On the Economic Value of Alphas

RAYMOND KAN and XIAOLU WANG*

This version: March 2015

*Kan is from the University of Toronto and Wang is from Iowa State University. Kan gratefully acknowledges financial support from the National Bank Financial of Canada and the Social Sciences and Humanities Research Council of Canada. Earlier version of the paper was circulated under the title "Seeking Positive Alpha." We thank Robert Grauer, Kevin Wang, Jason Wei, Jialin Yu, Chu Zhang, Guofu Zhou, seminar participants at Central University of Finance and Economics, Hong Kong Polytechnic University, Hong Kong University, Tsinghua University, University of Toronto, University of Wisconsin at Madison, Wilfrid Laurier University, and participants at the 2011 Asian Finance Association Annual Meeting and 2014 HKUST Finance Symposium on Asset Pricing for helpful discussions and comments. Corresponding author: Raymond Kan, Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario, Canada M5S 3E6; Tel: (416) 978-4291; Fax: (416) 978-5433; Email: kan@chass.utoronto.ca.

On the Economic Value of Alphas

Abstract

In this paper, we examine the benefit of incorporating test assets with nonzero alphas into an optimal portfolio when the mean and covariance matrix of asset returns are estimated with errors. Under the normality assumption, we derive the distribution of out-of-sample return of a portfolio that is optimized based on sample mean and covariance matrix. We show that as long as the benchmarks are not *ex ante* efficient, this sample optimal portfolio will generate positive alpha relative to the benchmarks. However, due to estimation errors, we need a very long estimation window for the sample optimal portfolio to outperform the benchmarks. We further consider a strategy that optimally combines the risk-free asset, the sample optimal portfolio, and the sample optimal portfolio based on just the benchmarks. This combining strategy consistently outperforms the benchmarks, providing a reliable way to realize the economic value of nonzero alphas.

I. Introduction

Since the seminal work of Jensen (1968), alpha has rapidly gained its popularity, and has become a widely used measure for performance evaluation, both in academic research and in practice. Today, in addition to the classic CAPM alpha, alphas with respect to many other asset pricing models, such as the Fama-French (1993) three-factor model and the Carhart (1997) four-factor model, are often used. Empirical asset pricing studies often find nonzero alphas in various test assets. These findings are typically interpreted as a rejection of the asset pricing model based on the benchmark portfolios. However, from the perspective of an investor, the more relevant question is how to improve the investment performance over the benchmark portfolios by using the test assets.

When the mean and the covariance matrix of the asset returns are known, one can improve the maximum Sharpe ratio of the benchmark portfolios by incorporating test assets with nonzero alphas. Specifically, Dybvig and Ross (1985) show that if an asset has a positive alpha relative to a benchmark, then buying some of the asset at the margin will result in a higher Sharpe ratio than that of the benchmark. In reality, the true mean and covariance matrix are unknown and need to be estimated. Therefore, the optimal portfolio constructed using the estimated mean and covariance matrix (which we call the *sample optimal portfolio*) contains estimation errors, and the benefit of incorporating the test assets into the optimal portfolio becomes unclear.[1]

In this paper, we study this problem by asking two related questions. First, given estimation errors, we examine whether the sample optimal portfolio will have positive alpha relative to the benchmarks and whether the sample optimal portfolio can outperform the benchmarks. Under the normality assumption, we derive the distribution of the excess return of the sample optimal portfolio. With this distribution, we show that as long as the benchmarks are not *ex ante* efficient, the sample optimal portfolio will generate positive unconditional alpha. In addition, we analyze the distribution of the t-ratio of the alpha of

[1]Brown (1976), Bawa and Klein (1976), Bawa, Brown, and Klein (1979), Jorion (1986), Pástor (2000), Pástor and Stambaugh (2000), MacKinlay and Pástor (2000), Tu and Zhou (2004), Kan and Zhou (2007), Tu and Zhou (2011) are examples of studies focusing on portfolio rules with parameter uncertainty. Pástor (2000) and Pástor and Stambaugh (2000) also take into account the benchmark portfolios while studying portfolio selection with parameter uncertainty.

the sample optimal portfolio and show that it is not unusual to find statistically significant positive alpha. However, due to estimation errors, the sample optimal portfolio does not always outperform the optimal portfolio based only on the benchmarks (which we call the *benchmark sample optimal portfolio*). In most cases, we need a very long estimation window in order for the sample optimal portfolio to outperform the benchmark sample optimal portfolio.

Given the poor performance of the sample optimal portfolio, the second question we ask is whether a combining rule — combining the risk-free asset, the sample optimal portfolio, and the benchmark sample optimal portfolio — can improve the out-of-sample performance and provide a reliable strategy to outperform the benchmarks.[2] When the parameters are known, the best strategy is to hold only the optimal portfolio and there is no point in considering a combining strategy. When the parameters are unknown, the estimation errors lower the benefit of including the test assets and we derive analytically the optimal weights on the risk-free asset, the sample optimal portfolio, and the benchmark sample optimal portfolio. As long as the benchmarks are not *ex ante* efficient, the optimal combining strategy always has positive weight on the sample optimal portfolio. In addition, we show that the optimal combining strategy consistently outperforms the benchmarks. Therefore, the optimal combining portfolio provides a reliable strategy to realize the economic value of nonzero alphas.

We check the robustness of our results to the departure from the assumption that excess returns of risky assets are i.i.d. multivariate normal. Specifically, we examine two alternative distributional assumptions: a multivariate t-distribution with five degrees of freedom and an empirical distribution. Test results suggest that our conclusions are robust to alternative distributional assumptions.

The remainder of the paper is organized as follows. Section II derives the distribution of out-of-sample return and conditional alpha of the sample optimal portfolio. Section III presents the optimal combining portfolio, and examines the performance of this combining strategy. Section IV investigates the robustness of our results under alternative distributional assumptions. Section V concludes and discusses future research opportunities.

[2]Jorion (1986), Kan and Zhou (2007), Tu and Zhou (2011) consider some alternative three-fund rules.

II. The Sample Optimal Portfolio

A. The Setting

Consider a portfolio choice problem of an investor in a universe with a risk-free asset, K benchmark portfolios, and N test assets. Let $r_t = [r'_{1,t},\ r'_{2,t}]'$, where $r_{1,t}$ and $r_{2,t}$ are the excess returns (in excess of risk-free rate) of the benchmark portfolios and the test assets at time t, respectively. We assume that r_t follows a multivariate normal distribution and is independent and identically distributed (i.i.d.) over time with mean

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \tag{1}$$

and covariance matrix

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}. \tag{2}$$

The investor is assumed to choose a portfolio q in order to maximize the mean-variance utility function

$$U_q = \mu_q - \frac{\gamma}{2}\sigma_q^2, \tag{3}$$

where γ is the investor's risk aversion coefficient, and μ_q and σ_q^2 are the mean and variance of portfolio q.

When the benchmarks are *ex ante* efficient, the N test assets will have zero alphas, i.e.,

$$\alpha = \mu_2 - V_{21}V_{11}^{-1}\mu_1 = 0_N. \tag{4}$$

The optimal portfolio consists of just the K benchmarks with weights

$$w_{s^*} = \frac{1}{\gamma}V_{11}^{-1}\mu_1. \tag{5}$$

We call this portfolio the benchmark optimal portfolio and denote it as s^*. The utility of holding portfolio s^* is

$$U_{s^*} = \frac{\theta_1^2}{2\gamma}, \tag{6}$$

where $\theta_1^2 = \mu_1' V_{11}^{-1}\mu_1$ is the squared Sharpe ratio of the *ex ante* tangency portfolio of the K benchmarks.

When the benchmarks are inefficient, the N test assets will have nonzero alphas

$$\alpha = \mu_2 - V_{21}V_{11}^{-1}\mu_1 \neq 0_N. \tag{7}$$

It is possible for the investor to improve the utility by including the N test assets into the portfolio. The optimal portfolio, p^*, has weights

$$w_{p^*} = \frac{1}{\gamma}V^{-1}\mu \tag{8}$$

on the $M \equiv N + K$ assets. The utility of holding portfolio p^* is

$$U_{p^*} = \frac{\theta^2}{2\gamma}, \tag{9}$$

where $\theta^2 = \mu' V^{-1}\mu$ is the squared Sharpe ratio of the *ex ante* tangency portfolio of the M assets. The utility improvement by including the N test assets is

$$U_{p^*} - U_{s^*} = \frac{\delta^2}{2\gamma}, \tag{10}$$

where[3]

$$\delta^2 = \theta^2 - \theta_1^2 = \mu' V^{-1}\mu - \mu_1' V_{11}^{-1}\mu_1 = \alpha'\Sigma^{-1}\alpha, \tag{11}$$

and $\Sigma = V_{22} - V_{21}V_{11}^{-1}V_{12}$. In addition, it can be readily shown that the alpha of portfolio p^* relative to the K benchmark portfolios is

$$\alpha_{p^*} = \frac{\delta^2}{\gamma}. \tag{12}$$

Therefore, the existence of nonzero alphas for the test assets indicates that p^* has positive alpha and nonzero weights in the test assets.

However, μ and V are unknown in practice. Therefore, portfolios s^* and p^* are unattainable to the investor. We assume the investor estimates μ and V using a window of h months of historical data of excess returns, and the estimates of μ and V at time t are given by

$$\hat{\mu}_t = \begin{bmatrix} \hat{\mu}_{1,t} \\ \hat{\mu}_{2,t} \end{bmatrix} = \frac{1}{h}\sum_{\tau=t-h+1}^{t} r_\tau, \tag{13}$$

$$\hat{V}_t = \begin{bmatrix} \hat{V}_{11,t} & \hat{V}_{12,t} \\ \hat{V}_{21,t} & \hat{V}_{22,t} \end{bmatrix} = \frac{1}{h}\sum_{\tau=t-h+1}^{t} (r_\tau - \hat{\mu}_t)(r_\tau - \hat{\mu}_t)'. \tag{14}$$

[3] See Jobson and Korkie (1982) and Gibbons, Ross, and Shanken (1985).

Natural estimators of w_{s^*} and w_{p^*} are the sample counterparts of (5) and (8):

$$w_{s,t} = \frac{1}{\gamma}\hat{V}_{11,t}^{-1}\hat{\mu}_{1,t}, \tag{15}$$

$$w_{p,t} = \frac{1}{\gamma}\hat{V}_t^{-1}\hat{\mu}_t. \tag{16}$$

We call these two portfolios as the benchmark sample optimal portfolio and the sample optimal portfolio, and denote them as portfolio s and portfolio p, respectively. Given estimation errors, we will examine whether portfolio p also has positive alpha with respect to the benchmarks, and more importantly, whether the investor is better off by holding portfolio p instead of portfolio s.

Note that $w_{s,t}$ and $w_{p,t}$ are random variables as functions of the historical returns data, and therefore, the conditional mean and variance of portfolios s and p are also random variables. The conditional out-of-sample utility of portfolio s and p at time t are given by

$$U_{s,t} = w_{s,t}'\mu_1 - \frac{\gamma}{2}w_{s,t}'V_{11}w_{s,t}, \tag{17}$$

$$U_{p,t} = w_{p,t}'\mu - \frac{\gamma}{2}w_{p,t}'Vw_{p,t}. \tag{18}$$

It is natural then to evaluate a portfolio rule q with random weights $w_{q,t}$ based on its expected out-of-sample utility[4]

$$E[U_q] = E\left[w_{q,t}'\mu - \frac{\gamma}{2}w_{q,t}'Vw_{q,t}\right]. \tag{19}$$

B. Distribution of Out-of-Sample Returns of the Sample Optimal Portfolio

Since there are estimation errors in $\hat{\mu}_t$ and $\hat{V}_t$, the distribution of the out-of-sample return of the sample optimal portfolio p could significantly differ from that of the true optimal portfolios p^*. In this subsection, we obtain the distribution of the out-of-sample return of the sample optimal portfolio p.

Conditional on portfolio weights $w_{p,t}$, the excess return of the sample optimal portfolio

[4]Brown (1976), Jorion (1986), Frost and Savarino (1986), Stambaugh (1997), Ter Horst, De Roon, and Werkerzx (2006), Kan and Zhou (2007), and Tu and Zhou (2011) are examples of using expected out-of-sample utility to evaluate portfolio rules.

at time $t+1$ is given by

$$r_{p,t+1} = w'_{p,t} r_{t+1}. \tag{20}$$

Under the normality assumption on r_{t+1}, $r_{p,t+1}$ is conditionally normally distributed with conditional mean and variance given by

$$\mu_{p,t} = w'_{p,t}\mu, \tag{21}$$

$$\sigma^2_{p,t} = w'_{p,t} V w_{p,t}. \tag{22}$$

Note that although $r_{p,t+1}$ is normally distributed when condtional on $w_{p,t}$, the unconditional distribution of $r_{p,t+1}$ is not normally distributed. In the following Proposition, we present the unconditional distribution of $r_{p,t+1}$.

Proposition 1: *Let $x_1 \sim N(0,1)$, $z_1 \sim N(\sqrt{h}\theta, 1)$, $u_0 \sim \chi^2_{M-1}$, $u_1 \sim \chi^2_{h-M}$, $u_2 \sim \chi^2_{h-M+1}$, $u_3 \sim \chi^2_{M-2}$, and they are independent of each other. When $h > M > 1$, the joint distribution of $\mu_{p,t}$ and $\sigma^2_{p,t}$ can be obtained using*

$$\mu_{p,t} = \frac{\sqrt{h}\theta}{\gamma u_1}\left(z_1 + \frac{x_1\sqrt{u_0}}{\sqrt{u_2}}\right), \tag{23}$$

$$\sigma^2_{p,t} = \frac{h(z_1^2 + u_0)}{\gamma^2 u_1^2}\left(1 + \frac{x_1^2 + u_3}{u_2}\right). \tag{24}$$

When $M = 1$, we have

$$\mu_{p,t} = \frac{\sqrt{h}\theta z_1}{\gamma u_1}, \qquad \sigma^2_{p,t} = \frac{h z_1^2}{\gamma^2 u_1^2}. \tag{25}$$

The out-of-sample excess returns of portfolio p can be simulated using

$$r_{p,t+1} = \mu_{p,t} + \sigma_{p,t} y, \tag{26}$$

where $y \sim N(0,1)$ and it is independent of x_1, z_1, u_0, u_1, u_2, and u_3.

Proposition 1 shows that although $w_{p,t}$ is a function of $\hat{\mu}_t$ and $\hat{V}_t$, the joint distribution of $\mu_{p,t}$ and $\sigma^2_{p,t}$ only depends on two normal random variables (x_1 and z_1) and four central chi-squared random variables (u_0 to u_3). In the proof of Proposition 1, we can see that z_1 and u_0 are due to the randomness of $\hat{\mu}_t$ and x_1, u_1, u_2, and u_3 are due to the randomness of $\hat{V}_t$. In addition, Proposition 1 reveals that the distribution of $r_{p,t+1}$ is completely determined by γ,

h, M, and θ. There is no need to specify μ and V for us to obtain the distribution of $r_{p,t+1}$. Therefore, Proposition 1 provides us a speedy approach for simulating the distribution of $r_{p,t+1}$ without the need to simulate the excess returns on the M assets. By replacing M with K and θ with θ_1, Proposition 1 also allows us to obtain the unconditional distribution of $r_{s,t+1}$.

In Figure 1, we plot the density function of the excess return of the sample optimal portfolio p together with that of the true optimal portfolio p^* ($h = \infty$) for different combinations of h (60 or 120) and M (13 or 28), assuming that the investor has a relative risk aversion of 10 and the tangency portfolio of the M assets has a Sharpe ratio of $\theta = 0.3$. When comparing with the distribution of the true optimal portfolio (which is normally distributed), the excess return of the sample optimal portfolio exhibits a higher volatility, especially when the length of the estimation window is short and when there are more risky assets. This is because with short estimation window or more risky assets, there are more estimation errors, which makes $r_{p,t+1}$ more volatile.

Figure 1 about here

Using the results of Proposition 1, we can obtain the noncentral and central moments of $r_{p,t+1}$. The first four noncentral moments are given by[5]

$$E[r_{p,t+1}] = E[\mu_{p,t}] = \frac{h\theta^2}{\gamma(h-M-2)} \quad \text{for } h > M+2, \tag{27}$$

$$\begin{aligned} E[r_{p,t+1}^2] &= E[\mu_{p,t}^2] + E[\sigma_{p,t}^2] \\ &= \frac{h^2(h-M-1)\theta^4 + h(h-2)[(h+1)\theta^2 + M]}{\gamma^2(h-M-1)(h-M-2)(h-M-4)} \quad \text{for } h > M+4, \end{aligned} \tag{28}$$

$$\begin{aligned} E[r_{p,t+1}^3] &= E[\mu_{p,t}^3] + 3E[\mu_{p,t}\sigma_{p,t}^2] \\ &= \frac{h^3(h-M-1)\theta^6 + 3h^2(h-2)\theta^2[(h+1)\theta^2 + M + 2]}{\gamma^3(h-M-1)(h-M-2)(h-M-4)(h-M-6)} \quad \text{for } h > M+6, \end{aligned} \tag{29}$$

$$\begin{aligned} E[r_{p,t+1}^4] &= E[\mu_{p,t}^4] + 6E[\mu_{p,t}^2\sigma_{p,t}^2] + 3E[\sigma_{p,t}^4] \\ &= \frac{h^4(h-M-1)(h-M-3)\theta^8}{\gamma^4(h-M-4)_4(h-M-6)(h-M-8)} \end{aligned}$$

[5]Using the inverse first and second moments of central Wishart distributions, Kan and Zhou (2007) obtain $E[\mu_{p,t}]$ and $E[\sigma_{p,t}^2]$. However, it is difficult to generalize their method to obtain the higher order moments of $r_{p,t+1}$.

$$+\frac{6h^3(h-2)(h-M-3)[(h+1)\theta^2+M+4]\theta^4}{\gamma^4(h-M-4)_4(h-M-6)(h-M-8)}$$
$$+\frac{3h^2(h-2)(h-4)\left[((h+1)\theta^2+M+2)^2-2(M+2)\right]}{\gamma^4(h-M-4)_4(h-M-6)(h-M-8)} \quad \text{for } h>M+8, \quad (30)$$

where $(a)_r = a(a+1)\cdots(a+r-1)$.

