), Frost and Savarino (1986), Stambaugh (1997), Ter Horst, De Roon, and Werker (2006), DeMiguel, Garlappi, and Uppal (2009), among others, use to evaluate portfolio rules.

The $1/N$ rule is chosen as the starting point of our combinations because it is simple, and yet can perform remarkably well when the sample size is small. Moreover, as is well-known in

[4]If the riskless asset is also included, the $1/N$ rule may be adjusted to $w_e = 1_N/(N+1)$. This, worsening from the earlier $1/N$ rule slightly, makes an insignificant difference in what follows.

[5]Note that the performances of most institutional managers are benchmarked to an index, say the S&P500. Then the return on the S&P500 index portfolio can be viewed as the riskless asset to apply the same framework. For active portfolio management with benchmarks, see Grinold and Kahn (1999), for example.

statistics (e.g., Lehmann and Casella, 1998), $1/N$ is one common choice of a good shrinkage point for improving the estimation of the mean of a multivariate distribution. However, the $1/N$ rule makes no use of any sample information, and will always fail to converge to the true optimal rule if it does not happen to be equal to it. In contrast, the combination rule always converges, and is designed to be better than either the $1/N$ rule or $\tilde{w}$ theoretically.

In practice, though, the true optimal combination coefficient δ is unknown. What is feasible is only a combination rule based on an estimated optimal δ, whose performance will then generally vary over applications. However, since the estimation errors in estimating the optimal δ, which is one single parameter, are usually small, the estimated optimal combination rule can generally improve both the $1/N$ rule and $\tilde{w}$ in our later analysis.

2.2. Combining with the Markowitz rule

The simplest case to start is to combine the $1/N$ rule with the standard ML rule or the (estimated) Markowitz rule. Let $\hat{\mu}$ and $\hat{\Sigma}$ be the sample mean and covariance matrix of R_{T+1}, then the ML rule is given by $\hat{w}^{\mathrm{ML}} = \hat{\Sigma}^{-1}\hat{\mu}/\gamma$. Instead of using $\hat{w}^{\mathrm{ML}}$, we use a scaled one:

$$\bar{w} = \frac{1}{\gamma}\tilde{\Sigma}^{-1}\hat{\mu}, \tag{4}$$

where $\tilde{\Sigma} = \frac{T}{T-N-2}\hat{\Sigma}$. The scaled $\bar{w}$ is unbiased and performs slightly better than $\hat{w}^{\mathrm{ML}}$.

According to (1), the combination rule is

$$\hat{w}_c = (1-\delta)w_e + \delta\bar{w}. \tag{5}$$

Then the expected loss associated with $\hat{w}_c$ is (all proofs are in the Appendices)

$$L(w^*, \hat{w}_c) = \frac{\gamma}{2}\left[(1-\delta)^2\pi_1 + \delta^2\pi_2\right], \tag{6}$$

where

$$\pi_1 = (w_e - w^*)'\Sigma(w_e - w^*), \qquad \pi_2 = E[(\bar{w} - w^*)'\Sigma(\bar{w} - w^*)].$$

Note that π_1 measures the impact from the bias of the $1/N$ rule, and π_2 measures the impact from the variance of $\bar{w}$. Thus, the combination coefficient δ determines the tradeoff between the bias and the variance. The optimal choice is easily shown as

$$\delta^* = \frac{\pi_1}{\pi_1 + \pi_2}. \tag{7}$$

Summarizing the result, we have

Proposition 1: *If* $\pi_1 > 0$*, then there exists an optimal* δ^**,* $0 < \delta^* < 1$*, such that*

$$L(w^*, \hat{w}_c) < \min[L(w^*, w_e), L(w^*, \bar{w})], \tag{8}$$

i.e., the optimal combination rule $\hat{w}_c$ *strictly dominates both the* $1/N$ *rule and* $\bar{w}$*.*

The condition $\pi_1 > 0$ is trivially satisfied in practice because the $1/N$ rule will not be equal to the true optimal rule with probability one. Proposition 1 says that the optimal combination rule $\hat{w}_c$ indeed provides strict improvements over both the $1/N$ rule and $\bar{w}$.[6] Suppose $\pi_1 = \pi_2$, then $\delta^* = 1/2$, and the loss of $\hat{w}_c$ will be only one half of the loss of either the $1/N$ rule or $\bar{w}$. This works exactly like a diversification over two independent and identically distributed risky assets.

To estimate δ^*, we only need to estimate π_1 and π_2, which can be done as follows:

$$\hat{\pi}_1 = w_e'\hat{\Sigma}w_e - \frac{2}{\gamma}w_e'\hat{\mu} + \frac{1}{\gamma^2}\tilde{\theta}^2, \tag{9}$$

$$\hat{\pi}_2 = \frac{1}{\gamma^2}(c_1 - 1)\tilde{\theta}^2 + \frac{c_1}{\gamma^2}\frac{N}{T}, \tag{10}$$

where $\tilde{\theta}^2$ is an estimator of $\theta^2 = \mu'\Sigma^{-1}\mu$ given by Kan and Zhou (2007), and $c_1 = (T-2)(T-N-2)/((T-N-1)(T-N-4))$. The condition of $T > N+4$ is needed here to ensure the existence of the second moment of $\hat{\Sigma}^{-1}$. Summarizing, we have

Proposition 2: *Assume* $T > N + 4$*. On the combination of the* $1/N$ *rule with* $\bar{w}$*,* $\hat{w}_c = (1-\delta)w_e + \delta\bar{w}$*, the estimated optimal one is*

$$\hat{w}^{\mathrm{CML}} = (1-\hat{\delta})w_e + \hat{\delta}\bar{w}, \tag{11}$$

where $\hat{\delta} = \hat{\pi}_1/(\hat{\pi}_1 + \hat{\pi}_2)$ *with* $\hat{\pi}_1$ *and* $\hat{\pi}_2$ *given by (9) and (10).*

Proposition 2 provides a simple way to optimally combine the $1/N$ rule with the unbiased ML rule $\bar{w}$. This combination rule is easy to carry out in practice since it is only a given function of the data. However, due to the errors in estimating δ^*, there is no guarantee that the estimated optimal combination rule, $\hat{w}^{\mathrm{CML}}$, will always be better than either the $1/N$

[6]Proposition 1 can be extended to allow any fixed constant rules, and can be adapted to allow biased estimated rules as well.

rule or $\bar{w}$. Nevertheless, in our later simulations, the magnitude of the errors in estimating δ^*, though varying over different scenarios, are generally small. Hence, $\hat{w}^{\mathrm{CML}}$ does improve upon $\bar{w}$, and can either outperform the $1/N$ rule or achieve close performances in most scenarios. Therefore, the combination does provide improvements overall. In addition, as T goes to infinity, $\hat{w}^{\mathrm{CML}}$ converges to the true optimal portfolio rule.