With the first four noncentral moments of $r_{p,t+1}$ available, the variance and the third and fourth central moments of $r_{p,t+1}$ can be obtained as

$$\mathrm{Var}[r_{p,t+1}] = E[r_{p,t+1}^2] - E[r_{p,t+1}]^2 \text{ for } h>M+4, \quad (31)$$

$$E[(r_{p,t+1}-E[r_{p,t+1}])^3] = E[r_{p,t+1}^3] - 3E[r_{p,t+1}^2]E[r_{p,t+1}] + 2E[r_{p,t+1}]^3 \text{ for } h>M+6, \quad (32)$$

$$E[(r_{p,t+1}-E[r_{p,t+1}])^4] = E[r_{p,t+1}^4] - 4E[r_{p,t+1}^3]E[r_{p,t+1}] + 6E[r_{p,t+1}^2]E[r_{p,t+1}]^2 - 3E[r_{p,t+1}]^4 \text{ for } h>M+8. \quad (33)$$

With the expressions of the unconditional mean $\mu_p \equiv E[r_{p,t+1}]$ and variance $\sigma_p^2 \equiv \mathrm{Var}[r_{p,t+1}]$ of the sample optimal portfolio available, we can easily compute the unconditional Sharpe ratio (denoted as $\theta_p = \mu_p/\sigma_p$) of the sample optimal portfolio. In Figure 2, we plot θ_p/θ as a function of h for four different combinations of θ (0.2 and 0.3) and M (13 and 28). Figure 2 clearly shows that the estimation errors of $\hat{\mu}_t$ and $\hat{V}_t$ significantly reduce the Sharpe ratio of the sample optimal portfolio. This is particularly the case when θ is small and M is large. When $\theta = 0.2$ and $M = 28$, θ_p is still less than 67% of θ even using a very long estimation window of $h = 600$.

Figure 2 about here

From Figure 1, we observe that while $r_{p,t+1}$ is not normally distributed, its distribution is almost symmetric. To understand whether this is a general property of $r_{p,t+1}$, we plot the coefficient of skewness of $r_{p,t+1}$ as a function of h for different combinations of θ (0.2 or 0.3) and M (13 or 28) in Figure 3. The figure suggests that while the excess return of the sample optimal portfolio has positive skewness, its coefficient of skewness is tiny even when the estimation window is short. The observation that $r_{p,t+1}$ does not have much skewness is important because it shows that the sample optimal portfolio is not a dynamic trading strategy that aims at replicating the payoff of an option.

Figure 3 about here

In Figure 4, we show the corresponding plot of coefficient of excess kurtosis of $r_{p,t+1}$ as a function of h. Unlike the coefficient of skewness, the coefficient of excess kurtosis of $r_{p,t+1}$ is rather large, especially when h is small. The positive excess kurtosis indicates that $r_{p,t+1}$ has a fatter tail than the normal distribution. If investors do not like return distributions that have fat tails, then the sample optimal portfolio will be less desirable due to the excess kurtosis introduced by the estimation errors on the optimal weights.

Figure 4 about here

C. Distribution of Conditional Alpha of the Sample Optimal Portfolio

Under our i.i.d. setting, the conditional alpha and beta of the true optimal portfolio p^* are constant over time because w_{p^*} is constant over time. However, for the sample optimal portfolio, its weights $w_{p,t}$ are time varying, so its conditional alpha and beta are no longer constant over time. The conditional alpha of the sample optimal portfolio at time t is given by

$$\alpha_{p,t} = w_{p,t}' \begin{bmatrix} 0_K \\ \alpha \end{bmatrix} = \frac{1}{\gamma}\hat{\mu}_t'\hat{V}_t^{-1} \begin{bmatrix} 0_K \\ \alpha \end{bmatrix}. \tag{34}$$

In the following Proposition, we provide a simplification of the distribution of the conditional alpha for the sample optimal portfolio.

Proposition 2: *Let $y \sim N(0,1)$, $u_0 \sim \chi^2_{M-1}(h\theta_1^2)$, $u_1 \sim \chi^2_{h-M}$, $u_2 \sim \chi^2_{h-M+1}$, and they are independent of each other. When $h > M > 1$, the distribution of $\alpha_{p,t}$ can be generated using*

$$\alpha_{p,t} = \frac{\sqrt{h}\delta}{\gamma u_1}\left[\sqrt{h}\delta + \left(1 + \frac{u_0}{u_2}\right)^{\frac{1}{2}} y\right]. \tag{35}$$

Proposition 2 reveals that the distribution of $\alpha_{p,t}$ depends only on γ, h, M, θ, and θ_1. When $\theta = \theta_1$ (i.e., $\delta = 0$), the benchmark portfolios are *ex ante* efficient, and $\alpha_{p,t} = 0$. Otherwise, $\alpha_{p,t}$ is nonzero.

$\boxed{\text{Figure 5 about here}}$

In Figure 5, we plot the density function of $\alpha_{p,t}$ for an investor with $\gamma = 10$. We assume $\theta_1 = 0.1$ and $\theta = 0.3$, so the benchmark portfolios are not *ex ante* efficient. Density functions for different combinations of h (60 or 120) and M (13 or 28) are plotted in the figure. Figure 5 shows that there is very high likelihood that $\alpha_{p,t}$ is positive. However, due to estimation errors, unlike the alpha of the true optimal portfolio (α_{p^*}) which is always positive, $\alpha_{p,t}$ can take both positive and negative values. The probability for $\alpha_{p,t}$ to be negative is non-trivial when h is small and M is large. In addition, Figure 5 shows that the distribution of $\alpha_{p,t}$ is more volatile when there are more risky assets ($M = 28$) and when the estimation window is short ($h = 60$) due to more estimation errors.

In general, the unconditional alpha of a portfolio is not equal to the expectation of the conditional alpha of the portfolio. However, under our setting, the conditional mean and variance of the benchmark portfolios are constant over time. As a result, the unconditional alpha of the sample optimal portfolio is indeed the same as the expectation of its conditional alpha. Therefore, the unconditional alpha of the sample optimal portfolio is

$$\alpha_p = E[\alpha_{p,t}] = \frac{h\delta^2}{\gamma(h - M - 2)} = \frac{h\alpha_{p^*}}{h - M - 2} \text{ when } h > M + 2. \tag{36}$$

This result shows that estimation errors of the optimal weights do not destroy the unconditional alpha of the sample optimal portfolio. As long as the benchmark portfolios are not *ex ante* efficient (i.e., $\delta > 0$), the unconditional alpha of the sample optimal portfolio is positive. In fact, the unconditional alpha of the sample portfolio is actually larger than that of the true optimal portfolio α_{p^*}. This is due to the fact that

$$E[w_{p,t}] = \frac{1}{\gamma}E[\hat{V}_t^{-1}\hat{\mu}_t] = \frac{h}{(h - M - 2)\gamma}V^{-1}\mu = \frac{h}{h - M - 2}w_{p^*}, \tag{37}$$

so the sample optimal portfolio on average invests more in the risky assets than the true optimal portfolio, which explains why $\alpha_p > \alpha_{p^*}$.

Although we have established that the sample optimal portfolio has positive unconditional alpha, we have not established that the t-ratio of the estimated alpha can be significant. The alpha of a portfolio depends on how much leverage is used in the portfolio, so it is rather difficult to directly compare alphas across portfolios. In contrast, the t-ratio of estimated alpha is independent of the degree of leverage used in the portfolio, so comparison of t-ratios across different portfolios is more meaningful.

In addition, we can interpret t-ratio of estimated alpha as a measure of value-added to the benchmark portfolios. In order to see that, we cite the following well known result (which is a special case of (11)):

$$\theta_*^2 = \theta_1^2 + AR_q^2, \tag{38}$$

where θ_* is the Sharpe ratio of the tangency portfolio from combining the benchmark portfolios with a given portfolio q, and

$$AR_q = \frac{\alpha_q}{\sigma(\epsilon_q)}, \tag{39}$$

is the appraisal ratio of portfolio q, with $\sigma(\epsilon_q)$ being the standard deviation of the residuals of regressing excess returns of portfolio q on excess returns of the benchmark portfolios. From (38), we can see that the appraisal ratio of a portfolio tells us how much it can help to improve the Sharpe ratio of the benchmark portfolios. It is straightforward to show that the t-ratio of estimated alpha of a portfolio is proportional to the sample measure of its appraisal ratio, so comparing t-ratios of estimated alphas across portfolios is the same as comparing their sample appraisal ratios.

Figure 6 about here

It is rather difficult to obtain the exact distribution of t-ratio of estimated alpha of the sample optimal portfolio, $t(\hat{\alpha}_p)$, because of overlapping estimation windows. However, it can be readily shown that the distribution of $t(\hat{\alpha}_p)$ depends only on h, T, K, N, θ_1, and θ. With the choice of these six parameters, we can readily simulate r_t for $t = 1, \ldots T$ to construct a time series of $r_{p,t}$ for $t = h+1, \ldots, T$. We can then run a regression of $r_{p,t}$ on a constant term and $r_{1,t}$ to obtain $t(\hat{\alpha}_p)$. We assume $K = 3$, $T = 990$, $\theta_1 = 0.1$ and $\theta = 0.3$ in our simulation experiment. Using 1,000,000 simulations, we plot in Figure 6 the density function of $t(\hat{\alpha})$. Density functions for different combinations of h (60 or 120) and M (13 or 28) are plotted in the figure. Figure 6 shows that it is not difficult to find statistically significant alpha for the sample optimal portfolio as there is a fairly high probability for the t-ratio to be above two. In addition, the figure shows that the distribution of $t(\hat{\alpha}_p)$ has higher mean for larger h and smaller M. This is because with larger h and smaller M, there is less estimation error on $\hat{\alpha}_p$, which then leads to a larger $t(\hat{\alpha}_p)$.

Finally, we examine the alphas of sample optimal portfolio empirically using monthly return data over the period of 1931/7 to 2013/12. We consider five different sets of test assets. They are (1) 10 portfolios sorted by momentum, (2) 10 portfolios sorted by volatility, (3) 10 portfolios sorted by idiosyncratic volatility, (4) 10 short-term reversal portfolios sorted based on returns of prior month, and (5) 5×5 size and book-to-market ranked portfolios. The return data for the 10 momentum portfolios, 10 short-term reversal portfolios, and 5×5 size and book-to-market ranked portfolios are obtained from Ken French's website. For the volatility portfolios, we take all the common stocks in the combined NYSE/AMEX/NASDAQ at the end of each month and sort them into ten deciles based on the volatility computed using daily returns in the month. For each volatility portfolio, we compute its value-weighted return for the next month and the portfolio is rebalanced every month. The idiosyncratic volatility portfolios are formed in a similar fashion except the sorting of stocks is based on their idiosyncratic volatility with respect to the value-weighted market portfolio using daily returns in the month.[6]

Table I about here

Table II about here

Table I reports the estimated alphas (in percentage points) and their t-ratios (OLS and White) of the sample optimal portfolio with respect to four asset pricing models with an estimation window of $h = 120$, assuming $\gamma = 10$. The four asset pricing models that we consider are: (1) CAPM with value-weighted market index, (2) CAPM with equal-weighted market index, (3) Fama-French three-factor model, and (4) Carhart four-factor model.[7] The

[6]Ang, Hodrick, Xing, and Zhang (2006) show that high idiosyncratic volatility leads to low future stock returns. Unlike Ang et al., which compute the idiosyncratic volatility with respect to three Fama-French factors, we use just the value-weighted market portfolio because daily Fama-French factors are not available before 1963.

[7]The equal-weighted market index is on the combined NYSE/AMEX/NASDAQ, and its return data are obtained from the CRSP. All other factors are obtained from Ken French's website.

Sharpe ratios of the sample optimal portfolios are also presented in the table. In Table II, we repeat the same analysis with a shorter estimation window of $h = 60$. Tables I and II show that all alphas of sample optimal portfolios are positive with significant t-ratios, regardless of the test assets used and the length of estimation window. The sample Sharpe ratio shows variation across different sets of test assets, and a higher Sharpe ratio is often associated with a larger alpha. In addition, we can see that the magnitude of alpha is larger when the estimation window is short ($h = 60$) but the t-ratio is larger when the estimation window is long ($h = 120$), consistent with our previous discussion.

D. Performance Comparison of the Sample Optimal Portfolio and the Benchmark Sample Optimal Portfolio

In the previous subsection, we show that estimation errors do not destroy the positive alpha of the sample optimal portfolio, and it is not difficult to find statistically significant alpha for the sample optimal portfolio. In this subsection, we examine whether the positive alpha can lead to better performance, that is, whether the investor is better off by holding the sample optimal portfolio p instead of the benchmark sample optimal portfolio s.

Using the results from Proposition 1, we obtain the expected out-of-sample performance of portfolios p and s in the following Proposition.

Proposition 3: *The expected out-of-sample utility of the sample optimal portfolio p and the benchmark sample optimal portfolio s are given by*

$$E[U_p] = \frac{h}{\gamma(h-M-2)}\left[\theta^2 - \frac{(h-2)(M+h\theta^2)}{2(h-M-1)(h-M-4)}\right] \quad \text{for } h > M+4, \tag{40}$$

$$E[U_s] = \frac{h}{\gamma(h-K-2)}\left[\theta_1^2 - \frac{(h-2)(K+h\theta_1^2)}{2(h-K-1)(h-K-4)}\right] \quad \text{for } h > K+4. \tag{41}$$

Using the results from Proposition 3, we can obtain the difference in expected out-of-sample performance of the sample optimal portfolio and the benchmark sample optimal portfolio

$$\Delta_p = E[U_p] - E[U_s]. \tag{42}$$

Figure 7 plots Δ_p as a function of the length of the estimation window (h). We assume $K = 3$, $\theta_1 = 0.1$, and $\gamma = 10$. Different combinations of θ (0.2 and 0.3) and N (10 or 25) are examined. Figure 7 shows that unlike the true optimal portfolio p^*, the sample optimal portfolio p does not always outerform the benchmark sample optimal portfolio s. When the estimation window is relatively short, portfolio p underperforms portfolio s. For example, when $\theta = 0.3$ and $N = 10$, the estimation window needs to be longer than 180 months in order for portfolio p to outperform portfolio s. The required estimation window becomes even longer when θ goes down and N goes up. For $\theta = 0.2$ and $N = 25$, portfolio p underperforms portfolio s even for estimation window as long as 600 months. The underperformance of portfolio p is due to the fact that portfolio p involves more risky assets than portfolio s and hence more estimation errors. When the length of the estimation window is short and when N is large, the effect of estimation errors on the test assets outweighs the effect of improved Sharpe ratio.

Figure 7 about here

III. The Optimal Combining Portfolio

The previous section shows that although the sample optimal portfolio has positive unconditional alpha, the investor is not necessarily better off by holding the sample optimal portfolio p instead of the benchmark sample optimal portfolio s as shown in Figure 7. In this section, we examine whether we can improve portfolio performance by combining the sample optimal portfolio p, the benchmark sample optimal portfolio s, and the risk-free asset.

A. The Optimal Combining Rule

We consider a combining portfolio which is a linear combination of p and s, so the excess return of the combining portfolio c at time $t+1$ is

$$r_{c,t+1} = \lambda_1 r_{p,t+1} + \lambda_2 r_{s,t+1} = \lambda_1 w'_{p,t} r_{t+1} + \lambda_2 w'_{s,t} r_{1,t+1}, \tag{43}$$

where λ_1 and λ_2 are two scalar combining coefficients. The expected out-of-sample performance of the combining portfolio c is

$$E[U_c(\lambda_1, \lambda_2)] = \lambda_1 E[w'_{p,t}\mu] + \lambda_2 E[w'_{s,t}\mu_1]$$

$$-\frac{\gamma}{2}\left(\lambda_1^2 E[w_{p,t}' V w_{p,t}] + \lambda_2^2 E[w_{s,t}' V_{11} w_{s,t}] + 2\lambda_1\lambda_2 E\left[w_{p,t}' \begin{bmatrix} V_{11} \\ V_{21} \end{bmatrix} w_{s,t}\right]\right). \quad (44)$$

In the following Proposition, we obtain the optimal values of λ_1 and λ_2 as well as the expected out-of-sample performance of the optimal combining portfolio c^*.[8]

Proposition 4: *The λ_1 and λ_2 that maximize $E[U_c(\lambda_1, \lambda_2)]$ are given by:*

$$\lambda_1^* = \frac{(h-M-2)\delta^2}{B-C}, \quad (45)$$

$$\lambda_2^* = (h-K-2)\left(\frac{\theta_1^2}{C} - \frac{\delta^2}{B-C}\right), \quad (46)$$

where

$$B = \frac{(h-2)(h-M-2)(M+h\theta^2)}{(h-M-1)(h-M-4)}, \quad (47)$$

$$C = \frac{(h-2)(h-K-2)(K+h\theta_1^2)}{(h-K-1)(h-K-4)}. \quad (48)$$

The expected out-of-sample performance of the optimal combining portfolio c^* is

$$E[U_{c^*}] = E[U_c(\lambda_1^*, \lambda_2^*)] = \frac{h\theta_1^4}{2\gamma C} + \frac{h\delta^4}{2\gamma(B-C)}. \quad (49)$$

When $h > M+4$, we have $B > 0$ and $C > 0$. In addition, we have $B > C$ because

$$\begin{aligned}
& 0 < \frac{h-M-4}{h-K-4} < \frac{h-M-2}{h-K-2} < \frac{h-M-1}{h-K-1} < 1 \\
\Rightarrow\ & 0 < \frac{(h-M-1)}{(h-K-1)}\frac{(h-M-4)(h-K-2)}{(h-K-4)(h-M-2)} < 1 \\
\Rightarrow\ & 0 < \frac{C}{B} = \frac{(K+h\theta_1^2)(h-M-1)}{(M+h\theta^2)(h-K-1)}\frac{(h-M-4)(h-K-2)}{(h-K-4)(h-M-2)} < 1 \\
\Rightarrow\ & B > C. \qquad (50)
\end{aligned}$$

Unlike $E[U_p]$ which can take negative values, Proposition 4 suggests that the optimal combining portfolio c^* always has positive expected out-of-sample performance when $h > M+4$.

[8] Tu and Zhou (2011) also consider a combining portfolio rule of p and a fixed weight portfolio. However, their combining portfolio rule imposes a constraint of $\lambda_1 + \lambda_2 = 1$, and this constraint can lead to substantial utility loss. In addition, unlike other portfolio rules, the out-of-sample utility of their combining portfolio rule is not linear in $1/\gamma$, so whether their combining portfolio rule outperforms other portfolio rules or not depends on the choice of γ.

Moreover, when the benchmark portfolios are not *ex ante* efficient (i.e., $\delta > 0$), Proposition 4 indicates that the optimal combining portfolio c^* has positive weight on the sample optimal portfolio (i.e., $\lambda_1^* > 0$), suggesting that there are benefits of incorporating test assets with nonzero alphas.

Figure 8 about here

In Figure 8, we plot λ_1^* as a function of the length of estimation window (h) for an investor with $\gamma = 10$. The number of benchmark portfolios is assumed to be three ($K = 3$), and the Sharpe ratio of the benchmark optimal portfolio is assumed to be 0.1 ($\theta_1 = 0.1$). The figure plots the results for different combinations of θ (0.2 or 0.3) and N (10 or 25). Figure 8 shows that although the optimal weight on the sample optimal portfolio is positive, the weight is far less than 100%, suggesting that when there are estimation errors, holding only the sample optimal portfolio is not the best strategy. The optimal weight λ_1^* increases as h goes up, N goes down, and θ goes up. With a higher value of θ, incorporating the test assets can improve portfolio performance more, and as a result, the optimal combining portfolio has more weight on the sample optimal portfolio. As h goes up or N goes down, there are fewer estimation errors in $w_{p,t}$. Therefore, the benefit of improved Sharpe ratio of the sample optimal portfolio outweighs the effect of estimation errors, and the optimal combining portfolio has more weight on the sample optimal portfolio.

Figure 9 about here

In Figure 9, we plot λ_2^* as a function of the length of estimation window (h) for different combinations of θ (0.2 or 0.3) and N (10 or 25), using the same assumptions as those for Figure 8. Figure 9 shows that the optimal combining strategy has nonzero weight on the benchmark sample optimal portfolio. This is because due to estimation errors, holding just the sample optimal portfolio is not optimal, and including another portfolio with random weights (i.e., the benchmark sample optimal portfolio) helps to diversify the estimation errors. Unlike λ_1^*, λ_2^* decreases as N goes down and θ goes up. This is because smaller N and higher θ indicate more benefit of incorporating the sample optimal portfolio into the combining strategy. As a result, less weight is put on the benchmark sample optimal portfolio. In addition, Figure 9 shows that λ_2^* is not a monotonic function of h. This can be explained by two facts. First, when h is small, there are significant estimation errors involved in both portfolios p and s. Since we do not restrict $\lambda_1^* + \lambda_2^* = 1$, the optimal strategy is to shift more weight to the risk-free asset. Therefore, the magnitudes of both λ_1^* and λ_2^* are small when h is small. As h goes up, there are increased benefits for investing in both p and s, so the optimal combining portfolio will see an increase in the magnitudes of λ_1^* and λ_2^*. Second, as h goes to infinity, we can estimate the parameters accurately and the benefit of incorporating the sample optimal portfolio becomes dominant. As a result, for very large h, the optimal combining portfolio will have close to 100% on the sample optimal portfolio and almost zero weight on the benchmark sample optimal portfolio. These two effects lead to a non-monotonic behavior of λ_2^* as h increases.

Comparing the expected out-of-sample performance of the optimal combining portfolio in (49) with that of the benchmark sample optimal portfolio in (41), we obtain the performance improvement of portfolio c^* relative to portfolio s

$$\Delta_{c^*} = E[U_{c^*}] - E[U_s] = \frac{h}{2\gamma C}\left(\theta_1^2 - \frac{C}{h-K-2}\right)^2 + \frac{h\delta^4}{2\gamma(B-C)}. \tag{51}$$

Note that both terms in the above equation are positive. The first term reflects the utility gain due to the possibility to optimize between portfolio s and the risk-free asset,[9] and the second term reflects the utility gain due to the improved Sharpe ratio by incorporating the sample optimal portfolio. Intuitively, the effect of the first term dominates when h is small, and the effect of the second term dominates when h is large.

Figure 10 about here

In Figure 10, we plot Δ_{c^*} as a function of h. The figure shows that when h is small (e.g., $h = 60$), the combining portfolio greatly outperforms portfolio s. As discussed previously, this is due to the effect of the first term in (51). As h goes up, this effect attenuates and

[9]It can be shown that the expected out-of-sample performance of the optimal portfolio combining the benchmark sample optimal portfolio s and the risk-free asset is $h\theta_1^4/(2\gamma C)$. Relative to portfolio s, the performance improvement is given by the first term in (51).

the effect coming from the improved Sharpe ratio (i.e., the second term in (51)) increases, which converges to $\delta^2/(2\gamma)$ as $h \to \infty$.

B. Implementable Combining Rules

Since θ_1^2 and δ^2 are unknown in practice, the optimal combining portfolio c^* is not feasible. We need to estimate θ_1^2 and δ^2 in order to implement the combining rule. We consider two different sets of estimators. The first set of estimator of θ_1^2 and δ^2 is their sample counterparts

$$\hat{\theta}_1^2 = \hat{\mu}_{1,t}' \hat{V}_{11,t}^{-1} \hat{\mu}_{1,t}, \tag{52}$$

$$\hat{\delta}^2 = \hat{\mu}_t \hat{V}_t^{-1} \hat{\mu}_t - \hat{\mu}_{1,t}' \hat{V}_{11,t}^{-1} \hat{\mu}_{1,t}. \tag{53}$$

In the proof of Proposition 5, we show that

$$\hat{\theta}_1^2 \sim \frac{K}{h-K} F_{K,h-K}(h\theta_1^2), \tag{54}$$

$$\hat{\delta}^2 | \hat{\theta}_1^2 \sim \frac{(1+\hat{\theta}_1^2)N}{(h-M)} F_{N,h-M}\left(\frac{h\delta^2}{1+\hat{\theta}_1^2}\right), \tag{55}$$

where $F_{n_1,n_2}(\lambda)$ represents the noncentral F distribution with degrees of freedom of n_1 and n_2, and a noncentrality parameter of λ. It is well known that these estimators can be heavily biased when h is small.

In order to come up with improved estimators of θ_1^2 and δ^2, we first note that the problem of estimating θ_1^2 and δ^2 is equivalent to the problem of estimating the noncentrality parameter of a noncentral F-distribution using a single observation. This problem has been studied by a number of researchers in statistics. Our second set of estimators of θ_1^2 and δ^2 is based on an adjusted estimator suggested by Kubokawa, Robert, and Saleh (1993):

$$\hat{\theta}_{1,a}^2 = \frac{(h-K-2)\hat{\theta}_1^2 - K}{h} + \frac{2(\hat{\theta}_1^2)^{\frac{K}{2}}(1+\hat{\theta}_1^2)^{-\frac{h-2}{2}}}{hB_{\hat{\theta}_1^2/(1+\hat{\theta}_1^2)}(K/2,(h-K)/2)}, \tag{56}$$

$$\hat{\delta}_a^2 = \frac{(h-M-2)\hat{\delta}^2 - N(1+\hat{\theta}_1^2)}{h} + \frac{2(1+\hat{\theta}_1^2)\Xi^{\frac{N}{2}}(1+\Xi)^{-\frac{h-K-2}{2}}}{hB_{\Xi/(1+\Xi)}(N/2,(h-M)/2)}, \tag{57}$$

where $\Xi = \hat{\delta}^2/(1+\hat{\theta}_1^2)$ and

$$B_z(a,b) = \int_0^z y^{a-1}(1-y)^{b-1}\mathrm{d}y \tag{58}$$

is the incomplete beta function.[10]

With the estimators of θ_1^2 and δ^2 available, we can obtain the estimators of λ_1 and λ_2. For the first set of estimator of θ_1^2 and δ^2, we have

$$\hat{\lambda}_1 = \frac{(h-M-2)\hat{\delta}^2}{\hat{B}-\hat{C}}, \tag{59}$$

$$\hat{\lambda}_2 = (h-K-2)\left(\frac{\hat{\theta}_1^2}{\hat{C}} - \frac{\hat{\delta}^2}{\hat{B}-\hat{C}}\right), \tag{60}$$

where

$$\hat{B} = \frac{(h-2)(h-M-2)[M+h(\hat{\theta}_1^2+\hat{\delta}^2)]}{(h-M-1)(h-M-4)}, \tag{61}$$

$$\hat{C} = \frac{(h-2)(h-K-2)(K+h\hat{\theta}_1^2)}{(h-K-1)(h-K-4)}. \tag{62}$$

For the second set of estimator of θ_1^2 and δ^2, we have

$$\hat{\lambda}_{1,a} = \frac{(h-M-2)\hat{\delta}_a^2}{\hat{B}_a-\hat{C}_a}, \tag{63}$$

$$\hat{\lambda}_{2,a} = (h-K-2)\left(\frac{\hat{\theta}_{1,a}^2}{\hat{C}_a} - \frac{\hat{\delta}_a^2}{\hat{B}_a-\hat{C}_a}\right), \tag{64}$$

where

$$\hat{B}_a = \frac{(h-2)(h-M-2)[M+h(\hat{\theta}_{1,a}^2+\hat{\delta}_a^2)]}{(h-M-1)(h-M-4)}, \tag{65}$$

$$\hat{C}_a = \frac{(h-2)(h-K-2)(K+h\hat{\theta}_{1,a}^2)}{(h-K-1)(h-K-4)}. \tag{66}$$

We call the combining portfolio that uses $\hat{\lambda}_1$ and $\hat{\lambda}_2$ as portfolio c_s, and the one that uses $\hat{\lambda}_{1,a}$ and $\hat{\lambda}_{2,a}$ as portfolio c_a.

Analytical computation of the expected out-of-sample performance of the combining portfolios c_s and c_a is difficult. In the following Proposition, we present a fast simulation approach to obtain the expected out-of-sample performance of these rules, which requires simulating only a small number of independent univariate random variables. Besides providing an efficient method for computing $E[U_c(\hat{\lambda}_1, \hat{\lambda}_2)]$ and $E[U_c(\hat{\lambda}_{1,a}, \hat{\lambda}_{2,a})]$, Proposition 5 also shows that these two quantities only depend on h, K, N, γ, θ_1, and θ.

[10] This adjusted estimator has also been applied in Kan and Zhou (2007).

Proposition 5: *The expected out-of-sample performance of the implementable combining rules can be simulated using the following steps:*

1. *Let* $\tilde{y}_0 \sim N(0,1)$, $z_{11} \sim N(\sqrt{h}\theta_1, 1)$, $u_1 \sim \chi^2_{h-K}$, $u_2 \sim \chi^2_{h-M}$, $v_1 \sim \chi^2_{K-1}$, $v_2 \sim \chi^2_{N-1}$, *and all these random variables are independent of each other.*[11]

2. *Set* $\hat{\theta}_1^2 = \tilde{u}_1/u_1$, *where* $\tilde{u}_1 = z_{11}^2 + v_1$.

3. *Set* $\hat{\delta}^2 = (1+\hat{\theta}_1^2)\tilde{u}_2/u_2$, *where* $\tilde{u}_2 = \tilde{y}_1^2 + v_2$ *and* $\tilde{y}_1 = \tilde{y}_0 - \sqrt{h}\delta/\sqrt{1+\hat{\theta}_1^2}$.

4. *With* $\hat{\theta}_1^2$ *and* $\hat{\delta}^2$ *available, we can obtain* $\hat{\lambda}_1$ *and* $\hat{\lambda}_2$ *using (59)–(60).*

5. *Obtain the expressions for the following five terms using simulation:*

$$E_1 = \sqrt{h}\theta_1 E[\hat{\lambda}_1 z_{11} q] - \sqrt{h}\delta E\left[\frac{\hat{\lambda}_1\sqrt{1+\hat{\theta}_1^2}\tilde{y}_1}{u_2}\right], \tag{67}$$

$$E_2 = \sqrt{h}\theta_1 E\left[\frac{\hat{\lambda}_2 z_{11}}{u_1}\right], \tag{68}$$

$$E_3 = \frac{h(h-2)}{h-K-1}\left(E[\hat{\lambda}_1^2\tilde{u}_1 q^2] + \frac{h-K-2}{h-M-1}E\left[\frac{\hat{\lambda}_1^2\hat{\delta}^2}{u_2}\left(1+\frac{1}{u_1(1+\hat{\theta}_1^2)}\right)\right]\right.$$
$$\left. + \frac{h\delta^2}{h-M-1}E\left[\frac{\hat{\lambda}_1^2\hat{\theta}_1^2 v_2}{(1+\hat{\theta}_1^2)u_1 u_2^2}\right]\right), \tag{69}$$

$$E_4 = \frac{h(h-2)}{h-K-1}E\left[\frac{\hat{\lambda}_2^2\hat{\theta}_1^2}{u_1}\right], \tag{70}$$

$$E_5 = \frac{h(h-2)}{h-K-1}E[\hat{\lambda}_1\hat{\lambda}_2\hat{\theta}_1^2 q]. \tag{71}$$

where

$$q = \frac{1}{u_1}\left(1+\frac{v_2+\tilde{y}_0\tilde{y}_1}{u_2}\right). \tag{72}$$

6. *The expected out-of-sample performance of portfolio* c_s *can be written as*

$$E[U_c(\hat{\lambda}_1,\hat{\lambda}_2)] = \frac{1}{\gamma}(E_1+E_2) - \frac{1}{2\gamma}(E_3+E_4+2E_5). \tag{73}$$

[11] We use the convention of $\chi_0^2 \equiv 0$ to deal with the case that $K=1$ or $N=1$.

Note that $\hat{\theta}^2_{1,a}$ and $\hat{\delta}^2_a$ are functions of $\hat{\theta}^2_1$ and $\hat{\delta}^2$. With the expressions of $\hat{\theta}^2_1$ and $\hat{\delta}^2$ available, we can easily obtain the expressions of $\hat{\theta}^2_{1,a}$ and $\hat{\delta}^2_a$, and therefore, the expressions of $\hat{\lambda}_{1,a}$ and $\hat{\lambda}_{2,a}$. Replacing $\hat{\theta}^2_1$ with $\hat{\theta}^2_{1,a}$, $\hat{\delta}^2$ with $\hat{\delta}^2_a$, $\hat{\lambda}_1$ with $\hat{\lambda}_{1,a}$, and $\hat{\lambda}_2$ with $\hat{\lambda}_{2,a}$, we get the expression of the out-of-sample performance of portfolio c_a, $E[U_c(\hat{\lambda}_{1,a}, \hat{\lambda}_{2,a})]$.

Figure 11 about here

Figure 12 about here

In Figures 11 and 12, we examine the performance improvement of the implementable combining portfolios relative to the benchmark sample optimal portfolio as a function of the length of estimation window for an investor with $\gamma = 10$. The number of benchmark portfolios is assumed to be three ($K = 3$), and the Sharpe ratio of the true benchmark optimal portfolio is assumed to be 0.1 ($\theta_1 = 0.1$). Figure 11 is based on the implementable combining portfolio with sample estimates of θ^2_1 and δ^2 (i.e., portfolio c_s), and Figure 12 is based on the implementable combining portfolio with adjusted estimates of θ^2_1 and δ^2 (i.e., portfolio c_a).

Figure 11 shows that when $N = 10$, portfolio c_s outperforms the benchmark sample optimal portfolio regardless of the length of the estimation window. However, when $N = 25$, there are more estimation errors in the combining coefficients, and portfolio c_s does not outperform the benchmark sample optimal portfolio when h is small. With the adjusted estimates, Figures 12 shows that portfolio c_a consistently outperforms the benchmark sample optimal portfolio regardless of the values of h, N, and θ. Therefore, the implementable combining portfolio with adjusted estimates provides the investor a portfolio strategy that reliably outperforms the benchmark sample optimal portfolio.

Table III about here

Table IV about here

Finally, we empirically examine the performance improvement over the benchmark sample optimal portfolio for different portfolio rules using return data over the period of 1931/7 to 2013/12. We consider the same four sets of benchmark portfolios and the five sets of test assets as in Tables I and II. Table III reports the results with an estimation window of 120 months, and Table IV is based on an estimation window of 60 months. The results in the two tables are consistent with our theoretical findings. Both tables show that the sample optimal portfolio p grossly underperforms the benchmark sample optimal portfolio s, but the two combining portfolio rules (c_s and c_a) often offer improvement over portfolio s. In all the cases, the combining portfolio with adjusted estimates (i.e., portfolio c_a) performs better than the portfolio with sample estimates (i.e., portfolio c_s), confirming the findings from Figures 11 and 12. When the estimation window is 60 months, the performance improvement of portfolio c_s is negative in some cases, and that of portfolio c_a is mostly positive. With a longer estimation window ($h = 120$), the performance improvement increases in general. Nevertheless, we notice that for the Fama-French three-factor model and the Carhart four-factor model, the performance improvement can be lower for $h = 120$ than that for $h = 60$. This pattern is also observed in Figure 10, which can be explained by the relative importance of the two effects in (51).

IV. Robustness Check: Departure from the Normality

So far, our analysis is based on the assumption that the excess returns of the risky assets are i.i.d. multivariate normal. It is natural to question the robustness of our results when the excess returns are not multivariate normally distributed. We address this concern by studying two alternative distributional assumptions on the excess returns. The first alternative is a multivariate t-distribution with five degrees of freedom. Under the multivariate t-distribution assumption, the excess returns of the risky assets have fat tails, which is often what we find in actual data. The second alternative is an empirical distribution based on the actual excess return data. We take the excess monthly returns of 10 momentum portfolios and the benchmark portfolios over the period of 1931/7 to 2013/12 as the population distribution and assume the excess returns of the risky assets are independently drawn from this empirical distribution. In order to have a proper comparison across the three distributional assumptions, we set the mean and covariance matrix of the excess returns to be identical in the three cases.

Figure 13 about here

In Figure 13, we present the distributions of the excess returns of the sample optimal portfolio under the three distributional assumptions, assuming the benchmark portfolios are the Fama-French three factors and the investor has $\gamma = 10$ and use $h = 60$ to construct the sample optimal portfolio. The distribution of $r_{p,t+1}$ under the two alternative distributional assumptions are generated based on 1,000,000 simulations. We notice that the distributions of $r_{p,t+1}$ under the two alternative distributional assumptions are quite similar. They both show higher kurtosis than the one from the normality assumption. Nevertheless, the distribution of $r_{p,t+1}$ under the three distributional assumptions are not too far off from each other. Therefore, we consider the distribution of $r_{p,t+1}$ derived under the normality assumption as a good working approximation for the case of monthly data.

Figure 14 about here

In Figure 14, we examine the robustness of the distribution of the t-ratio of alpha to departure from the multivariate normality assumption. Similarly, we assume $\gamma = 10$ and $h = 60$, and obtain the distribution of t-ratio under the two alternative distributional assumptions using 1,000,000 simulations. Figure 14 shows that the distribution of t-ratio is not particularly sensitive to the distributional assumption on the excess returns of the risky assets and the benchmark portfolios.