2.3. Combining with the Kan and Zhou (2007) rule

Consider now the combination of the $1/N$ rule with the Kan and Zhou (2007) rule, $\hat{w}^{\mathrm{KZ}}$, which is motivated to minimize the impact of estimation errors via a three-fund portfolio. With $\hat{\eta}$ and $\hat{\mu}_g$ as defined in their paper (as estimators of the squared slope of the asymptote to the minimum-variance frontier and the expected excess return of the global minimum-variance portfolio), we have

Proposition 3: *Assume $T > N + 4$. On the combination of the $1/N$ rule with the Kan and Zhou (2007) rule, $\tilde{w}_c = (1 - \delta_k)w_e + \delta_k \hat{w}^{\mathrm{KZ}}$, the estimated optimal one is*

$$\hat{w}^{\mathrm{CKZ}} = (1 - \hat{\delta}_k)w_e + \hat{\delta}_k \hat{w}^{\mathrm{KZ}}, \tag{12}$$

where $\hat{\delta}_k = (\hat{\pi}_1 - \hat{\pi}_{13})/(\hat{\pi}_1 - 2\hat{\pi}_{13} + \hat{\pi}_3)$ with $\hat{\pi}_1$ given by (9), and $\hat{\pi}_{13}$ and $\hat{\pi}_3$ given by

$$\begin{aligned} \hat{\pi}_{13} &= \frac{1}{\gamma^2}\tilde{\theta}^2 - \frac{1}{\gamma} w_e' \hat{\mu} + \frac{1}{\gamma c_1}\Big([\hat{\eta} w_e' \hat{\mu} + (1 - \hat{\eta})\hat{\mu}_g w_e' 1_N] \\ &\qquad - \frac{1}{\gamma}[\hat{\eta}\hat{\mu}'\tilde{\Sigma}^{-1}\hat{\mu} + (1 - \hat{\eta})\hat{\mu}_g \hat{\mu}'\tilde{\Sigma}^{-1} 1_N]\Big), \end{aligned} \tag{13}$$

$$\hat{\pi}_3 = \frac{1}{\gamma^2}\tilde{\theta}^2 - \frac{1}{\gamma^2 c_1}\left(\tilde{\theta}^2 - \frac{N}{T}\hat{\eta}\right). \tag{14}$$

Proposition 3 provides the estimated optimal combination rule that combines the $1/N$ rule with $\hat{w}^{\mathrm{KZ}}$. By design, it should be better than the $1/N$ rule if the errors in estimating the true optimal δ_k are small and if the $1/N$ rule is not exactly identical to the true optimal portfolio rule. This is indeed often the case in the performance evaluations in Section 3.

2.4. Combining with the Jorion (1986) rule

Consider now the combination of the $1/N$ rule with the Jorion (1986) rule, $\hat{w}^{\mathrm{PJ}}$, which is motivated from both the shrinkage and Bayesian perspectives. Assume $T > N + 4$ as before.

The optimal combination coefficient can be solved analytically in terms of the moments of $\hat{w}^{\text{PJ}}$,

$$\delta_j = \frac{\pi_1 - (w_e - w^*)'\Sigma E[\hat{w}^{\text{PJ}} - w^*]}{\pi_1 - 2(w_e - w^*)'\Sigma E[\hat{w}^{\text{PJ}} - w^*] + E[(\hat{w}^{\text{PJ}} - w^*)'\Sigma(\hat{w}^{\text{PJ}} - w^*)]}. \tag{15}$$

However, due to the complexity of $\hat{w}^{\text{PJ}}$, the analytical evaluation of the moments is intractable. In Appendix B, we provide an approximate estimator, $\hat{\delta}_j$, of δ_j, so that the estimated optimal combination rule,

$$\hat{w}^{\text{CPJ}} = (1 - \hat{\delta}_j)w_e + \hat{\delta}_j \hat{w}^{\text{PJ}}, \tag{16}$$

can be implemented easily in practice.

2.5. Combining with the MacKinlay and Pástor (2000) rule

In order to provide a more efficient estimator of the expected returns, MacKinlay and Pástor (2000) utilize an extension of the CAPM:

$$R_t = \alpha + \beta f_t + \epsilon_t, \tag{17}$$

where f_t is a latent factor. Let $\hat{\mu}^{\text{MP}}$ and $\hat{\Sigma}^{\text{MP}}$ be the maximum likelihood estimators of the parameters in their latent factor model (see Appendix C for the details), then the (estimated) MacKinlay and Pástor portfolio rule is given by the standard Markowitz formula, $\hat{w}^{\text{MP}} = (\hat{\Sigma}^{\text{MP}})^{-1}\hat{\mu}^{\text{MP}}/\gamma$. To optimally combine the $1/N$ rule with $\hat{w}^{\text{MP}}$, we need to evaluate the optimal combination coefficient:

$$\delta_m = \frac{\pi_1 - (w_e - w^*)'\Sigma E[\hat{w}^{\text{MP}} - w^*]}{\pi_1 - 2(w_e - w^*)'\Sigma E[\hat{w}^{\text{MP}} - w^*] + E[(\hat{w}^{\text{MP}} - w^*)'\Sigma(\hat{w}^{\text{MP}} - w^*)]}. \tag{18}$$

This requires the evaluation of the expectation terms associated with $\hat{w}^{\text{MP}}$. Since it is difficult to obtain them analytically, we use a Jackknife approach (e.g., Shao and Tu, 1996) to obtain an estimator, $\hat{\delta}_m$, of δ_m, via

$$E[\hat{w}^{\text{MP}} - w^*] \approx T(\hat{w}^{\text{MP}} - w^*) - \frac{T-1}{T}\sum_{t=1}^{T}(\hat{w}_{-t}^{\text{MP}} - w^*), \tag{19}$$

$$\begin{aligned} E[(\hat{w}^{\text{MP}} - w^*)'\Sigma(\hat{w}^{\text{MP}} - w^*)] \approx\; & T[(\hat{w}^{\text{MP}} - w^*)'\tilde{\Sigma}(\hat{w}^{\text{MP}} - w^*)] \\ & -\frac{T-1}{T}\sum_{t=1}^{T}[(\hat{w}_{-t}^{\text{MP}} - w^*)'\tilde{\Sigma}(\hat{w}_{-t}^{\text{MP}} - w^*)], \end{aligned} \tag{20}$$

where $\hat{w}_{-t}^{\mathrm{MP}}$ is the (estimated) MacKinlay and Pástor rule when the t-th observation ($t = 1, \ldots, T$) is deleted from the data. Then the estimated optimal combination rule is

$$\hat{w}^{\mathrm{CMP}} = (1 - \hat{\delta}_m) w_e + \hat{\delta}_m \hat{w}^{\mathrm{MP}}. \tag{21}$$

With the preparations thus far, it is ready to assess the performances of $\hat{w}^{\mathrm{CMP}}$ and other rules in both simulations and real data sets.

2.6. Alternative combinations and criteria

Before evaluating the combination rules provided above, we conclude this section by discussing some broader perspectives on the combination methodology.[7] First, on various ways of combining, what are the gains with combining more than two rules and with combining two rules not including the $1/N$ rule? Theoretically, if the true optimal combination coefficients are known, combining more than two rules must dominate combining any subset of them. However, the true optimal combination coefficients are unknown and have to be estimated. As more rules are combined, more combination coefficients need to be estimated and the estimation errors can grow. Hence combining more than two rules may not improve the performance. In addition, combining two rules not including the $1/N$ rule is usually not as good as including the $1/N$ rule, as done by our approach here. Nevertheless, certain optimal estimation methods might be developed to improve the performances of the more general combination approaches, which is an interesting subject of future research.

Second, on the objective of combining, what happens if the combination is to maximize a different objective function? The Sharpe ratio is such a natural objective which seems at least as popular as the utilities or risk-adjusted returns. When the true parameters are known, maximizing the Sharpe ratio and maximizing the expected utility are equivalent, a well-known fact. However, once the true parameters are unknown, the two are different. In this paper, we focus on maximizing the expected utility as it is easier to solve than maximizing the Sharpe ratio because the latter is to maximize a highly nonlinear function of the portfolio weights and there are no closed-form solutions available in the presence of estimation errors. Interestingly, though, due to their equivalence in the parameter certainty

[7]We are grateful to an anonymous referee for these and many other insights that help to improve the paper enormously.

case, the combination strategies of this paper, designed to maximize the expected utility, do in general have higher Sharp ratios than their uncombined components.

3. Performance evaluation

In this section, we evaluate the performances of the four combination rules and compare them with their uncombined counterparts and the $1/N$ rule, based on both simulated data sets (10,000 of them) and real data sets.

3.1. Comparison based on simulated data sets

Following MacKinlay and Pástor (2000), and DeMiguel, Garlappi, and Uppal (2009), we assume first the CAPM model with an annual excess return of 8% and an annual standard deviation of 16% on the market factor. The factor loadings, β's, are evenly spread between 0.5 and 1.5. The residual variance-covariance matrix is assumed to be diagonal, with the diagonal elements drawn from a uniform distribution with a support of $[0.10, 0.30]$ so that the cross-sectional average annual idiosyncratic volatility is 20%. In addition, we make two extensions. First, we examine not only a case of the risk aversion coefficient $\gamma = 3$, but also a case of $\gamma = 1$. Second, we allow nonzero alphas as well to assess the impact of mispricing on the results. This seems of practical interest because a given one-factor model (or any given K-factor models in general) may not hold exactly in the real world.

Table 1 provides the average expected utilities of the various rules over the simulated data sets without mispricing and with $N = 25$ assets, where the risk-free rate is set as zero without affecting the relative performances of different rules. Panel A of the table corresponds to the case studied by DeMiguel, Garlappi, and Uppal (2009) with $\gamma = 3$. The true expected utility is 4.17 (all utility values are annualized and in percentage points), greater than those from the estimated rules as expected due to estimation errors. But the four combination rules all have better performances than their uncombined counterparts, respectively. However, in comparison with the $1/N$ rule, which achieves a good value of 3.89, the combination rules have lower utility values, 1.68, 1.42, 2.19, and 3.71, when $T = 120$. Despite the improvements over their estimated uncombined counterparts, the combination rules suffer from estimation errors and still underperform the $1/N$ rule when T is small.

Why does the $1/N$ rule perform so well in the above case? This is because the assumed data-generating process happens to be in its favor: holding the $1/N$ portfolio is roughly equivalent to a 100% investment in the true optimal portfolio. To see why, we note first that the betas are evenly spread between 0.5 and 1.5, and so the $1/N$ portfolio 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 true optimal portfolio must be proportional to it. The proportion depends on γ. With $\gamma = 3$, the $1/N$ portfolio happens to be close to the true optimal portfolio, as evidenced by its utility value of 3.89 that is close to the maximum possible. It is therefore difficult for any other rules, which are estimated from the data, to outperform the $1/N$ rule in the above particular case.

However, when $\gamma = 1$, the $1/N$ rule will no longer be close to the true optimal portfolio. This is also evident from Panel B of Table 1. In this case, 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 $1/N$ rule is not optimal, it still outperforms the other rules when $T = 120$. The reason is that the $1/N$ rule now still holds correctly the efficient portfolio, though the proportion is incorrect. In contrast, the other rules depend on the estimated weights, which approximate the efficient portfolio weights with estimation errors. Nevertheless, $\hat{w}^{\text{CMP}}$ and $\hat{w}^{\text{CKZ}}$ have close results to the $1/N$ rule when $T = 120$, and they do better than it when $T \geq 240$. Overall, the combination rules improve the performances in this case as well and they do better in outperforming the $1/N$ rule than previously. After understanding the sensitivity of the $1/N$ rule to γ, we assume $\gamma = 3$ as usual in what follows.

When there is mispricing, Panel A of Table 2 reports the results where the annualized mispricing alphas are evenly spread between -2% to 2%. The combination rules again generally have better performances than their uncombined counterparts. Now the $1/N$ rule gets not only the proportion but also the composition of the optimal portfolio incorrect, since the factor portfolio is no longer on the efficient frontier. In this case, the expected utility of the $1/N$ rule, 3.89, is not close to but is about 40% less than the expected utility of the true optimal rule, 6.50. Now the combination rules not only improve, they also outperform the

$1/N$ rule more easily than before (Panel A of Table 1).