Figure 15 about here

In Figure 15, we plot the performance improvement of the implementable combining portfolio with adjusted estimates relative to the benchmark sample optimal portfolio (Δ_{c_a}) as a function of the length of estimation window (h) under the two alternative distributional assumptions for an investor with $\gamma = 10$ based on 1,000,000 simulations. The figure shows that the performance improvement under the multivariate t-distribution is very similar to that under the multivariate normal distribution, whereas the performance improvement under the empirical distribution is slightly lower compared to that under the other two distributional assumptions. Nevertheless, the performance improvement is positive under all the three distributional assumptions. Therefore, our conclusion that the implementable combining portfolio with adjusted estimates consistently outperforms the benchmark sample optimal portfolio is robust to alternative distributional assumptions.

V. Conclusion

In this paper, we study the benefit of incorporating test assets with nonzero alphas into an optimal portfolio where the parameters are estimated with errors. Our theoretical results suggest that as long as the benchmark portfolios are not *ex ante* efficient, the sample optimal portfolio can generate a positive alpha. In addition, we demonstrate that the t-ratio of the estimated alpha of such a portfolio strategy is often statistically significant. However, due to estimation errors, the sample optimal portfolio does not outperform the benchmarks unless we have a very long estimation window. To explore the economic value of nonzero alphas, we investigate an optimal combining rule, and show that this portfolio strategy consistently outperforms the benchmarks.

An important assumption in our theoretical analysis is that the excess returns on the risky assets are independent and identically distributed. Nevertheless, we believe qualitatively similar results can be obtained even when the expected return and covariance matrix of risky assets are time-varying, provided that they evolve over time in a slowly moving fashion. Although analytical results are difficult to obtain beyond the i.i.d. multivariate normality assumption, more simulation analyses could provide us additional insights on the impact of time-varying return distribution.

The results of our paper also provide an explanation for the persistence of alphas of test assets. When investors do not know the true alphas of certain assets, they will optimally scale back their investments on such assets, especially when the estimation window is short.[12] Without enough money moving into these assets, it will take a long time for their prices to adjust, and this can lead to persistent alphas for these assets.

[12] In our context, this is represented by a small λ_1^* in the sample optimal portfolio p.

From an asset pricing perspective, the alpha of a sample optimal portfolio offers us a potential way to test the mean-variance efficiency of a given set of benchmark portfolios. This is because the unconditional alpha of a sample optimal portfolio is zero if and only if the benchmark portfolios are *ex ante* efficient. This provides an alternative to the popular Gibbons-Ross-Shanken (1989) test of mean-variance efficiency of a given set of benchmark portfolios. Unlike the Gibbons-Ross-Shanken test which is performed using in-sample alphas of the test assets, our test is attractive in the sense that it makes use of out-of-sample alpha from a trading strategy, which is potentially more relevant for investors.

Appendix

Proof of Proposition 1: Under the multivariate normality assumption, it is well known that $\hat{\mu}_t$ and $\hat{V}_t$ are independent of each other and have the following distributions:

$$\hat{\mu}_t \sim N(\mu, V/h), \tag{A1}$$

$$\hat{V}_t \sim W_M(h-1, V/h), \tag{A2}$$

where $W_M(h-1, V/h)$ is a Wishart distribution with $h-1$ degree of freedom and covariance matrix V/h. Define $\eta = V^{-\frac{1}{2}}\mu/\theta$, we have $\eta'\eta = 1$. Let P be an $M \times M$ orthonormal matrix with its first column equals to η. By defining

$$z = \sqrt{h}P'V^{-\frac{1}{2}}\hat{\mu}_t \sim N\left(\left[\begin{array}{c} \sqrt{h}\theta \\ 0_{M-1} \end{array}\right], I_M\right), \tag{A3}$$

$$W = hP'V^{-\frac{1}{2}}\hat{V}_t V^{-\frac{1}{2}}P \sim W_M(h-1, I_M), \tag{A4}$$

we can write the weights of the sample optimal portfolio p as

$$w_{p,t} = \frac{1}{\gamma}\hat{V}_t^{-1}\hat{\mu}_t = \frac{\sqrt{h}}{\gamma}V^{-\frac{1}{2}}PW^{-1}z. \tag{A5}$$

It follows that

$$\mu_{p,t} = \mu' w_{p,t} = \frac{\sqrt{h}\theta}{\gamma}e_1' W^{-1}z, \tag{A6}$$

$$\sigma^2_{p,t} = w'_{p,t}Vw_{p,t} = \frac{h}{\gamma^2}z'W^{-2}z, \tag{A7}$$

where $e_1 = [1,\ 0'_{M-1}]'$. Define an $M \times M$ orthonormal matrix $Q = [\tilde{z},\ Q_1]$ with its first column equals to $\tilde{z} = z/(z'z)^{\frac{1}{2}}$. Let

$$A = (Q'W^{-1}Q)^{-1} = \left[\begin{array}{cc} \tilde{z}'W^{-1}\tilde{z} & \tilde{z}'W^{-1}Q_1 \\ Q_1 W^{-1}\tilde{z} & Q_1 W^{-1}Q_1 \end{array}\right]^{-1} = \left[\begin{array}{cc} A_{11} & A_{12} \\ A_{21} & A_{22} \end{array}\right] \sim W_M(h-1, I_M), \tag{A8}$$

where A_{11} is the $(1,1)$ element of A. Theorem 3.2.10 of Muirhead (1982) suggests that

$$u_1 \equiv A_{11\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21} \sim \chi^2_{h-M}, \tag{A9}$$

and it is independent of A_{12} and A_{22}. In addition, using the result of Dickey (1967), we can show that

$$-A_{22}^{-1}A_{21} \sim \frac{x}{\sqrt{u_2}}, \tag{A10}$$

where $x \sim N(0_{M-1}, I_{M-1})$, $u_2 \sim \chi^2_{h-M+1}$, and they are independent of each other and u_1. Using the partitioned matrix inverse formula, we can easily verify that

$$\tilde{z}'W^{-1}\tilde{z} = A_{11\cdot 2}^{-1} = \frac{1}{u_1}, \tag{A11}$$

$$Q_1'W^{-1}\tilde{z} = -A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1} = \frac{x}{u_1\sqrt{u_2}}. \tag{A12}$$

With these identities, we can write

$$z'W^{-2}z = z'W^{-1}\left(\tilde{z}\tilde{z}' + Q_1Q_1'\right)W^{-1}z = (z'z)\left(\frac{1}{u_1^2} + \frac{x'x}{u_1^2 u_2}\right). \tag{A13}$$

Let $z_1 \sim N(\sqrt{h}\theta, 1)$ and $x_1 \sim N(0,1)$ be the first element of z and x, respectively. We can write $z'z = z_1^2 + u_0$ and $x'x = x_1^2 + u_3$ where $u_0 \sim \chi^2_{M-1}$ and $u_3 \sim \chi^2_{M-2}$. It follows that

$$\sigma_{p,t}^2 = \frac{h(z_1^2 + u_0)}{\gamma^2 u_1^2}\left(1 + \frac{x_1^2 + u_3}{u_2}\right). \tag{A14}$$

Without loss of generality, let the first column of Q_1 be

$$\xi = \frac{(I_M - \tilde{z}\tilde{z}')e_1}{[e_1'(I_M - \tilde{z}\tilde{z}')e_1]^{\frac{1}{2}}} = \frac{(I_M - \tilde{z}\tilde{z}')e_1}{\sqrt{1 - \frac{z_1^2}{z'z}}}. \tag{A15}$$

From (A12), we know that

$$\frac{x_1}{u_1\sqrt{u_2}} = \xi'W^{-1}\tilde{z} = \frac{e_1'W^{-1}\tilde{z} - \frac{e_1'\tilde{z}}{u_1}}{\sqrt{1 - \frac{z_1^2}{z'z}}} = \frac{e_1'W^{-1}z - \frac{z_1}{u_1}}{\sqrt{u_0}}, \tag{A16}$$

and hence

$$e_1'W^{-1}z = \frac{z_1}{u_1} + \frac{x_1\sqrt{u_0}}{u_1\sqrt{u_2}}. \tag{A17}$$

It follows that

$$\mu_{p,t} = \frac{\sqrt{h}\theta}{\gamma u_1}\left(z_1 + \frac{x_1\sqrt{u_0}}{\sqrt{u_2}}\right). \tag{A18}$$

When $M = 1$, u_0 and x vanish and we have

$$\mu_{p,t} = \frac{\sqrt{h}\theta z_1}{\gamma u_1}, \qquad \sigma_{p,t}^2 = \frac{h z_1^2}{\gamma^2 u_1^2}. \tag{A19}$$

This completes the proof.

Proof of Proposition 2: We first deal with the case that $\delta > 0$. Let P be an $M \times M$ orthonormal matrix with its first column equals to

$$\eta = \frac{1}{\delta} V^{-\frac{1}{2}} \begin{bmatrix} 0_K \\ \alpha \end{bmatrix}. \tag{A20}$$

By defining

$$z = \sqrt{h} P' V^{-\frac{1}{2}} \hat{\mu}_t \sim N(\sqrt{h} P' V^{-\frac{1}{2}} \mu, I_M), \tag{A21}$$

$$W = h P' V^{-\frac{1}{2}} \hat{V}_t V^{-\frac{1}{2}} P \sim W_M(h-1, I_M), \tag{A22}$$

we can write

$$\alpha_{p,t} = \frac{1}{\gamma} \hat{\mu}_t' \hat{V}_t^{-1} \begin{bmatrix} 0_K \\ \alpha \end{bmatrix} = \frac{\sqrt{h}\delta}{\gamma} z' W^{-1} e_1. \tag{A23}$$

Following the proof of Proposition 1, in particular (A17), we can write

$$z' W^{-1} e_1 = \frac{z_1}{u_1} + \frac{x_1 \sqrt{u_0}}{u_1 \sqrt{u_2}}, \tag{A24}$$

where $z_1 \sim N(\sqrt{h}\delta, 1)$, $x_1 \sim N(0,1)$, $u_0 \sim \chi^2_{M-1}(h\theta_1^2)$, $u_1 \sim \chi^2_{h-M}$, $u_2 \sim \chi^2_{h-M+1}$, and they are independent of each other. The mean of z_1 and the noncentrality parameter of u_0 are obtained as follows:

$$E[z_1] = \sqrt{h} \eta' V^{-\frac{1}{2}} \mu = \frac{\sqrt{h}}{\delta} \alpha' (-\Sigma^{-1} V_{21} V_{11}^{-1} \mu_1 + \Sigma^{-1} \mu_2) = \frac{\sqrt{h} \alpha' \Sigma^{-1} \alpha}{\delta} = \sqrt{h}\delta, \tag{A25}$$

$$E[z]' E[z] - E[z_1]^2 = h\theta^2 - h\delta^2 = h\theta_1^2. \tag{A26}$$

Conditional on u_0, u_1, and u_2, we have

$$z' W^{-1} e_1 \sim N\left(\frac{\sqrt{h}\delta}{u_1}, \frac{1}{u_1^2} \left(1 + \frac{u_0}{u_2}\right) \right), \tag{A27}$$

and hence we can write

$$\alpha_{p,t} = \frac{\sqrt{h}\delta}{\gamma u_1} \left[\sqrt{h}\delta + \left(1 + \frac{u_0}{u_2}\right)^{\frac{1}{2}} y \right], \tag{A28}$$

where $y \sim N(0,1)$ and it is independent of u_0, u_1, and u_2. Note that when $\delta = 0$, we have $\alpha_{p,t} = 0$ from (34), so (A28) also holds for the $\delta = 0$ case. This completes the proof.

Proof of Proposition 3: $E[U_p]$ and $E[U_s]$ can be obtained by using Eq.(20) of Kan and Zhou (2007). Here we provide an alternative proof. Based on the results of Proposition 1, the expected out-of-sample utility of portfolio p can be obtained as

$$E[U_p] = E\left[\mu_{p,t} - \frac{\gamma}{2}\sigma_{p,t}^2\right] = \frac{\sqrt{h}\theta}{\gamma}E\left[\frac{z_1}{u_1} + \frac{x_1\sqrt{u_0}}{u_1\sqrt{u_2}}\right] - \frac{h}{2\gamma}E\left[\frac{z_1^2+u_0}{u_1^2}\left(1+\frac{x_1^2+u_3}{u_2}\right)\right]. \quad \text{(A29)}$$

Note that x_1 has zero mean and is independent of u_0, u_1, and u_2. In addition, $z'z = z_1^2+u_0 \sim \chi_M^2(h\theta^2)$, $x'x = x_1^2 + u_3 \sim \chi_{M-1}^2$, and $z'z$, $x'x$, u_1, and u_2 are independent of each other. Therefore,

$$\begin{aligned} E[U_p] &= \frac{\sqrt{h}\theta}{\gamma}E\left[\frac{z_1}{u_1}\right] - \frac{h}{2\gamma}E[z'z]E\left[\frac{1}{u_1^2}\right]\left(1+E[x'x]E\left[\frac{1}{u_2}\right]\right) \\ &= \frac{h\theta^2}{\gamma(h-M-2)} - \frac{h}{2\gamma}\frac{(M+h\theta^2)}{(h-M-2)(h-M-4)}\frac{(h-2)}{(h-M-1)} \\ &= \frac{h}{\gamma(h-M-2)}\left[\theta^2 - \frac{(h-2)(M+h\theta^2)}{2(h-M-1)(h-M-4)}\right]. \end{aligned} \quad \text{(A30)}$$

The derivation of $E[U_s]$ is obtained by changing M to K and θ to θ_1 in the proof above. This completes the proof.

Proof of Proposition 4: We can write

$$E[U_c(\lambda_1,\lambda_2)] = \lambda_1 E[\mu_{p,t}] + \lambda_2 E[\mu_{s,t}] - \frac{\gamma}{2}\left(\lambda_1^2 E[\sigma_{p,t}^2] + \lambda_2^2 E[\sigma_{s,t}^2] + 2\lambda_1\lambda_2 E[\sigma_{ps,t}]\right), \quad \text{(A31)}$$

where $\mu_{p,t}$ and $\mu_{s,t}$ are the conditional mean of portfolio p and portfolio s at time t, $\sigma_{p,t}^2$ and $\sigma_{s,t}^2$ are the conditional variance of the two portfolios, and $\sigma_{ps,t}$ is the conditional covariance between the two portfolio. Differentiating $E[U_c(\lambda_1,\lambda_2)]$ with respect to λ_1 and λ_2 and setting them equal to zero, we have

$$\frac{\partial E[U_c(\lambda_1,\lambda_2)]}{\partial \lambda_1} = E[\mu_{p,t}] - \gamma\lambda_1 E[\sigma_{p,t}^2] - \gamma\lambda_2 E[\sigma_{ps,t}] = 0, \quad \text{(A32)}$$

$$\frac{\partial E[U_c(\lambda_1,\lambda_2)]}{\partial \lambda_2} = E[\mu_{s,t}] - \gamma\lambda_2 E[\sigma_{s,t}^2] - \gamma\lambda_1 E[\sigma_{ps,t}] = 0. \quad \text{(A33)}$$

Solving for λ_1 and λ_2, we obtain

$$\lambda_1^* = \frac{1}{\gamma}\left(\frac{E[\mu_{p,t}]E[\sigma_{s,t}^2] - E[\mu_{s,t}]E[\sigma_{ps,t}]}{E[\sigma_{s,t}^2]E[\sigma_{p,t}^2] - E[\sigma_{ps,t}]^2}\right), \quad \text{(A34)}$$

$$\lambda_2^* = \frac{1}{\gamma}\left(\frac{E[\mu_{s,t}]E[\sigma_{p,t}^2] - E[\mu_{p,t}]E[\sigma_{ps,t}]}{E[\sigma_{s,t}^2]E[\sigma_{p,t}^2] - E[\sigma_{ps,t}]^2}\right). \tag{A35}$$

From the proof of Proposition 3, we have

$$E[\mu_{p,t}] = \frac{h\theta^2}{\gamma(h-M-2)}, \tag{A36}$$

$$E[\sigma_{p,t}^2] = \frac{h(h-2)(M+h\theta^2)}{\gamma^2(h-M-1)(h-M-2)(h-M-4)}, \tag{A37}$$

$$E[\mu_{s,t}] = \frac{h\theta_1^2}{\gamma(h-K-2)}, \tag{A38}$$

$$E[\sigma_{s,t}^2] = \frac{h(h-2)(K+h\theta_1^2)}{\gamma^2(h-K-1)(h-K-2)(h-K-4)}. \tag{A39}$$

It remains to obtain an explicit expression of $E[\sigma_{ps,t}]$. Let $P = [P_1,\ P_2]$ be an $M \times M$ orthonormal matrix with its first K columns equal to

$$P_1 = V^{\frac{1}{2}}\begin{bmatrix} I_K \\ 0_{N\times K}\end{bmatrix} V_{11}^{-\frac{1}{2}}. \tag{A40}$$

By defining

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \sqrt{h}P'V^{-\frac{1}{2}}\hat{\mu}_t \sim N(\sqrt{h}P'V^{-\frac{1}{2}}\mu, I_M), \tag{A41}$$

$$W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22}\end{bmatrix} = hP'V^{-\frac{1}{2}}\hat{V}_tV^{-\frac{1}{2}}P \sim W_M(h-1, I_M), \tag{A42}$$

where z_1 is a $K\times 1$ vector and W_{11} is a $K \times K$ submatrix of W, we can write

$$E[\sigma_{ps,t}] = \frac{1}{\gamma^2}E\left[\hat{\mu}_t'\hat{V}_t^{-1}V\begin{bmatrix} I_K \\ 0_{N\times K}\end{bmatrix}\hat{V}_{11,t}^{-1}\hat{\mu}_{1,t}\right] = \frac{h}{\gamma^2}E\left[z'W^{-1}\begin{bmatrix} I_K \\ 0_{N\times K}\end{bmatrix}W_{11}^{-1}z_1\right]. \tag{A43}$$

Applying the partitioned matrix inverse formula, we have

$$E[\sigma_{ps,t}] = \frac{h}{\gamma^2}E\left[z_1'W_{11}^{-2}z_1 + (z_1'W_{11}^{-1}W_{12} - z_2')W_{22\cdot 1}^{-1}W_{21}W_{11}^{-2}z_1\right], \tag{A44}$$

where $W_{22\cdot 1} = W_{22} - W_{21}W_{11}^{-1}W_{12}$. Using Theorem 3.2.10 of Muirhead (1982), we have $W_{22\cdot 1} \sim W_N(h-K-1, I_N)$, $\mathrm{vec}(W_{21}W_{11}^{-\frac{1}{2}}) \sim N(0_{NK}, I_{NK})$, $W_{11} \sim W_K(h-1, I_K)$, and $W_{22\cdot 1}$, W_{11}, and $Z = W_{21}W_{11}^{-\frac{1}{2}}$ are independent of each other. In addition, $E[z_2'W_{22\cdot 1}^{-1}W_{21}W_{11}^{-2}z_1] = E[z_2'W_{22\cdot 1}^{-1}ZW_{11}^{-\frac{3}{2}}z_1] = 0$ because $E[Z] = 0_{N\times K}$ and Z is independent of W_{11}, $W_{22\cdot 1}$ and z.