For the interest of comparison, we now study how the rules perform in a three-factor model. We use the same assumptions as before, except now we have three factors, whose means and covariance matrix are calibrated based on the monthly data from July 1963 to August 2007 on the market factor and Fama-French's (1993) size and book-to-market portfolios. The asset factor loadings are randomly paired and evenly spread between 0.9 and 1.2 for the market β's, between -0.3 and 1.4 for the size portfolio β's, and between -0.5 and 0.9 for the book-to-market portfolio β's. Panel B of Table 2 provides the results with the same mispring distribution as before (Panel A of Table 2). Once again, the combination rules are generally better than their estimated uncombined components. Since the $1/N$ rule is now far away from being the true optimal portfolio, it is outperformed by some of the combination rules even with $T = 120$. As T increases, the combination rules perform even better. Overall, combination improves performance, and some combination rules can outperform the $1/N$ rule in general. This suggests that there is indeed value-added through combining rules and by using portfolio theory to guide portfolio choice over the use of the naive $1/N$ diversification.[8]

3.2. Other properties of the combinations

In this subsection, we explore two aspects about the combination rules. First, while the combination rules are designed to maximize the expected utility, we also examine their performances in terms of the Sharpe ratio, and provide the standard errors for both the utilities and Sharpe ratios of the rules over the simulated data sets. Second, we study the estimation errors of the combination coefficients.

Table 3 provides in percentage points the Sharpe ratios in the one-factor model. Panel A of the table corresponds to the earlier case studied in Panel A of Table 1. Similar to the case in utilities, the combination rules generally improve the Sharpe ratio substantially, despite that maximizing the expected utility may not maximize the Sharpe ratio simultaneously in the presence of parameter uncertainty as discussed in Section 2.6. Prior to combining, all

[8]The same conclusion holds when the number of assets is 50, or in a model without factor structures. The results are available upon request.

the estimated rules, except the MacKinlay and Pástor rule, have Sharpe ratios less than 5 when $T = 120$. In contrast, the combination rules have Sharpe ratios close to that of the $1/N$ rule, 13.95, which in turn is close to the Sharpe ratio of the true optimal rule, 14.43. As discussed earlier on utilities, the reason why the $1/N$ rule does so well is because it is set roughly equal to the true optimal portfolio in this particular simulation design. When some mispricing is allowed (Panel B of Table 3), generally speaking, the combination rules again improve, and they are better than before.[9]

So far, the combination rules improve significantly across various simulation models. Hypothetically, this might happen with large standard errors in the utilities across data sets. To address this issue, Table 4 reports the standard errors of all the strategies when the data are drawn from a three-factor model with the annualized mispricing α's ranging from -2% to 2%, the case corresponding to Panel B of Table 2.[10] Both the true and the $1/N$ rules are data-independent, and so their expected utilities are the same across data sets. For the estimated rules, their expected utilities are data-dependent and vary across data sets with their standard errors ranging from 0.29% to 12.37%, when $T = 120$. The combination rules in general have smaller standard errors than their estimated component rules, especially when the sample size is less than 480. Similar results are also true for the standard errors of the Sharpe ratios, as reported in Panel B of Table 4.

To see how the $1/N$ rule contributes to the combination strategies, Table 5 reports both the true and the average estimated optimal combination coefficients for the four combination rules, with the data simulated in the same way as in Table 4. Consider first $\hat{w}^{\mathrm{CML}}$ and $\hat{w}^{\mathrm{CKZ}}$. When $T = 120$, the true optimal coefficient δ for $\hat{w}^{\mathrm{CML}}$, denoted simply by δ in the table, is 15.74%, but the average estimated one is 20.56%, biased upward. So the latter uses 79.44% ($= 1 - 20.56\%$) of the $1/N$ rule. In contrast, the true optimal δ for $\hat{w}^{\mathrm{CKZ}}$, 53.78%, is much larger, and the average estimated value is 56.18%, slightly biased upward with much less usage of the $1/N$ rule. The standard error of the estimated δ is also relatively smaller for $\hat{w}^{\mathrm{CKZ}}$. As T increases, the true optimal δ's are increasing as expected. It is of interest to note that the $1/N$ rule remains to possess a few percentage points in the weighting even

[9]Similar results hold, though not reported, in the three-factor model as well as in the non-factor model.
[10]The results in other simulation models are similar, and are omitted for brevity.

when the sample size is 6000.

On $\hat{w}^{\mathrm{CPJ}}$ and $\hat{w}^{\mathrm{CMP}}$, the estimates of their optimal combination coefficients have larger biases. This is because now we do not have analytical and accurate estimation formulas for them, unlike the case for $\hat{w}^{\mathrm{CML}}$ and $\hat{w}^{\mathrm{CKZ}}$. In particular, the bias in estimating the optimal combination coefficient for $\hat{w}^{\mathrm{CMP}}$ is quite large, which explains why the combination rule $\hat{w}^{\mathrm{CMP}}$ barely improves. Clearly, if better estimation methods are found, the performances of $\hat{w}^{\mathrm{CPJ}}$ and $\hat{w}^{\mathrm{CMP}}$ should improve. This will be yet another direction of future research.

Because of the small improvements of $\hat{w}^{\mathrm{CMP}}$ over $\hat{w}^{\mathrm{MP}}$ due to the inaccurate estimate of the true optimal combination coefficient, it may make sense to consider a simple naive combination of the MacKinlay and Pástor rule with the $1/N$ rule by using a 50% weight. As it turns out in the next section, this naive combination rule can improve over both $\hat{w}^{\mathrm{MP}}$ and $\hat{w}^{\mathrm{CMP}}$ and can perform well consistently across all real data sets in our study for practical sample sizes of 120 or 240. However, the same naive procedure does not improve $\hat{w}^{\mathrm{CPJ}}$ consistently because the difference (not reported here) between $\hat{w}^{\mathrm{CPJ}}$ and its true optimal combination rule (using the true optimal combination coefficient) is in general much smaller than that in the case of the MacKinlay and Pástor rule. As a result, we consider the naive combination only for the MacKinlay and Pástor rule in the next section.

3.3. Empirical application

Now we apply the various rules to the real data sets, which are those used by DeMiguel, Garlappi, and Uppal (2009),[11] as well as the Fama-French 49 industry portfolios plus the Fama-French three factors and the earlier Fama-French 25 portfolios plus the Fama-French three factors.

Given a sample size of T, we use a rolling estimation approach with two estimation windows of length $M = 120$ and 240 months, respectively. 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

[11] We thank Victor DeMiguel for the data. A detailed description of the data can be found in DeMiguel, Garlappi, and Uppal (2009).

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 average value of the $T-M$ realized returns, $\hat{\mu}_z$, and the standard deviation, $\hat{\sigma}_z$. The *certainty-equivalent* return is thus given by

$$CER_z = \hat{\mu}_z - \frac{\gamma}{2}\hat{\sigma}_z^2, \tag{22}$$

which can be interpreted as the risk-free rate of return that an investor is willing to accept instead of adopting the given risky portfolio rule z. Clearly the higher the CER, the better the rule. As before, we set the risk aversion coefficient γ to 3. Note that all the CERs have a common term of the average realized risk-free rate, which cancels out in their differences. Hence, as in the case for the expected utilities, we report the CERs by ignoring the risk-free rate term.

With the real data, the true optimal rule is unknown. We approximate it by using the ML estimator based on the entire sample. 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 in advance. Its performance may serve as a useful benchmark to measure how the estimation errors affect the out-of-sample results. Table 6 reports the results. Due to substantially less information in the rolling sample, all rules have CERs (annualized and in percentage points as before) less than half of those from the in-sample ML rule in most cases.

The first real data set, the 10 industry portfolios plus the market portfolio, is a good example that highlights the problem of the existing estimated rules. When $M = 120$, the in-sample ML rule has a CER of 8.42, the $1/N$ rule has a decent value of 3.66, and $\hat{w}^{\mathrm{CPJ}}$, $\hat{w}^{\mathrm{CMP}}$, and $\hat{w}^{\mathrm{CKZ}}$ have 3.15, 2.21, and 3.02. But the others including all the four uncombined estimated rules have negative CERs, ranging from -38.18 to -0.76, that is, they lose money on a risk-adjusted basis. For the second real data set, the international portfolios, the $1/N$ rule remains hard to beat. Unlike the other uncombined estimated rules, the MacKinlay and Pástor rule, and three combination rules have positive CERs. For all the remaining five data sets, the four combination rules work well in most cases. Overall, both $\hat{w}^{\mathrm{CMP}}$ and $\hat{w}^{\mathrm{CKZ}}$ have positive CERs consistently across all the seven data sets. This is an obvious improvement over the four uncombined theoretical rules, which can have negative CERs or lose money on a risk-adjusted basis. When $M = 240$, the results are even better. Both $\hat{w}^{\mathrm{CMP}}$ and $\hat{w}^{\mathrm{CKZ}}$ now still have positive CERs consistently across all the seven data sets. Moreover, most of the combination rules not only improve, but also outperform the $1/N$ rule most of the time.

In short, when applied to the real data sets, the combination rules generally improve from their uncombined Markowitz type counterparts and can perform consistently well, and some of them can outperform the $1/N$ rule in most of the cases.

4. Conclusion

The modern portfolio theory pioneered by Markowitz (1952) is widely used in practice and extensively taught to MBAs. However, due to parameter uncertainty or estimation errors, many studies show that the naive $1/N$ investment strategy performs much better than those recommended from the theory. Moreover, the existing theory-based portfolio strategies, except that of MacKinlay and Pástor (2000), perform poorly when applied to many real data sets used in our study. These findings raise a serious doubt on the usefulness of the investment theory. In this paper, we provide new theory-based portfolio strategies which are the combinations of the naive $1/N$ rule with the sophisticated theory-based strategies. We find that the combination rules are substantially better than their uncombined counterparts in general even when the sample size is small. In addition, some of the combination rules can perform consistently well and outperform the $1/N$ rule significantly. Overall, our study reaffirms the usefulness of the investment theory and shows that combining portfolio rules can potentially add significant value in portfolio management under estimation errors.

Since parameter uncertainty appears in almost every financial decision-making problem, our ideas and results may be applied to various other areas. For example, they may be applied to turn many practical quantitative investing strategies (e.g., Lo and Patel, 2008) into those more robust to estimation errors; they may also be applied to hedge derivatives optimally in the presence of parameter uncertainty; or be applied to make optimal capital structure decisions with unknown investors' expectations and macroeconomic determinants. While studies of these issues go beyond the scope of this paper, they seem interesting topics for future research.

Appendix A. Proofs of Propositions and Equations

A.1. Proof of Proposition 1

Based on (6), we need only to show

$$f(\delta) \equiv (1-\delta)^2\pi_1 + \delta^2\pi_2 = \pi_1 - 2\delta\pi_1 + \delta^2(\pi_1+\pi_2) \tag{23}$$

satisfies $f'(\delta^*) = 0$ and $f''(\delta^*) > 0$ at δ^*, which are easy to verify. Then the claim follows. Q.E.D.

A.2. Proof of Equation (10)

In many expectation evaluations below, a key is to apply two equalities about the inverse of the sample covariance matrix (e.g., Haff, 1979), i.e., the formulas for $E[\Sigma^{\frac{1}{2}}\hat{\Sigma}^{-1}\Sigma^{\frac{1}{2}}]$ and $E[\Sigma^{\frac{1}{2}}\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\Sigma^{\frac{1}{2}}]$. Expanding out the quadratic form of π_2 into three terms, and applying the formulas to the two terms involving $\bar{w}$, we have

$$\pi_2 \quad = \quad \frac{1}{\gamma^2}(c_1-1)\theta^2 + \frac{c_1}{\gamma^2}\frac{N}{T}. \tag{24}$$

Then, plugging the estimator for θ^2 into this equation yields the desired claim. Q.E.D.

A.3. Proof of Proposition 3

Now, we have

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

where $\tilde{w}$ denotes $\hat{w}^{\mathrm{KZ}}$ for brevity. Let $a = w_e - w^*$ and $b = \tilde{w} - w^*$, then 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 obtain the optimal choice of δ,

$$\delta = \frac{a'\Sigma a - a'\Sigma E[b]}{a'\Sigma a - 2a'\Sigma E[b] + E[b'\Sigma b]}. \tag{25}$$

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 π_{13} with $\hat{\pi}_{13}$ as given by (13). Finally, let $\pi_3 = E[b'\Sigma b]$. Using Eq. (63) of Kan and Zhou (2007), we can estimate π_3 with $\hat{\pi}_3$ as given by (14). Q.E.D.

Appendix B. Combining with the Jorion (1986) rule

Equation (15) follows from (25). To compute $E[(\hat{\Sigma}^{\mathrm{PJ}})^{-1}\hat{\mu}^{PJ}]$, we rewrite

$$\hat{\Sigma}^{\mathrm{PJ}} = d\tilde{\Sigma} + DD', \tag{26}$$

where d and D are defined accordingly from (26). Inverting this matrix, we have

$$(\hat{\Sigma}^{\mathrm{PJ}})^{-1} = \tilde{\Sigma}^{-1}/d - \tilde{\Sigma}^{-1}DD'\tilde{\Sigma}^{-1}/(d^2 + dD'\tilde{\Sigma}^{-1}D) = \tilde{\Sigma}^{-1}/d - B, \tag{27}$$

where B is defined as the second term. Since it is relatively small, we treat it as a constant. Then, we approximately evaluate $E[(\hat{\Sigma}^{\mathrm{PJ}})^{-1}\hat{\mu}^{PJ}]$ as the product of the expectations.