Using the expression for the moments and inverse moments of a Wishart distribution from Haff (1979), we obtain

$$E[Z'Z] = NI_K, \tag{A45}$$

$$E[W_{22\cdot 1}^{-1}] = \frac{1}{h-M-2} I_N, \tag{A46}$$

$$E[W_{11}^{-2}] = \frac{h-2}{(h-K-1)(h-K-2)(h-K-4)} I_K. \tag{A47}$$

Together with the fact that $z_1'z_1 \sim \chi_K^2(h\theta_1^2)$ and $E[z_1'z_1] = K + h\theta_1^2$, we obtain an explicit expression of $E[\sigma_{ps,t}]$ as

$$\begin{aligned}
E[\sigma_{ps,t}] &= \frac{h}{\gamma^2}\left(E[z_1'W_{11}^{-2}z_1] + \frac{E[z_1'W_{11}^{-1}W_{12}W_{21}W_{11}^{-2}z_1]}{h-M-2}\right) \\
&= \frac{h}{\gamma^2}\left(E[z_1'W_{11}^{-2}z_1] + \frac{E[z_1'W_{11}^{-\frac{1}{2}}Z'ZW_{11}^{-\frac{3}{2}}z_1]}{h-M-2}\right) \\
&= \frac{h}{\gamma^2}\left(E[z_1'W_{11}^{-2}z_1] + \frac{NE[z_1'W_{11}^{-2}z_1]}{h-M-2}\right) \\
&= \frac{h(h-2)(K+h\theta_1^2)}{\gamma^2(h-K-1)(h-K-4)(h-M-2)}.
\end{aligned} \tag{A48}$$

Substituting (A36)–(A39) and (A48) in (A34) and (A35), we obtain the expressions of λ_1^* and λ_2^*. Finally, using the expressions of λ_1^* and λ_2^* and substituting (A36)–(A39) and (A48) in (A31), we obtain the expected out-of-sample performance of the optimal combining portfolio. This completes the proof.

Proof of Proposition 5: Let $P = [P_1,\ P_2]$ be an $M \times M$ orthonormal matrix as defined in the proof of Proposition 4. Further define an $M \times M$ orthonormal matrix

$$Q = \begin{bmatrix} Q_1 & 0_{K\times N} \\ 0_{N\times K} & Q_2 \end{bmatrix}, \tag{A49}$$

where Q_1 is a $K \times K$ orthonormal matrix with its first column equals to $V_{11}^{-\frac{1}{2}}\mu_1/\theta_1$, and Q_2 is an $N \times N$ orthonormal matrix with its first column equals to $P_2'V^{-\frac{1}{2}}\mu/\delta$. Define

$$z \equiv \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \sqrt{h}Q'P'V^{-\frac{1}{2}}\hat{\mu}_t \sim N\left(\begin{bmatrix} \sqrt{h}\theta_1 \\ 0_{K-1} \\ \sqrt{h}\delta \\ 0_{N-1} \end{bmatrix}, I_M\right), \tag{A50}$$

$$W = hQ'P'V^{-\frac{1}{2}}\hat{V}_t V^{-\frac{1}{2}}PQ \equiv \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \sim W_M(h-1, I_M), \tag{A51}$$

where z_1 is the first K elements of z, and W_{11} is the upper left $K \times K$ submatrix of W. Using Theorem 3.2.10 of Muirhead (1982), we have $W_{22\cdot 1} \equiv W_{22} - W_{21}W_{11}^{-1}W_{12} \sim W_N(h - K - 1, I_N)$, $\text{vec}(Z) \equiv \text{vec}(W_{21}W_{11}^{-\frac{1}{2}}) \sim N(0_{NK}, I_{NK})$, $W_{11} \sim W_K(h-1, I_K)$, and they are independent of each other. With the above transformations, we can write

$$\hat{\mu}_t = \frac{1}{\sqrt{h}} V^{\frac{1}{2}}PQz, \tag{A52}$$

$$\hat{V}_t^{-1} = hV^{-\frac{1}{2}}PQW^{-1}Q'P'V^{-\frac{1}{2}}. \tag{A53}$$

Using the definition of P_1, we have

$$[I_K,\ 0_{K\times N}]V^{\frac{1}{2}}PQ = V_{11}^{\frac{1}{2}}P_1'PQ = V_{11}^{\frac{1}{2}}[I_K,\ 0_{K\times N}]Q = [V_{11}^{\frac{1}{2}}Q_1,\ 0_{K\times N}], \tag{A54}$$

and we can express $\hat{\mu}_{1,t}$ and $\hat{V}_{11,t}^{-1}$ as

$$\hat{\mu}_{1,t} = [I_K,\ 0_{K\times N}]\hat{\mu}_t = \frac{1}{\sqrt{h}} V_{11}^{\frac{1}{2}}Q_1 z_1, \tag{A55}$$

$$\hat{V}_{11,t}^{-1} = ([I_K,\ 0_{K\times N}]\hat{V}_t[I_K,\ 0_{K\times N}]')^{-1} = hV_{11}^{-\frac{1}{2}}Q_1W_{11}^{-1}Q_1'V_{11}^{-\frac{1}{2}},. \tag{A56}$$

Let $S_1 = [\eta_1,\ T_1]$ be a $K \times K$ orthonormal matrix with its first column equals to $\eta_1 = z_1/(z_1'z_1)^{\frac{1}{2}}$, and

$$A = (S_1'W_{11}^{-1}S_1)^{-1} \equiv \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \sim W_K(h-1, I_K), \tag{A57}$$

where A_{11} is the (1,1) element of A. Theorem 3.2.10 of Muirhead (1982) suggests that $u_1 \equiv A_{11} - A_{12}A_{22}^{-1}A_{21} \sim \chi^2_{h-K}$, and it is independent of A_{12} and A_{22}. Since $\eta_1'W_{11}^{-1}\eta_1 = 1/u_1$, we can write

$$\hat{\theta}_1^2 = z_1'W_{11}^{-1}z_1 = (z_1'z_1)\eta_1'W_{11}^{-1}\eta_1 = \frac{\tilde{u}_1}{u_1}, \tag{A58}$$

where $\tilde{u}_1 \equiv z_1'z_1 = z_{11}^2 + v_1$ with $z_{11} \sim N(\sqrt{h}\theta_1, 1)$ being the first element of z_1 and $v_1 \sim \chi^2_{K-1}$, and u_1 and $\tilde{u}_1$ are independent of each other. Using the result of Dickey (1967), we have

$$T_1'W_{11}^{-1}\eta_1 = \frac{s_1}{u_1\sqrt{w_1}}, \tag{A59}$$

where $s_1 \sim N(0_{K-1}, I_{K-1})$ and $w_1 \sim \chi^2_{h-K+1}$, and s_1, w_1, u_1, and $\tilde{u}_1$ are independent of each other. Therefore,

$$z_1' W_{11}^{-2} z_1 = z_1' W_{11}^{-1} \left(\eta_1 \eta_1' + T_1 T_1'\right) W_{11}^{-1} z_1 = \frac{\tilde{u}_1}{u_1^2} \left(1 + \frac{s_1' s_1}{w_1}\right), \tag{A60}$$

Let the first column of T_1 be

$$\frac{(I_K - \eta_1 \eta_1')\tilde{e}_1}{[\tilde{e}_1'(I_K - \eta_1\eta_1')\tilde{e}_1]^{\frac{1}{2}}} = \frac{(I_K - \eta_1 \eta_1')\tilde{e}_1}{\sqrt{1 - \frac{z_{11}^2}{z_1' z_1}}} = \frac{(I_K - \eta_1 \eta_1')\tilde{e}_1}{\sqrt{v_1/\tilde{u}_1}}, \tag{A61}$$

where $\tilde{e}_1 = [1,\ 0_{K-1}']'$. Following the proof of Proposition 1, in particular (A17), we obtain

$$\tilde{e}_1' W_{11}^{-1} z_1 = \frac{z_{11}}{u_1} + \frac{s_{11}\sqrt{v_1}}{u_1\sqrt{w_1}}, \tag{A62}$$

where s_{11} is the first element of s_1.

Let $\xi_1 = W_{11}^{-\frac{1}{2}} z_1/\hat{\theta}_1$ and $y \equiv W_{21} W_{11}^{-1} z_1 - z_2 = \hat{\theta}_1 Z \xi_1 - z_2$. Conditional on $W_{11}^{-\frac{1}{2}} z_1$, we have

$$y \sim N\left(-E[z_2], (1 + \hat{\theta}_1^2) I_N\right). \tag{A63}$$

Using the fact that $E[z_2] = \sqrt{h}\delta \hat{e}_1$ where $\hat{e}_1 = [1,\ 0_{N-1}']'$, we define

$$\tilde{y} \equiv \frac{y}{\sqrt{1 + \hat{\theta}_1^2}} \sim N\left(-\frac{\sqrt{h}\delta}{\sqrt{1+\hat{\theta}_1^2}}\hat{e}_1, I_N\right), \tag{A64}$$

and

$$\tilde{u}_2 \equiv \tilde{y}'\tilde{y} = \tilde{y}_1^2 + v_2 \tag{A65}$$

where $\tilde{y}_1$ is the first element of $\tilde{y}$ and $v_2 \sim \chi^2_{N-1}$. In addition, $\tilde{y}_1$ is independent of v_2, and we can write

$$\tilde{y}_1 = \tilde{y}_0 - \sqrt{h}\delta / \sqrt{1 + \hat{\theta}_1^2}, \tag{A66}$$

where $\tilde{y}_0 \sim N(0, 1)$,

Let $S_2 = [\eta_2,\ T_2]$ be an $N \times N$ orthonormal matrix with its first column equals to

$$\eta_2 = \frac{y}{(y'y)^{\frac{1}{2}}}, \tag{A67}$$

and

$$B = (S_2' W_{22\cdot 1}^{-1} S_2)^{-1} \equiv \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \sim W_N(h - K - 1, I_N), \tag{A68}$$

where B_{11} is the (1,1) element of B. Using Theorem 3.2.10 of Muirhead (1982), we have $u_2 \equiv B_{11} - B_{12}B_{22}^{-1}B_{21} \sim \chi^2_{h-M}$, and it is independent of B_{21} and B_{22}. Since $\eta_2' W_{22\cdot 1}^{-1}\eta_2 = 1/u_2$, we can write

$$\hat{\delta}^2 = \hat{\theta}^2 - \hat{\theta}_1^2 = z'W^{-1}z - z_1'W_{11}^{-1}z_1 = y'W_{22\cdot 1}^{-1}y = \frac{(1+\hat{\theta}_1^2)\tilde{u}_2}{u_2}, \tag{A69}$$

where $\tilde{u}_2$ and u_2 are independent of each other. In addition, we have

$$T_2'W_{22\cdot 1}^{-1}\eta_2 = \frac{s_2}{u_2\sqrt{w_2}}, \tag{A70}$$

where $s_2 \sim N(0_{N-1}, I_{N-1})$ and $w_2 \sim \chi^2_{h-M+1}$, and they are independent of each other and u_1, u_2, $\tilde{u}_1$, and $\tilde{u}_2$. We can write

$$y'W_{22\cdot 1}^{-2}y = y'W_{22\cdot 1}^{-1}(\eta_2\eta_2' + T_2T_2')W_{22\cdot 1}^{-1}y = \frac{(1+\hat{\theta}_1^2)\tilde{u}_2}{u_2^2}\left(1 + \frac{s_2's_2}{w_2}\right). \tag{A71}$$

Let the first column of T_2 be

$$\frac{(I_N - \eta_2\eta_2')\hat{e}_1}{[\hat{e}_1'(I_N - \eta_2\eta_2')\hat{e}_1]^{\frac{1}{2}}}. \tag{A72}$$

Following the proof of Proposition 1, we have

$$\hat{e}_1'W_{22\cdot 1}^{-1}y = \frac{\tilde{y}_1\sqrt{1+\hat{\theta}_1^2}}{u_2} + \frac{s_{21}\sqrt{(1+\hat{\theta}_1^2)v_2}}{u_2\sqrt{w_2}}, \tag{A73}$$

where s_{21} is the first element of s_2.

Note that s_1 and w_1 in (A60) and (A62) are independent of $\hat{\theta}_1^2$ and $\hat{\delta}^2$, and therefore, independent of $\hat{\lambda}_2$. Together with the fact that s_{11} has zero mean, we obtain

$$E_2 \equiv E[\hat{\lambda}_2\mu_1'\hat{V}_{11,t}^{-1}\hat{\mu}_{1,t}] = \sqrt{h}\theta_1 E[\hat{\lambda}_2\tilde{e}_1'W_{11}^{-1}z_1] = \sqrt{h}\theta_1 E\left[\frac{\hat{\lambda}_2 z_{11}}{u_1}\right], \tag{A74}$$

$$E_4 \equiv E[\hat{\lambda}_2^2\hat{\mu}_{1,t}\hat{V}_{11,t}^{-1}V_{11}\hat{V}_{11,t}^{-1}\hat{\mu}_{1,t}] = hE[\hat{\lambda}_2^2 z_1'W_{11}^{-2}z_1] = \frac{h(h-2)}{h-K-1}E\left[\frac{\hat{\lambda}_2^2\hat{\theta}_1^2}{u_1}\right]. \tag{A75}$$

In addition, note that s_2 and w_2 in (A71) and (A73) are also independent of $\hat{\theta}_1^2$ and $\hat{\delta}^2$, and s_2 has zero mean. We have

$$E[\hat{\lambda}_1\hat{e}_1'W_{22\cdot 1}^{-1}y] = E\left[\frac{\hat{\lambda}_1\tilde{y}_1\sqrt{1+\hat{\theta}_1^2}}{u_2}\right], \tag{A76}$$

$$E[\hat{\lambda}_1^2 y' W_{22\cdot 1}^{-2} y] = \frac{h-K-2}{h-M-1} E\left[\frac{\hat{\lambda}_1^2(1+\hat{\theta}_1^2)\tilde{u}_2}{u_2^2}\right]. \tag{A77}$$

To complete the proof, we still need to obtain the following terms

$$E_1 \equiv E[\hat{\lambda}_1 \mu' \hat{V}_t^{-1} \hat{\mu}_t], \tag{A78}$$

$$E_3 \equiv E\left[\hat{\lambda}_1^2 \hat{\mu}_t' \hat{V}_t^{-1} V \hat{V}_t^{-1} \hat{\mu}_t\right], \tag{A79}$$

$$E_5 \equiv E\left[\hat{\lambda}_1 \hat{\lambda}_2 \hat{\mu}_t' \hat{V}_t^{-1} \begin{bmatrix} V_{11} \\ V_{21} \end{bmatrix} \hat{V}_{11,t}^{-1} \hat{\mu}_{1,t}\right]. \tag{A80}$$

Using (A52), (A53), (A56), and applying the partitioned matrix inverse formula on W, we get

$$E_1 = \sqrt{h}\theta_1 E[\hat{\lambda}_1(\tilde{e}_1' W_{11}^{-1} z_1 + \tilde{e}_1' W_{11}^{-1} W_{12} W_{22\cdot 1}^{-1} y)] - \sqrt{h}\delta E[\hat{\lambda}_1 \hat{e}_1' W_{22\cdot 1}^{-1} y], \tag{A81}$$

$$E_3 = hE[\hat{\lambda}_1^2(z_1' W_{11}^{-2} z_1 + 2y' W_{22\cdot 1}^{-1} W_{21} W_{11}^{-2} z_1 + y' W_{22\cdot 1}^{-2} y + y' W_{22\cdot 1}^{-1} W_{21} W_{11}^{-2} W_{12} W_{22\cdot 1}^{-1} y)], \tag{A82}$$

$$E_5 = hE[\hat{\lambda}_1 \hat{\lambda}_2 (z_1' W_{11}^{-2} z_1 + y' W_{22\cdot 1}^{-1} W_{21} W_{11}^{-2} z_1)]. \tag{A83}$$

The expectation of the terms that involve $\tilde{e}_1' W_{11}^{-1} z_1$, $z_1' W_{11}^{-2} z_1$, $\hat{e}_1' W_{22\cdot 1}^{-1} y$, and $y' W_{22\cdot 1}^{-2} y$ have already been derived. It remains to obtain the following three terms

$$D_1 = E[\hat{\lambda}_1 \tilde{e}_1' W_{11}^{-1} W_{12} W_{22\cdot 1}^{-1} y], \tag{A84}$$

$$D_2 = E[\hat{\lambda}_1^2 y' W_{22\cdot 1}^{-1} W_{21} W_{11}^{-2} z_1], \tag{A85}$$

$$D_3 = E[\hat{\lambda}_1^2 y' W_{22\cdot 1}^{-1} W_{21} W_{11}^{-2} W_{12} W_{22\cdot 1}^{-1} y]. \tag{A86}$$