Finally, we have from (27) that

$$(\hat{\Sigma}^{\mathrm{PJ}})^{-1}\Sigma(\hat{\Sigma}^{\mathrm{PJ}})^{-1} = \tilde{\Sigma}^{-1}\Sigma\tilde{\Sigma}^{-1}/d^2 - 2(\tilde{\Sigma}^{-1}\Sigma B/d) + (B'\Sigma B). \tag{28}$$

The first term can be evaluated as in Proposition 3. The second and third terms are trivial. Hence, we can evaluate approximately $E[\hat{\mu}^{PJ\prime}(\hat{\Sigma}^{\mathrm{PJ}})^{-1}\Sigma(\hat{\Sigma}^{\mathrm{PJ}})^{-1}\hat{\mu}^{PJ}]$ by treating $\hat{\Sigma}^{\mathrm{PJ}}$ and $\hat{\mu}^{PJ}$ as independent variables.

Appendix C. Semi-analytical solution to the MacKinlay and Pástor (2000) rule

Assume normality with $E(f_t) = 0$, $\mathrm{var}(f_t) = \sigma_f^2$, $E(f_t, \epsilon_t) = 0_N$, and that the covariance matrix of the residuals is $\sigma^2 I_N$, with I_N as the identity matrix. Moreover, assume that an exact asset pricing relation holds with $\mu = \beta\gamma_f$, where γ_f is the factor risk premium. Then, $\Sigma = \sigma^2 I_N + a\mu\mu'$, where $a = \sigma_f^2/\gamma_f^2$. The maximum likelihood estimator, $\hat{\mu}^{\mathrm{MP}}$ and $\hat{\Sigma}^{\mathrm{MP}}$, of μ and Σ, are obtained by maximizing the log-likelihood function over σ^2, a and μ:

$$\ln\mathcal{L} = -\frac{NT}{2}\ln(2\pi) - \frac{T}{2}\ln\left(|a\mu\mu' + \sigma^2 I_N|\right) - \frac{1}{2}\sum_{t=1}^{T}(R_t - \mu)'(a\mu\mu' + \sigma^2 I_N)^{-1}(R_t - \mu). \tag{29}$$

The numerical solution to the optimization problem can be very demanding due to the number of parameters. Fortunately, there is available an almost analytical solution.[12]

Let $\hat{U} = \hat{\Sigma} + \hat{\mu}\hat{\mu}'$. Since

$$\ln\left(|a\mu\mu' + \sigma^2 I_N|\right) = (N-1)\ln(\sigma^2) + \ln(\sigma^2 + a\mu'\mu), \tag{30}$$

we can minimize

$$f(\mu, a, \sigma^2) = (N-1)\ln(\sigma^2) + \ln(\sigma^2 + a\mu'\mu) + \frac{1}{\sigma^2}\left[\operatorname{tr}(\hat{U}) + \frac{\sigma^2(\mu'\mu - 2\hat{\mu}'\mu) - a\mu'\hat{U}\mu}{\sigma^2 + a\mu'\mu}\right] \tag{31}$$

to obtain the ML estimator.

Let $\hat{Q}\hat{\Lambda}\hat{Q}'$ be the spectral decomposition of $\hat{U}$, where $\hat{\Lambda} = \text{Diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_N)$ 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

$$p(\phi) = \sum_{i=1}^{N} \frac{(\hat{\lambda}_i - c)\hat{z}_i^2}{[1 - \phi(\hat{\lambda}_i - c)]^2} = 0 \tag{32}$$

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:

$$g(c) = \ln\left(c - \sum_{i=1}^{N} \frac{\hat{z}_i^2}{1 - \tilde{\phi}(c)(\hat{\lambda}_i - c)}\right) + (N-1)\ln\left(\sum_{i=1}^{N}\hat{\lambda}_i - c\right), \tag{33}$$

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 μ is

$$\hat{\mu}^{\text{MP}} = \tilde{\mu} = \hat{Q}[I_N - \tilde{\phi}(c^*)(\hat{\Lambda} - c^* I_N)]^{-1}\hat{z}, \tag{34}$$

and hence the ML estimators of σ^2 and a are

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^{N}\hat{\lambda}_i - c^*}{N-1}, \qquad \tilde{a} = \frac{c^* - \tilde{\sigma}^2}{\tilde{\mu}'\tilde{\mu}} - 1. \tag{35}$$

Then the MacKinlay and Pástor (2000) portfolio rule is obtained easily. Q.E.D.

[12]We are grateful to Raymond Kan for this semi-analytical solution.

References

Avramov, D., 2002. Stock return predictability and model uncertainty. Journal of Financial Economics 64, 423–458.

Bawa, V.S., Brown, S.J., Klein, R.W., 1979. Estimation Risk and Optimal Portfolio Choice. North-Holland, Amsterdam.

Bawa, V.S., Klein, R.W., 1976. The effect of estimation risk on optimal portfolio choice. Journal of Financial Economics 3, 215–231.

Brandt, M.W., 2009. Portfolio choice problems. In: Aït-Sahalia, Y., Hansen, L. (Eds.), Handbook of Financial Econometrics. Elsevier, North-Holland, Amsterdam. pp. 269–336.

Brandt, M.W., Santa-Clara, P., Valkanov, R., 2009. Parametric portfolio policies: exploiting characteristics in the cross section of equity returns. Review of Financial Studies 22, 3411–3447.

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

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

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

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

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

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

Haff, L.R., 1979. An identity for the Wishart distribution with applications. Journal of Multivariate Analysis 9, 531–544.

Harvey, C.R., Liechty, J., Liechty, M.W., Müller, P., 2004. Portfolio selection with higher moments. Duke University, Working Paper.

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

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

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

Lehmann, E.L., Casella, G., 1998. Theory of Point Estimation. Springer-Verlag, New York.

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

Lo, A.W., Patel, P.N., 2008. 130/30: The new long-only. Journal of Portfolio Management 34, 12–38.

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

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

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

Michaud, R., 2008. Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation (second edition). Oxford University Press, New York.

Pástor, Ľ., 2000. Portfolio selection and asset pricing models. Journal of Finance 55, 179–223.