Using (A70) and the fact that $s_2 \sim N(0_{N-1}, I_{N-1})$ and $w_2 \sim \chi^2_{h-M+1}$, we can rewrite these three terms as

$$\begin{aligned} D_1 &= E[\hat{\lambda}_1 \tilde{e}_1' W_{11}^{-1} W_{12} (\eta_2 \eta_2' + T_2 T_2') W_{22\cdot 1}^{-1} y] \\ &= E\left[\hat{\lambda}_1 \left(\frac{\tilde{e}_1' W_{11}^{-1} W_{12} y}{u_2} + \frac{\tilde{e}_1' W_{11}^{-1} W_{12} T_2 s_2 \sqrt{(1+\hat{\theta}_1^2)\tilde{u}_2}}{u_2 \sqrt{w_2}}\right)\right] \\ &= E\left[\frac{\hat{\lambda}_1}{u_2} \tilde{e}_1' W_{11}^{-\frac{1}{2}} Z' y\right], \end{aligned} \tag{A87}$$

$$D_2 = E[\hat{\lambda}_1^2 y' W_{22\cdot 1}^{-1} (\eta_2 \eta_2' + T_2 T_2') W_{21} W_{11}^{-2} z_1]$$

$$
\begin{aligned}
&= E\left[\hat{\lambda}_1^2\left(\frac{y'W_{21}W_{11}^{-2}z_1}{u_2}+\frac{s_2\sqrt{(1+\hat{\theta}_1^2)\tilde{u}_2}T_2'W_{21}W_{11}^{-2}z_1}{u_2\sqrt{w_2}}\right)\right] \\
&= E\left[\frac{\hat{\lambda}_1^2\hat{\theta}_1}{u_2}y'ZW_{11}^{-1}\xi_1\right], \qquad \text{(A88)}
\end{aligned}
$$

$$
\begin{aligned}
D_3 &= E[\hat{\lambda}_1^2 y'W_{22\cdot 1}^{-1}(\eta_2\eta_2'+T_2T_2')W_{21}W_{11}^{-2}W_{12}(\eta_2\eta_2'+T_2T_2')W_{22\cdot 1}^{-1}y] \\
&= E\left[\hat{\lambda}_1^2\left(\frac{y'W_{21}W_{11}^{-2}W_{12}y}{u_2^2}+\frac{2y'W_{21}W_{11}^{-2}W_{12}T_2s_2\sqrt{(1+\hat{\theta}_1^2)\tilde{u}_2}}{u_2^2\sqrt{w_2}}\right.\right. \\
&\qquad\left.\left.+\frac{(1+\hat{\theta}_1^2)\tilde{u}_2s_2'(T_2'W_{21}W_{11}^{-2}W_{12}T_2)s_2}{u_2^2w_2}\right)\right] \\
&= E\left[\frac{\hat{\lambda}_1^2}{u_2^2}\left(y'W_{21}W_{11}^{-2}W_{12}y+\frac{(1+\hat{\theta}_1^2)\tilde{u}_2\mathrm{tr}(T_2'W_{21}W_{11}^{-2}W_{12}T_2)}{h-M-1}\right)\right] \\
&= E\left[\frac{\hat{\lambda}_1^2}{u_2^2}\left(\frac{(1+\hat{\theta}_1^2)\tilde{u}_2\mathrm{tr}(W_{21}W_{11}^{-2}W_{12})}{h-M-1}+\frac{(h-M-2)y'W_{21}W_{11}^{-2}W_{12}y}{h-M-1}\right)\right] \\
&= E\left[\frac{\hat{\lambda}_1^2}{u_2^2}\left(\frac{(1+\hat{\theta}_1^2)\tilde{u}_2\mathrm{tr}(ZW_{11}^{-1}Z')}{h-M-1}+\frac{(h-M-2)y'ZW_{11}^{-1}Z'y}{h-M-1}\right)\right]. \qquad \text{(A89)}
\end{aligned}
$$

The fourth equality of D_3 is obtained because

$$
\begin{aligned}
\mathrm{tr}(T_2'W_{21}W_{11}^{-2}W_{12}T_2) &= \mathrm{tr}(W_{21}W_{11}^{-2}W_{12}T_2T_2') \\
&= \mathrm{tr}\left(W_{21}W_{11}^{-2}W_{12}(I_N-\eta_2\eta_2')\right) \\
&= \mathrm{tr}(W_{21}W_{11}^{-2}W_{12})-\eta_2'W_{21}W_{11}^{-2}W_{12}\eta_2. \qquad \text{(A90)}
\end{aligned}
$$

Define a $K\times(K-1)$ orthonormal matrix

$$
H = W_{11}^{\frac{1}{2}}T_1(T_1'W_{11}T_1)^{-\frac{1}{2}}, \qquad \text{(A91)}
$$

and it can be seen that $HH' = I_K-\xi_1\xi_1'$. It is easy to show that

$$
\mathrm{vec}(ZH)\sim N\left(0_{N(K-1)}, I_{N(K-1)}\right) \qquad \text{(A92)}
$$

and ZH is independent of W_{11}, y, $\tilde{u}_1$, and u_2. In addition, we have

$$
H'W_{11}^{-1}H = (T_1'W_{11}T_1)^{-1} = A_{22}^{-1}, \qquad \text{(A93)}
$$

which is independent of u_1, u_2, $\tilde{u}_1$, and $y'y$, and

$$E\left[A_{22}^{-1}\right] = \frac{1}{h-K-1} I_{K-1}. \tag{A94}$$

Given the definition of H and the fact that ZH has zero mean, we can use (A60) and (A62) to rewrite D_1 and D_2 as

$$D_1 = E\left[\frac{\hat{\lambda}_1}{u_2}\tilde{e}_1' W_{11}^{-\frac{1}{2}}(\xi_1\xi_1' + HH')Z'y\right] = E\left[\frac{\hat{\lambda}_1 z_{11}}{u_1 u_2 \hat{\theta}_1} y'Z\xi_1\right], \tag{A95}$$

$$D_2 = E\left[\frac{\hat{\lambda}_1^2\hat{\theta}_1}{u_2} y'Z(\xi_1\xi_1' + HH')W_{11}^{-1}\xi_1\right] = \frac{h-2}{h-K-1} E\left[\frac{\hat{\lambda}_1^2 \tilde{u}_1}{u_1^2 u_2 \hat{\theta}_1} y'Z\xi_1\right] \tag{A96}$$

In addition, using

$$\begin{aligned}
& E\left[ZW_{11}^{-1}Z' \middle| u_1, u_2, \tilde{u}_1, y\right] \\
&= E\left[Z(\xi_1\xi_1' + HH')W_{11}^{-1}(\xi_1\xi_1' + HH')Z' \middle| u_1, u_2, \tilde{u}_1, y\right] \\
&= \frac{h-2}{h-K-1} E\left[\frac{Z\xi_1\xi_1'Z'}{u_1} \middle| u_1, u_2, \tilde{u}_1, y\right] + E\left[ZHA_{22}^{-1}H'Z' \middle| u_1, u_2, \tilde{u}_1, y\right] \\
&= \frac{h-2}{h-K-1} E\left[\frac{Z\xi_1\xi_1'Z'}{u_1} \middle| u_1, u_2, \tilde{u}_1, y\right] + \frac{K-1}{h-K-1} I_N,
\end{aligned} \tag{A97}$$

we can simplify D_3 to

$$\begin{aligned}
D_3 &= E\left[\frac{\hat{\lambda}_1^2}{u_2^2}\left(\frac{(1+\hat{\theta}_1^2)\tilde{u}_2}{(h-M-1)}\left[\frac{(h-2)\xi_1'Z'Z\xi_1}{u_1(h-K-1)} + \frac{N(K-1)}{h-K-1}\right]\right.\right. \\
&\quad \left.\left. + \frac{h-M-2}{h-M-1}\left(\frac{(h-2)(y'Z\xi_1)^2}{u_1(h-K-1)} + \frac{(K-1)(1+\hat{\theta}_1^2)\tilde{u}_2}{h-K-1}\right)\right)\right] \\
&= E\left[\frac{\hat{\lambda}_1^2}{u_2^2}\left(\frac{(h-2)(1+\hat{\theta}_1^2)\tilde{u}_2\xi_1'Z'Z\xi_1}{(h-K-1)(h-M-1)u_1} + \frac{(h-2)(h-M-2)(y'Z\xi_1)^2}{(h-K-1)(h-M-1)u_1}\right.\right. \\
&\quad \left.\left. + \frac{(K-1)(h-K-2)(1+\hat{\theta}_1^2)\tilde{u}_2}{(h-K-1)(h-M-1)}\right)\right].
\end{aligned} \tag{A98}$$

Conditional on $W_{11}^{-\frac{1}{2}}$ and y, we have

$$Z\xi_1 | y, W_{11}^{-\frac{1}{2}} z_1 \sim N\left(\frac{\hat{\theta}_1}{\sqrt{1+\hat{\theta}_1^2}}\left(\tilde{y} + \frac{\sqrt{h}\delta}{\sqrt{1+\hat{\theta}_1^2}}\hat{e}_1\right), \frac{1}{1+\hat{\theta}_1^2} I_N\right), \tag{A99}$$

$$y'Z\xi_1|y,\hat{\theta}_1^2 \sim N\left(\hat{\theta}_1\left(\tilde{u}_2+\frac{\sqrt{h}\delta\tilde{y}_1}{\sqrt{1+\hat{\theta}_1^2}}\right),\tilde{u}_2\right). \tag{A100}$$

Using (A99) and (A100), we can then express D_1 to D_3 as

$$D_1 = E\left[\frac{\hat{\lambda}_1 z_{11}}{u_1u_2}\left(\tilde{u}_2+\frac{\sqrt{h}\delta\tilde{y}_1}{\sqrt{1+\hat{\theta}_1^2}}\right)\right], \tag{A101}$$

$$D_2 = \frac{h-2}{h-K-1}E\left[\frac{\hat{\lambda}_1^2\tilde{u}_1}{u_1^2u_2}\left(\tilde{u}_2+\frac{\sqrt{h}\delta\tilde{y}_1}{\sqrt{1+\hat{\theta}_1^2}}\right)\right], \tag{A102}$$

$$\begin{aligned}
D_3 &= E\left[\frac{\hat{\lambda}_1^2}{u_2^2}\left(\frac{(h-2)\tilde{u}_2}{(h-K-1)(h-M-1)u_1}\left(\hat{\theta}_1^2\left(\tilde{u}_2+\frac{2\sqrt{h}\delta\tilde{y}_1}{\sqrt{1+\hat{\theta}_1^2}}+\frac{h\delta^2}{1+\hat{\theta}_1^2}\right)+N\right)\right.\right.\\
&\quad+\frac{(h-2)(h-M-2)}{(h-K-1)(h-M-1)u_1}\left(\hat{\theta}_1^2\left(\tilde{u}_2+\frac{\sqrt{h}\delta\tilde{y}_1}{\sqrt{1+\hat{\theta}_1^2}}\right)^2+\tilde{u}_2\right)\\
&\quad\left.\left.+\frac{(K-1)(h-K-2)(1+\hat{\theta}_1^2)\tilde{u}_2}{(h-K-1)(h-M-1)}\right)\right]\\
&= \frac{(h-2)(h-K-2)}{(h-K-1)(h-M-1)}E\left[\frac{\hat{\lambda}_1^2\tilde{u}_2}{u_1u_2^2}\right]+\frac{h-2}{h-K-1}E\left[\frac{\hat{\lambda}_1^2\hat{\theta}_1^2}{u_1u_2^2}\left(\tilde{u}_2+\frac{\sqrt{h}\delta\tilde{y}_1}{\sqrt{1+\hat{\theta}_1^2}}\right)^2\right]\\
&\quad+\frac{h(h-2)\delta^2}{(h-K-1)(h-M-1)}E\left[\frac{\hat{\lambda}_1^2\hat{\theta}_1^2v_2}{u_1u_2^2(1+\hat{\theta}_1^2)}\right]\\
&\quad+\frac{(K-1)(h-K-2)}{(h-K-1)(h-M-1)}E\left[\frac{\hat{\lambda}_1^2\tilde{u}_2(1+\hat{\theta}_1^2)}{u_2^2}\right].
\end{aligned} \tag{A103}$$

Putting all the terms together and after simplification, we obtain the expressions in Proposition 5. This completes the proof.

References

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259–299.

Bawa, Vijay S., Stephen J. Brown, and Roger W. Klein, 1979, *Estimation Risk and Optimal Portfolio Choice*, North-Holland, Amsterdam.

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

Brown, Stephen J., 1976, Optimal Portfolio Choice under Uncertainty, Ph.D. dissertation, University of Chicago.

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

Dickey, James M., 1967, Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution, *Annals of Mathematical Statistics* 38, 511–518.

Dybvig, Philip H., and Stephen Ross, 1985, The analytics of performance measurement using a security market line, *Journal of Finance* 40, 401–416.

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, Empirical Bayes approach to efficient portfolio selection, *Journal of Financial and Quantitative Analysis* 21, 293–305.

Gibbons, Michael, Stephen Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

Haff, L. R., 1979, An identity for the Wishart distribution with applications, *Journal of Multivariate Analysis* 9, 531–554.

Jensen, Michael C., 1968, The performance of mutual funds in the period 1945–1964, *Journal of Finance* 23, 389–446.

Jobson, J. D., and Bob Korkie, 1982, Potential performance and tests of portfolio efficiency, *Journal of Financial Economics* 10, 433–466.

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, Tasuya, Christian P. Robert, and A. K. Md. Ehasanes Saleh, 1993, Estimation of noncentrality parameters, *Canadian Journal of Statistics* 21, 45–57.

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

Muirhead, Robb J., 1982, *Aspects of Multivariate Statistical Theory* (Wiley, New York).

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

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

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

Ter Horst, Jenke, Frans A. De Roon, Bas J. M. Werker, 2006, Incorporating estimation risk in portfolio choice, in Renneboog, Luc (ed.), *Advances in Corporate Finance and Asset Pricing*, Elsevier, Amsterdam, 449–472.

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.

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

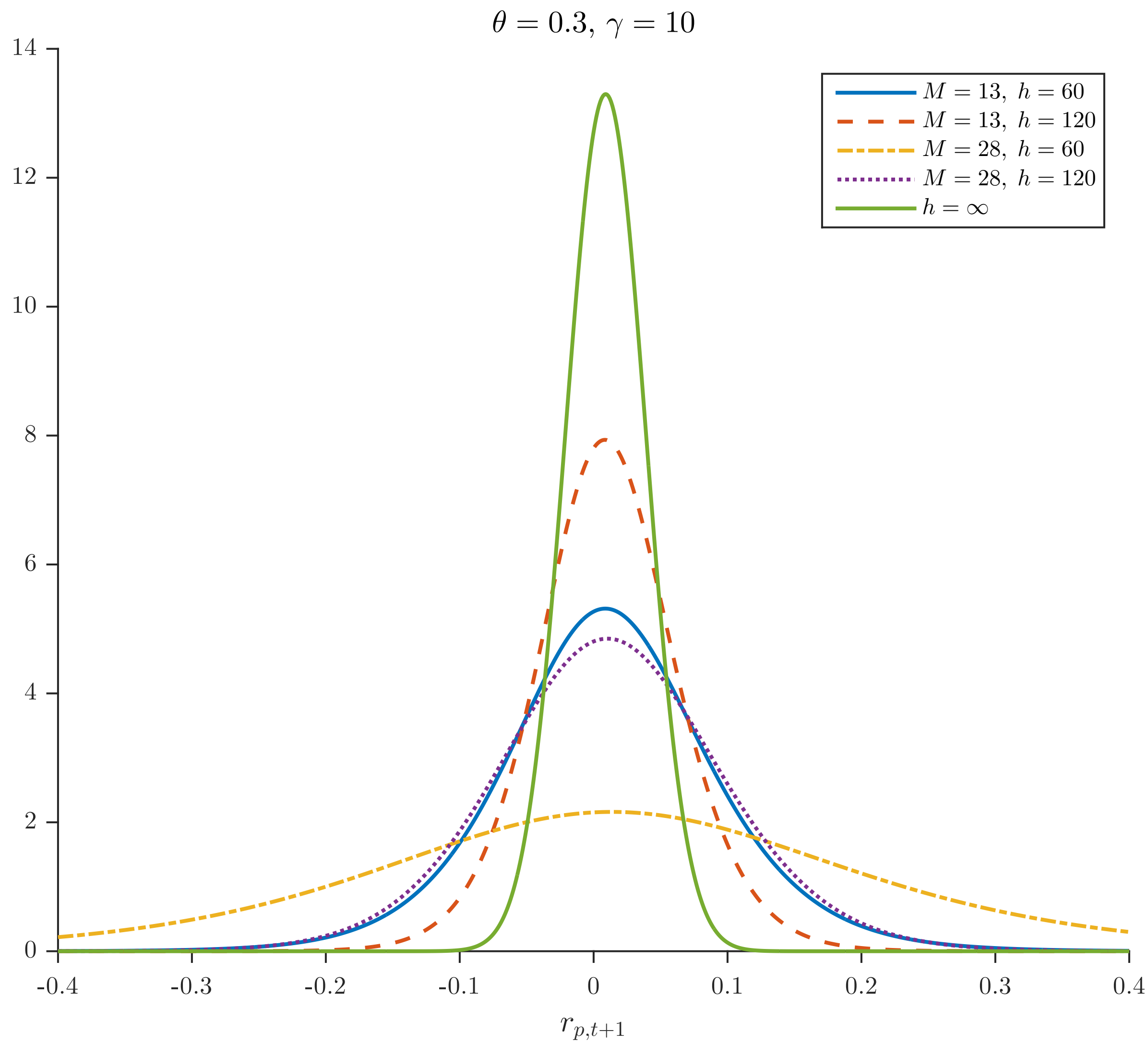

Figure 1
Distribution of Excess Return on the Sample Optimal Portfolio
The figure plots the unconditional distribution of excess return on the sample optimal portfolio for an investor with $\gamma = 10$ under the assumption that the excess returns of the risky assets and the benchmark portfolios are i.i.d. multivariate normal. The Sharpe ratio of the *ex ante* tangency portfolio is assumed to be 0.3 ($\theta = 0.3$). Two different values of the length of estimation window ($h = 60$ or 120) and number of risky assets ($M = 13$ or 28) are examined. For comparison, the figure also plots the distribution of the excess return on the true optimal portfolio ($h = \infty$).