Pástor, Ľ, Stambaugh, R.F., 2000. Comparing asset pricing models: An investment perspective. Journal of Financial Economics 56, 335–381.

Shao, J., Tu, D., 1996. The Jackknife and Bootstrap. Springer-Verlag, New York.

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

Ter Horst, J., De Roon, F., Werker, B.J.M., 2006. Incorporating estimation risk in portfolio choice. In: Renneboog, L. (Ed.), Advances in Corporate Finance and Asset Pricing. Elsevier, Amsterdam. pp. 449–472.

Tu, J., Zhou, G., 2004. Data-generating process uncertainty: What difference does it make in portfolio decisions? Journal of Financial Economics 72, 385–421.

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

Table 1

Utilities in a one-factor model without mispricing

This table reports the average utilities (annualized and in percentage points) of a mean-variance investor under various investment rules: the true optimal rule, the $1/N$ rule, the ML rule, the Jorion (1986) rule, the MacKinlay and Pástor (2000) rule, the Kan and Zhou (2007) rule, and the four combination rules, with 10,000 sets of sample size T simulated data from a one-factor model with zero mispricing alphas and with $N = 25$ assets. Panels A and B assume that the risk aversion coefficient γ is 3 and 1, respectively.

	T					
Rules	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
ML	-85.72	-25.81	-8.35	-1.61	2.42	3.30
Jorion	-12.85	-3.79	-0.18	1.55	2.98	3.47
MacKinlay-Pástor	2.11	3.00	3.44	3.65	3.79	3.83
Kan-Zhou	-2.15	-0.00	1.13	1.90	2.97	3.47
$\hat{w}^{\mathrm{CML}}$	1.68	2.95	3.42	3.60	3.81	3.90
$\hat{w}^{\mathrm{CPJ}}$	1.42	2.93	3.46	3.71	3.88	3.86
$\hat{w}^{\mathrm{CMP}}$	2.19	3.05	3.48	3.67	3.80	3.83
$\hat{w}^{\mathrm{CKZ}}$	3.71	3.77	3.81	3.85	3.91	3.95
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
ML	-257.16	-77.42	-25.05	-4.83	7.25	9.91
Jorion	-38.55	-11.38	-0.55	4.66	8.95	10.42
MacKinlay-Pástor	6.33	9.00	10.31	10.94	11.37	11.48
Kan-Zhou	-6.44	-0.01	3.38	5.69	8.92	10.40
$\hat{w}^{\mathrm{CML}}$	1.14	4.79	6.39	7.47	9.50	10.62
$\hat{w}^{\mathrm{CPJ}}$	1.28	5.68	6.97	7.11	7.46	10.34
$\hat{w}^{\mathrm{CMP}}$	6.57	9.16	10.49	11.09	10.95	11.43
$\hat{w}^{\mathrm{CKZ}}$	6.36	6.70	6.99	7.41	8.78	9.97

Table 2

Utilities in factor models with mispricing

This table reports the average utilities (annualized and in percentage points) of a mean-variance investor under various investment rules: the true optimal rule, the $1/N$ rule, the ML rule, the Jorion (1986) rule, the MacKinlay and Pástor (2000) rule, the Kan and Zhou (2007) rule, and the four combination rules, with $N = 25$ assets for 10,000 sets of sample size T simulated data from a one-factor model (Panel A) and a three-factor model (Panel B), respectively. The annualized mispricing α's are assumed to spread evenly between -2% to 2%. The risk aversion coefficient γ is 3.

	T					
Rules	120	240	480	960	3000	6000
Panel A: One-factor model						
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
ML	-84.75	-23.84	-6.18	0.65	4.73	5.62
Jorion	-12.36	-2.99	0.95	3.09	5.06	5.71
MacKinlay-Pástor	2.34	3.23	3.67	3.88	4.02	4.06
Kan-Zhou	-2.35	0.02	1.64	3.14	5.06	5.71
$\hat{w}^{\mathrm{CML}}$	2.02	3.32	3.91	4.43	5.38	5.82
$\hat{w}^{\mathrm{CPJ}}$	2.27	3.70	4.02	3.92	4.83	5.72
$\hat{w}^{\mathrm{CMP}}$	2.41	3.27	3.71	3.90	4.02	4.04
$\hat{w}^{\mathrm{CKZ}}$	3.84	3.95	4.12	4.41	5.14	5.62
Panel B: Three-factor model						
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
ML	-81.09	-17.11	1.39	8.52	12.76	13.69
Jorion	-7.85	2.84	7.65	10.45	12.99	13.75
MacKinlay-Pástor	1.78	2.66	3.09	3.30	3.44	3.48
Kan-Zhou	1.61	5.12	7.96	10.45	12.99	13.75
$\hat{w}^{\mathrm{CML}}$	3.84	6.15	8.44	10.63	13.02	13.76
$\hat{w}^{\mathrm{CPJ}}$	5.79	5.36	4.17	9.67	13.02	13.76
$\hat{w}^{\mathrm{CMP}}$	1.86	2.73	3.12	3.30	3.45	3.48
$\hat{w}^{\mathrm{CKZ}}$	5.09	6.06	7.57	9.59	12.56	13.58

Table 3

Sharpe ratios in a one-factor model

This table reports in percentage points the average Sharpe ratios of a mean-variance investor under various investment rules: the true optimal rule, the $1/N$ rule, the ML rule, the Jorion (1986) rule, the MacKinlay and Pástor (2000) rule, the Kan and Zhou (2007) rule, and the four combination rules, 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 annualized mispricing α's are zeros or between -2% to 2%, respectively.

	T					
Rules	120	240	480	960	3000	6000
Panel A: α=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
ML	3.88	5.59	7.54	9.54	12.19	13.18
Jorion	4.54	6.46	8.40	10.18	12.38	13.24
MacKinlay-Pástor	12.19	13.51	13.86	13.89	13.89	13.89
Kan-Zhou	4.97	7.03	8.80	10.27	12.34	13.24
$\hat{w}^{\text{CML}}$	12.04	12.88	13.34	13.53	13.83	13.98
$\hat{w}^{\text{CPJ}}$	10.40	12.36	13.22	13.67	13.94	13.90
$\hat{w}^{\text{CMP}}$	12.07	13.44	13.87	13.90	13.89	13.89
$\hat{w}^{\text{CKZ}}$	13.70	13.79	13.86	13.91	14.00	14.07
Panel B: α 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
ML	5.92	8.34	10.94	13.32	16.06	16.97
Jorion	5.61	8.03	10.69	13.16	16.03	16.95
MacKinlay-Pástor	12.70	13.98	14.28	14.30	14.31	14.31
Kan-Zhou	4.77	7.15	10.09	12.97	16.02	16.95
$\hat{w}^{\text{CML}}$	12.81	13.69	14.30	15.02	16.45	17.09
$\hat{w}^{\text{CPJ}}$	11.64	13.73	14.31	14.12	15.60	16.96
$\hat{w}^{\text{CMP}}$	12.52	13.89	14.26	14.28	14.27	14.25
$\hat{w}^{\text{CKZ}}$	14.02	14.23	14.54	15.04	16.21	16.91

Table 4

Standard errors of utilities 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. The annualized mispricing α's are assumed to spread evenly between -2% to 2%. The risk aversion coefficient γ is 3.