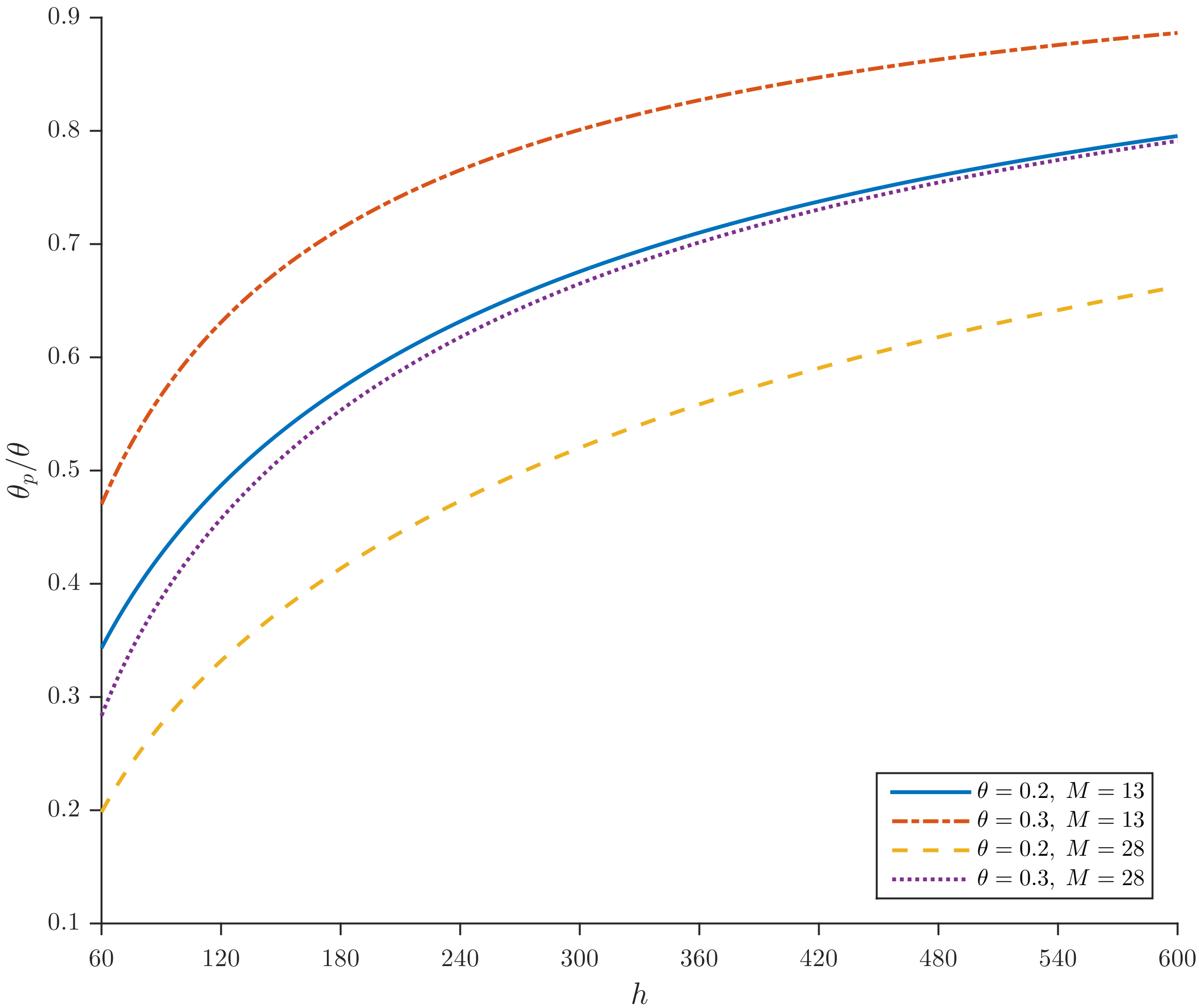

Figure 2
Sharpe Ratio of the Sample Optimal Portfolio as a Function of the Length of Estimation Window
The figure plots the ratio of the unconditonal Sharpe ratio of the sample optimal portfolio (θ_p) to the Sharpe ratio of the true optimal portfolio (θ) as a function of the length of estimation window (h). The excess returns are assumed to be i.i.d. multivariate normal. The four lines in the graph are for different combinations of θ (0.2 or 0.3) and number of risky assets M (13 or 28).

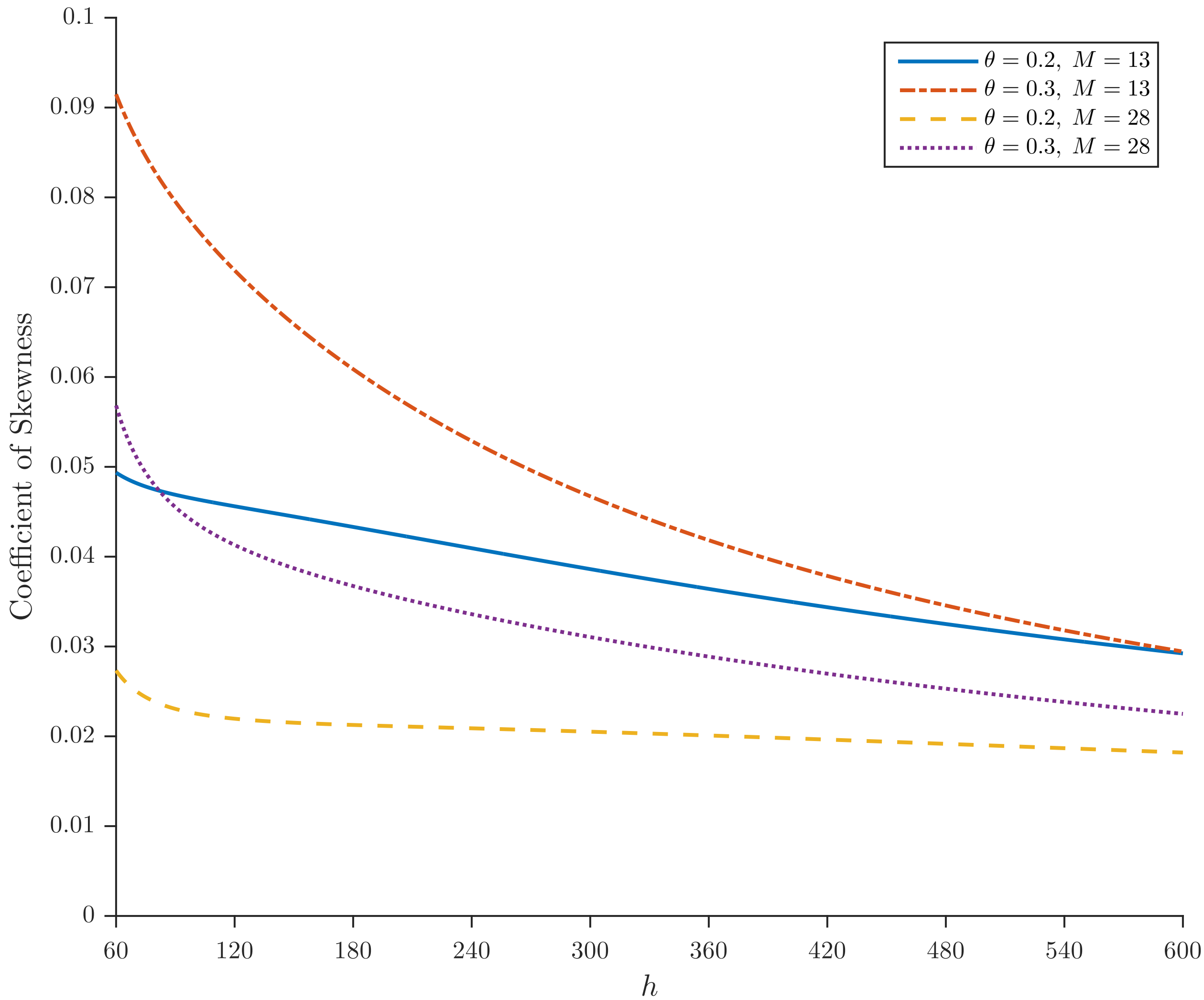

Figure 3
Coefficient of Skewness of the Sample Optimal Portfolio as a Function of the Length of Estimation Window
The figure plots the coefficient of skewness of the excess return of the sample optimal portfolio as a function of the length of estimation window (h). The excess returns are assumed to be i.i.d. multivariate normal. The four lines in the graph are for different combinations of θ (0.2 or 0.3) and number of risky assets M (13 or 28).

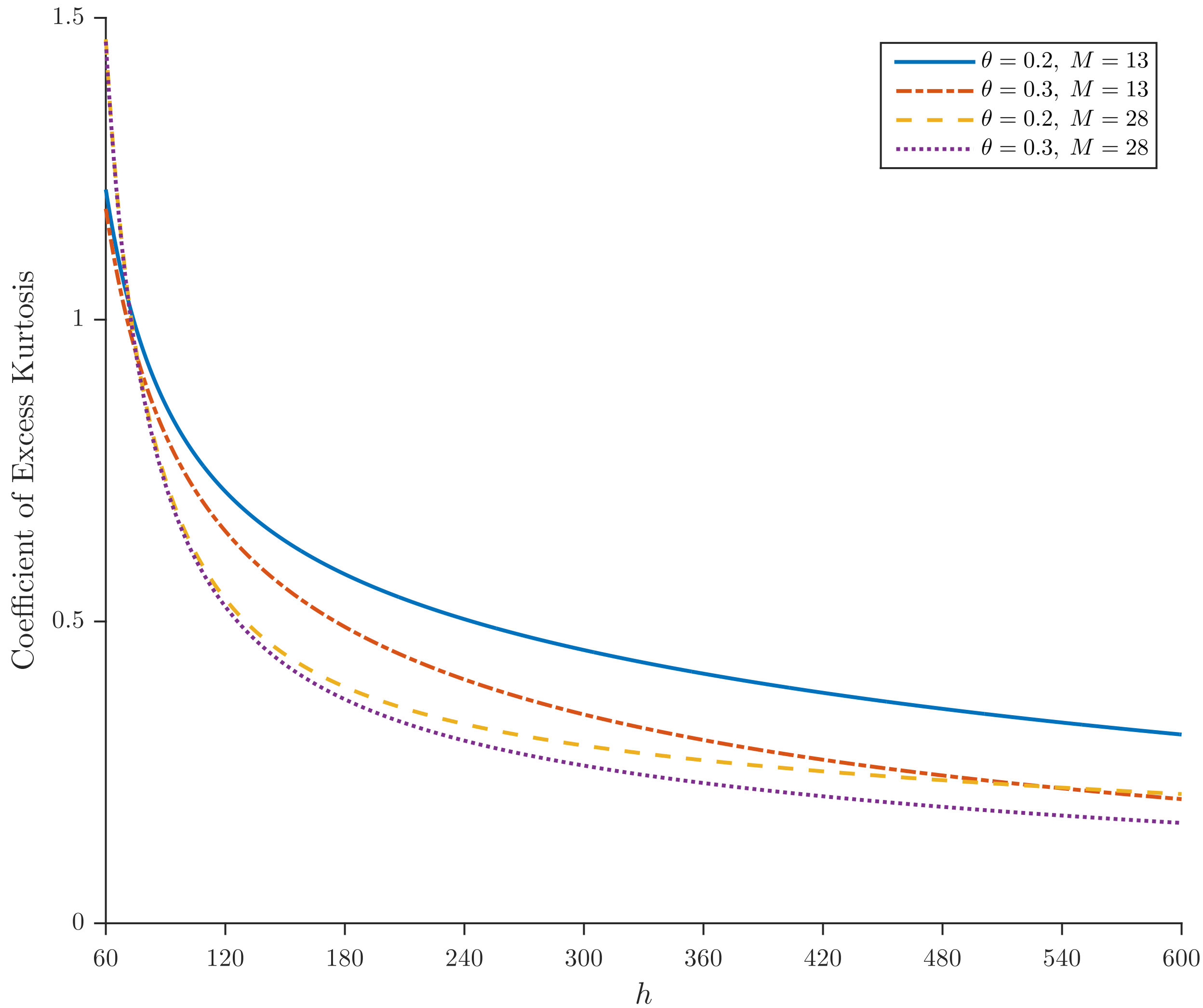

Figure 4
Coefficient of Excess Kurtosis of the Sample Optimal Portfolio as a Function of the Length of Estimation Window
The figure plots the coefficient of excess kurtosis of the excess return of the sample optimal portfolio as a function of the length of estimation window (h). The excess returns are assumed to be i.i.d. multivariate normal. The four lines in the graph are for different combinations of θ (0.2 or 0.3) and number of risky assets M (13 or 28).

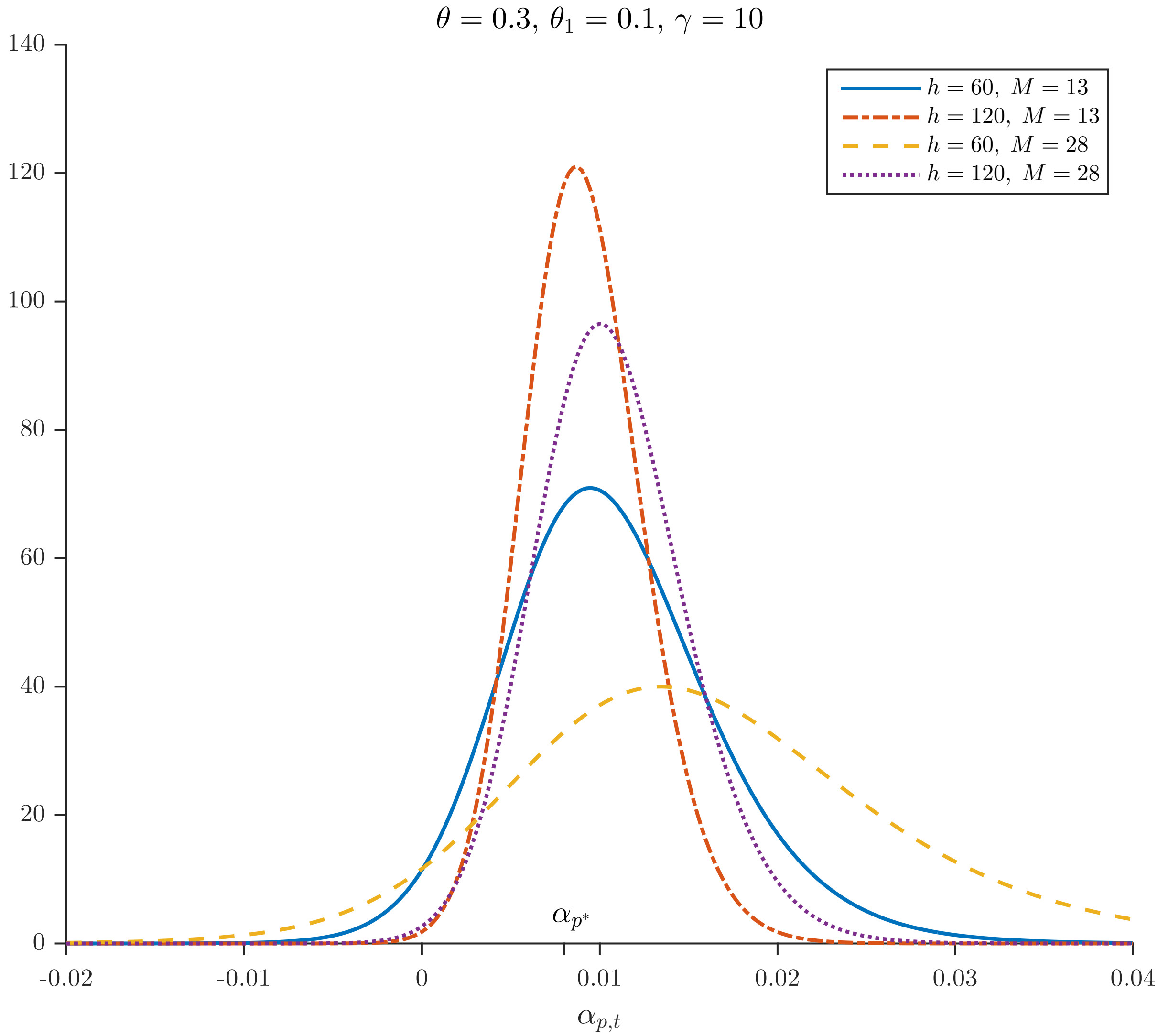

Figure 5
Distribution of Conditional Alpha of the Sample Optimal Portfolio
The figure plots the distribution of conditional alpha of the sample optimal portfolio for an investor with a relative risk aversion of 10 ($\gamma = 10$) under the assumption that the excess returns of the risky assets are i.i.d. multivariate normal. Sharpe ratios of the true optimal portfolio (θ) and the true benchmark optimal portfolio (θ_1) are assumed to be 0.3 and 0.1, respectively. The four lines in the graph are for different combinations of number of risky assets M (13 or 28) and the length of estimation window h (60 or 120). α_{p^*} is the alpha of the true optimal portfolio.

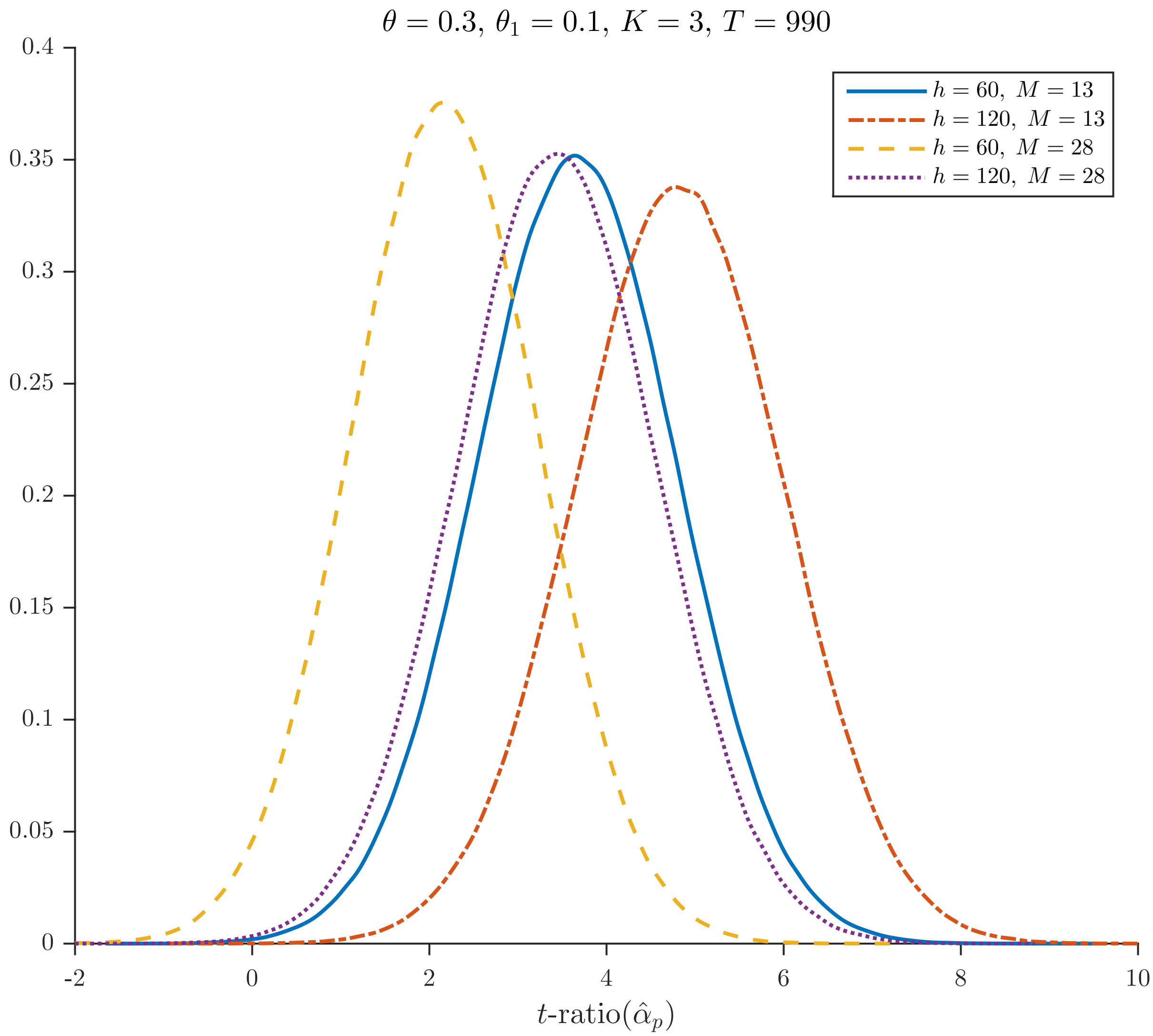

Figure 6
Distribution of OLS t-ratio of Sample Alpha of the Sample Optimal Portfolio
The figure plots the distribution of the OLS t-ratio of sample alpha of the sample optimal portfolio under the assumption that the excess returns are i.i.d. multivariate normal. Sharpe ratios of the true optimal portfolio (θ) and the true benchmark optimal portfolio (θ_1) are assumed to be 0.3 and 0.1, respectively. The number of benchmark portfolios is assumed to be three ($K = 3$) and the length of the time series is set equal to 990 ($T = 990$). The four lines in the graph are for different combinations of number of risky assets M (13 or 28) and the length of estimation window h (60 or 120). The distribution is based on 1,000,000 simulations.