	T					
Rules	120	240	480	960	3000	6000
Panel A: Standard errors of utilities						
True	0	0	0	0	0	0
1/N	0	0	0	0	0	0
ML	12.37	3.29	1.24	0.53	0.15	0.08
Jorion	3.55	1.26	0.61	0.33	0.13	0.07
MacKinlay-Pástor	0.75	0.36	0.17	0.09	0.03	0.01
Kan-Zhou	1.44	0.72	0.50	0.32	0.13	0.07
$\hat{w}^{\text{CML}}$	1.24	0.62	0.47	0.31	0.13	0.07
$\hat{w}^{\text{CPJ}}$	0.49	0.37	0.31	0.70	0.13	0.07
$\hat{w}^{\text{CMP}}$	0.67	0.33	0.16	0.08	0.03	0.01
$\hat{w}^{\text{CKZ}}$	0.29	0.35	0.40	0.36	0.17	0.08
Panel B: Standard errors of Sharp Ratios						
True	0	0	0	0	0	0
1/N	0	0	0	0	0	0
ML	4.02	3.01	2.00	1.19	0.43	0.22
Jorion	4.18	2.87	1.84	1.12	0.42	0.22
MacKinlay-Pástor	6.87	3.54	1.16	0.27	0.04	0.03
Kan-Zhou	4.51	2.82	1.90	1.16	0.42	0.22
$\hat{w}^{\text{CML}}$	2.44	2.22	1.85	1.12	0.42	0.22
$\hat{w}^{\text{CPJ}}$	2.57	2.25	2.28	3.22	0.43	0.22
$\hat{w}^{\text{CMP}}$	6.36	3.38	0.80	0.07	0.04	0.03
$\hat{w}^{\text{CKZ}}$	1.66	1.88	1.84	1.30	0.45	0.23

Table 5

Combination coefficients

This table reports in percentage points the true optimal combination coefficients, the average estimated optimal combination coefficients and their standard errors (in parentheses) for the four combination strategies. The data are simulated in the same way as in Table 4.

	T					
Parameters	120	240	480	960	3000	6000
Panel A: w^{CML}						
δ	15.74	29.93	47.57	65.12	85.60	92.27
$\hat{\delta}$	20.56	29.38	45.16	63.73	85.35	92.20
	(10.87)	(13.44)	(12.49)	(7.61)	(2.05)	(0.80)
Panel B: w^{CPJ}						
δ_j	27.56	46.65	65.29	80.32	93.90	97.17
$\hat{\delta}_j$	35.95	17.61	11.52	87.90	100.00	100.00
	(12.78)	(16.41)	(29.55)	(32.63)	(0.00)	(0.00)
Panel C: w^{CMP}						
δ_m	28.50	42.54	56.06	66.62	76.25	78.93
$\hat{\delta}_m$	97.02	97.27	98.03	99.37	100.00	100.00
	(1.14)	(0.89)	(0.87)	(0.75)	(0.00)	(0.00)
Panel D: w^{CKZ}						
δ_k	53.78	68.09	79.87	88.35	95.81	97.84
$\hat{\delta}_k$	56.18	57.37	63.49	72.26	86.43	92.27
	(6.37)	(7.70)	(7.88)	(6.49)	(3.09)	(1.52)

Table 6

Certainty-equivalent returns based on the real data sets

This table reports the certainty-equivalent returns (annualized and in percentage points) of a mean-variance investor under various investment rules: the true optimal rule, the $1/N$ rule, the ML rule, the Jorion (1986) rule, the MacKinlay and Pástor (2000) rule, the Kan and Zhou (2007) rule, and the four combination rules. While the in-sample ML rule uses all the data for its estimation, the other estimated rules are based on a rolling sample with an estimation window $M = 120$ (Panel A) or 240 (Panel B). The real data sets are the five data sets used by DeMiguel, Garlappi, and Uppal (2009), 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 γ 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	5.92	5.51	5.14
ML	-38.18	-18.30	4.90	-100.69	-128.59	-194.33	-1173.78
Jorion	-9.21	-5.80	9.51	0.82	1.99	-20.72	-152.10
MacKinlay-Pástor	-0.76	0.86	-0.20	0.47	0.37	1.02	1.45
Kan-Zhou	-3.59	-3.42	9.51	20.75	22.01	9.15	-17.77
$\hat{w}^{\mathrm{CML}}$	-1.39	-0.34	6.39	22.25	26.06	14.62	-6.40
$\hat{w}^{\mathrm{CPJ}}$	3.15	1.74	4.52	6.39	11.10	6.77	-1.20
$\hat{w}^{\mathrm{CMP}}$	2.21	2.26	2.64	3.31	3.54	3.67	3.57
$\hat{w}^{\mathrm{CKZ}}$	3.02	1.79	8.54	28.97	29.35	19.36	8.51
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	4.95	5.09	5.48
ML	-14.30	-6.94	12.08	-5.10	-38.63	-20.80	-158.40
Jorion	-0.76	-1.38	12.40	23.15	10.56	10.44	-18.70
MacKinlay-Pástor	2.84	-0.02	0.44	2.78	2.67	3.37	4.32
Kan-Zhou	1.89	-0.17	12.21	26.60	19.61	14.08	12.43
$\hat{w}^{\mathrm{CML}}$	4.58	0.29	11.96	18.73	18.97	16.70	6.29
$\hat{w}^{\mathrm{CPJ}}$	4.19	0.07	12.40	-19.01	-8.99	7.38	14.55
$\hat{w}^{\mathrm{CMP}}$	4.11	0.49	2.20	3.71	3.88	4.31	4.95
$\hat{w}^{\mathrm{CKZ}}$	5.40	0.88	11.03	26.84	30.25	20.09	16.28