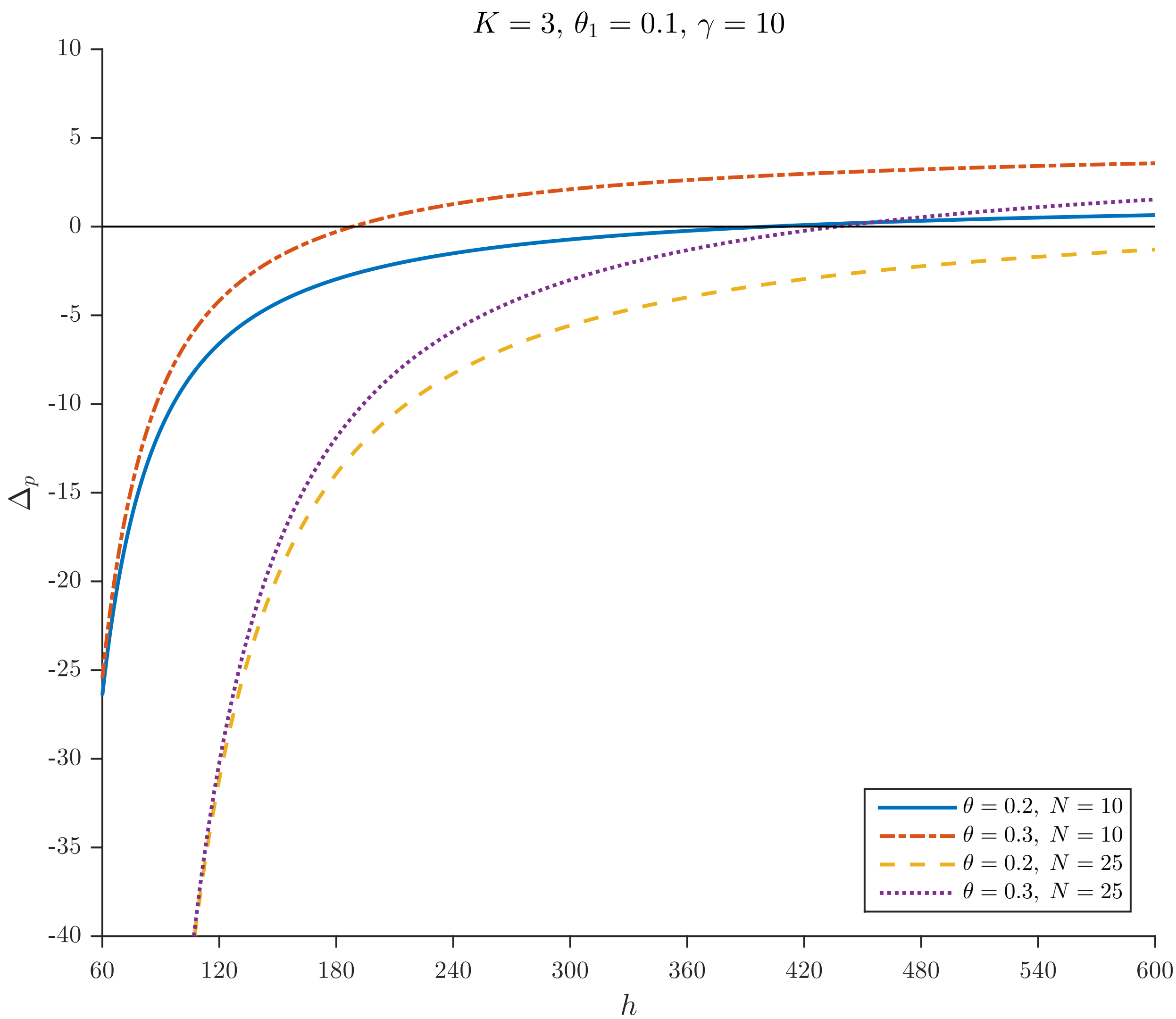

Figure 7
Difference in Out-of-sample Performance between Portfolio p and Portfolio s
The figure plots the difference in expeceted out-of-sample performance (annualized and in percentage points) between portfolio p and portfolio s ($\Delta_p = E[U_p] - E[U_s]$) as a function of the length of estimation window (h) for an investor with $\gamma = 10$. The excess returns are assumed to be i.i.d. multivariate normal. The number of benchmark portfolios is assumed to be three ($K = 3$). The Sharpe ratio of the true benchmark optimal portfolio is assumed to be 0.1 ($\theta_1 = 0.1$). The four lines in the graph are for different combinations of θ (0.2 or 0.3) and N (10 or 25).

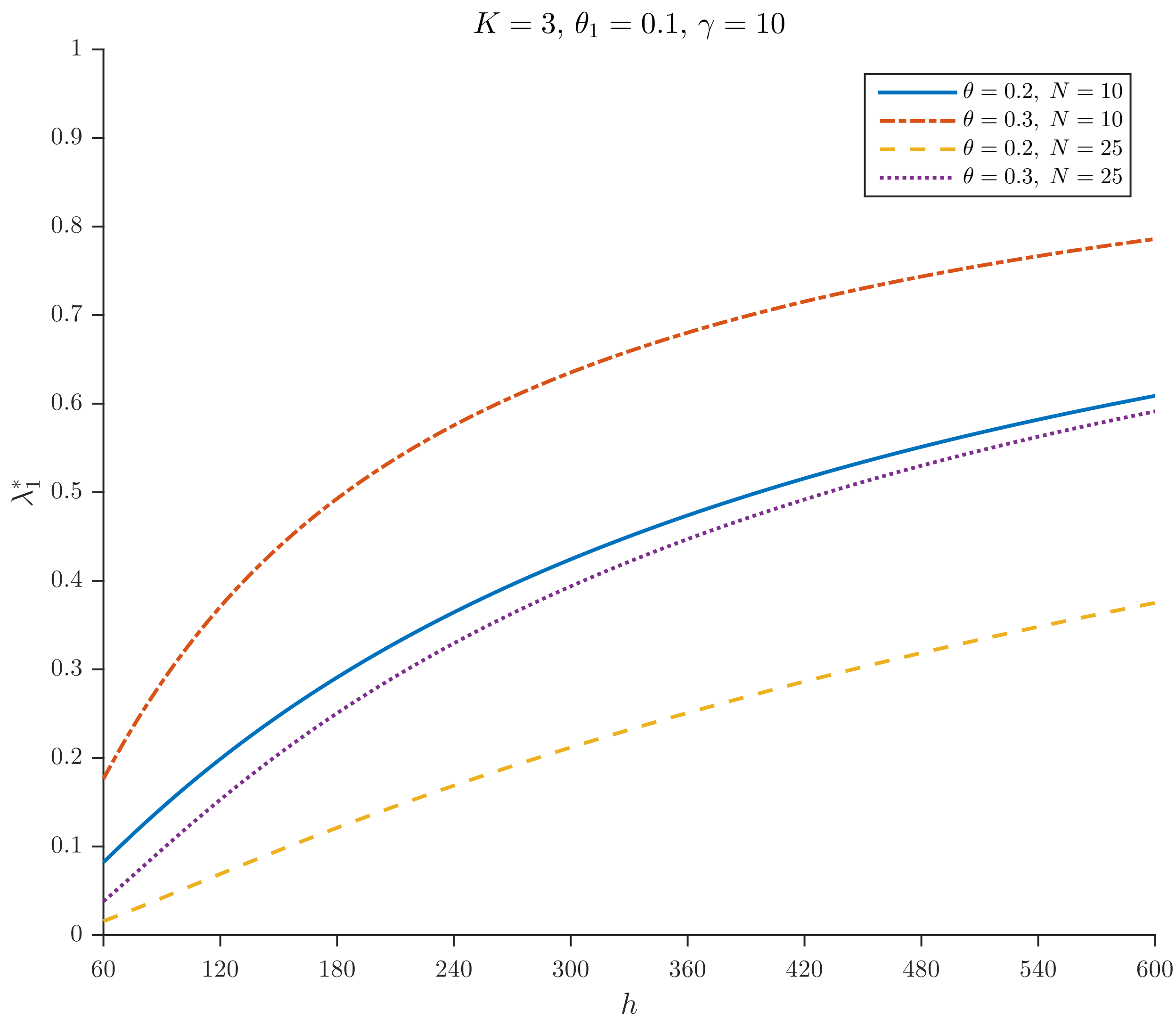

Figure 8
Weight on the Sample Optimal Portfolio in the Optimal Combining Portfolio
The figure plots the weight on the sample optimal portfolio in the optimal combining portfolio (λ_1^*) as a function of the length of estimation window (h) for an investor with $\gamma = 10$, under the assumption that the excess returns are i.i.d. multivariate normal. The number of benchmark portfolios is assumed to be three ($K = 3$). The Sharpe ratio of the true benchmark optimal portfolio is assumed to be 0.1 ($\theta_1 = 0.1$). The four lines in the graph are for different combinations of θ (0.2 or 0.3) and N (10 or 25).

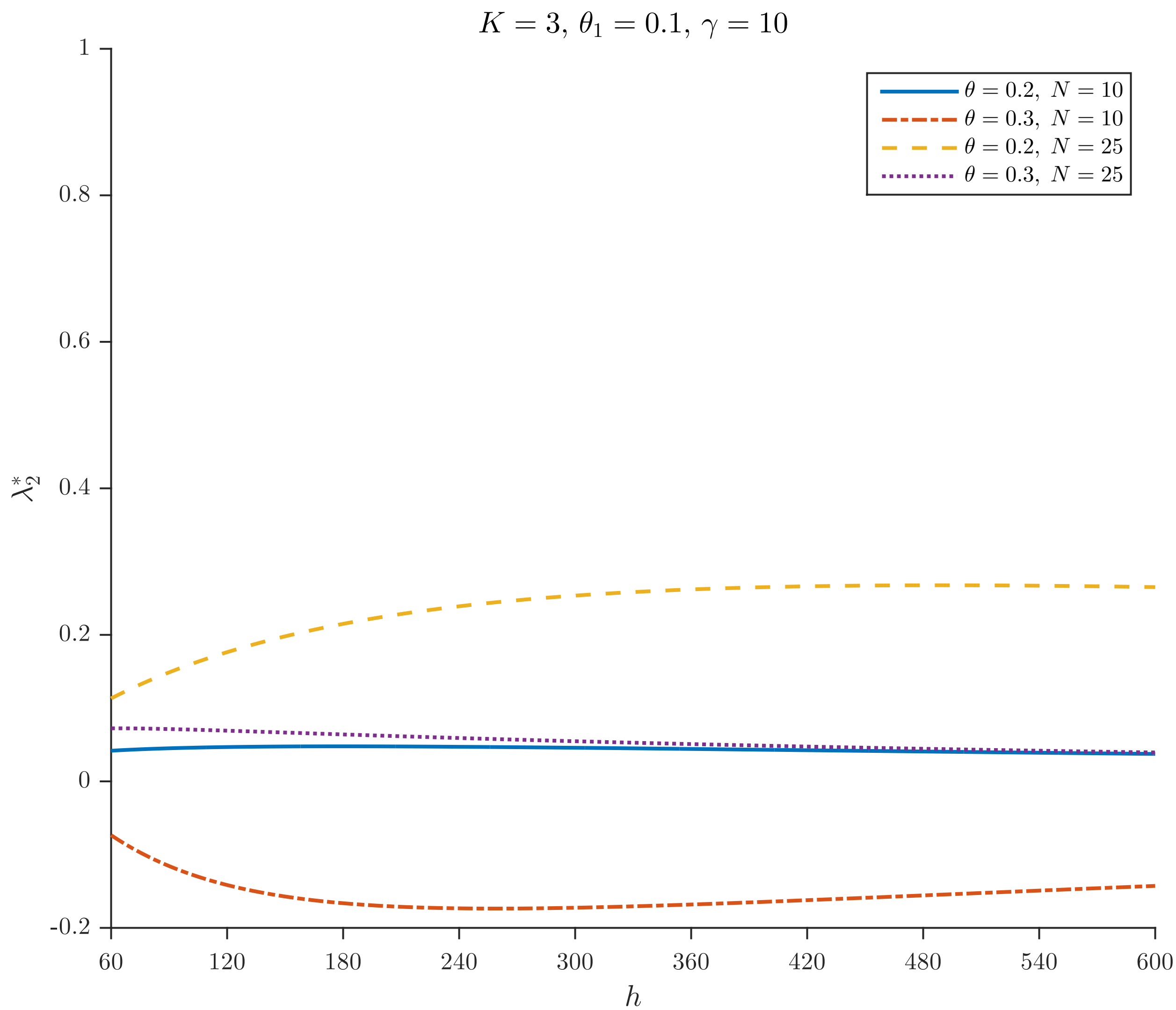

Figure 9
Weight on the Benchmark Sample Optimal Portfolio in the Optimal Combining Portfolio
The figure plots the weight on the benchmark sample optimal portfolio in the optimal combining portfolio (λ_2^*) as a function of the length of estimation window (h) for an investor with $\gamma = 10$, under the assumption that the excess returns are i.i.d. multivariate normal. The number of benchmark portfolios is assumed to be three ($K = 3$). The Sharpe ratio of the true benchmark optimal portfolio is assumed to be 0.1 ($\theta_1 = 0.1$). The four lines in the graph are for different combinations of θ (0.2 or 0.3) and N (10 or 25).

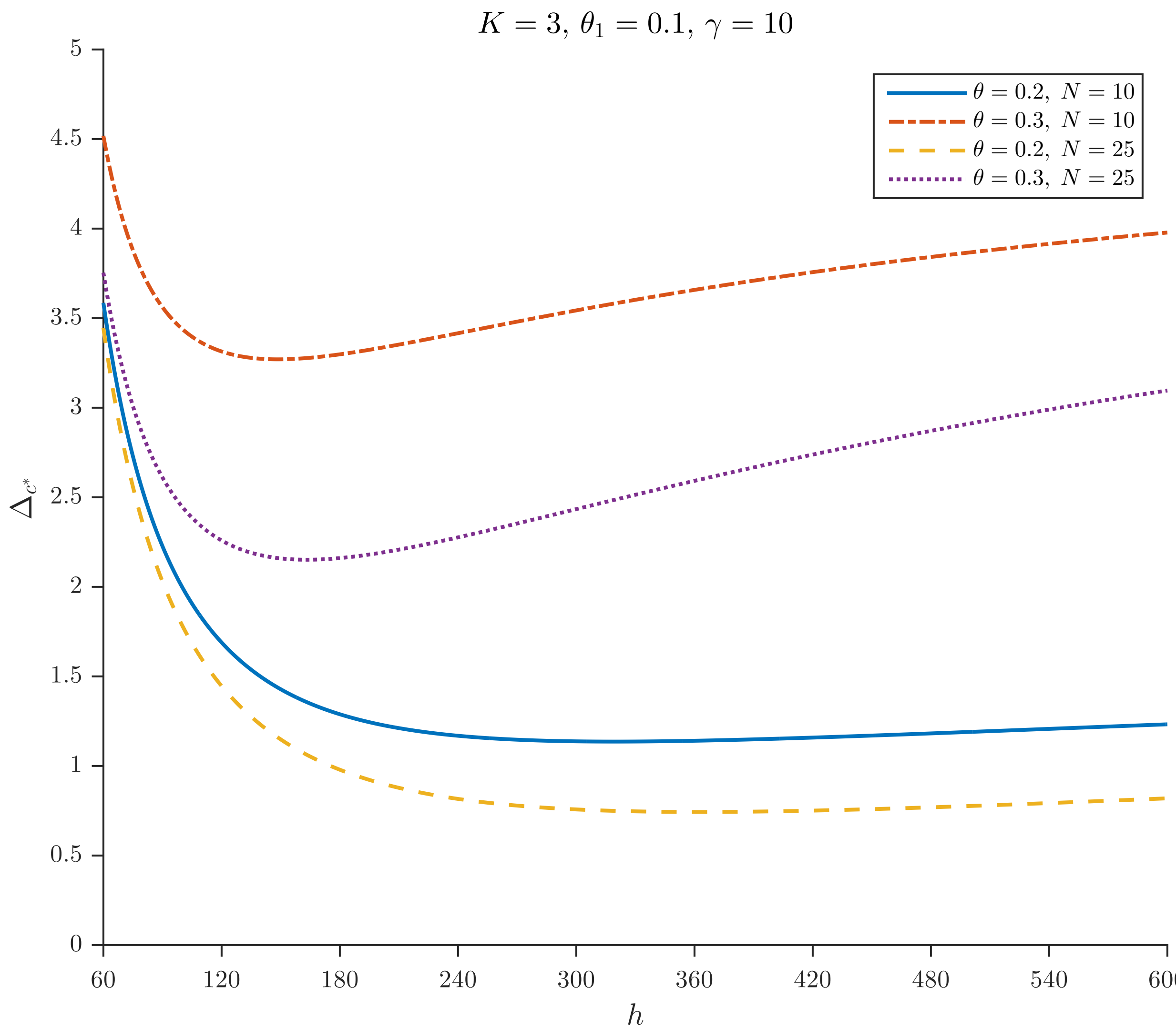

Figure 10
Difference in Expected Out-of-sample Performance between the Optimal Combining Portfolio and the Benchmark Sample Optimal Portfolio
The figure plots the difference in expected out-of-sample performance (annualized and in percentage points) between the optimal combining portfolio and the benchmark sample optimal portfolio ($\Delta_{c^*} = E[U_{c^*}] - E[U_s]$) as a function of the length of estimation window (h) for an investor with $\gamma = 10$, under the assumption that the excess returns are i.i.d. multivariate normal. The number of benchmark portfolios is assumed to be three ($K = 3$). The Sharpe ratio of the true benchmark optimal portfolio is assumed to be 0.1 ($\theta_1 = 0.1$). The four lines in the graph are for different combinations of θ (0.2 or 0.3) and N (10 or 25).

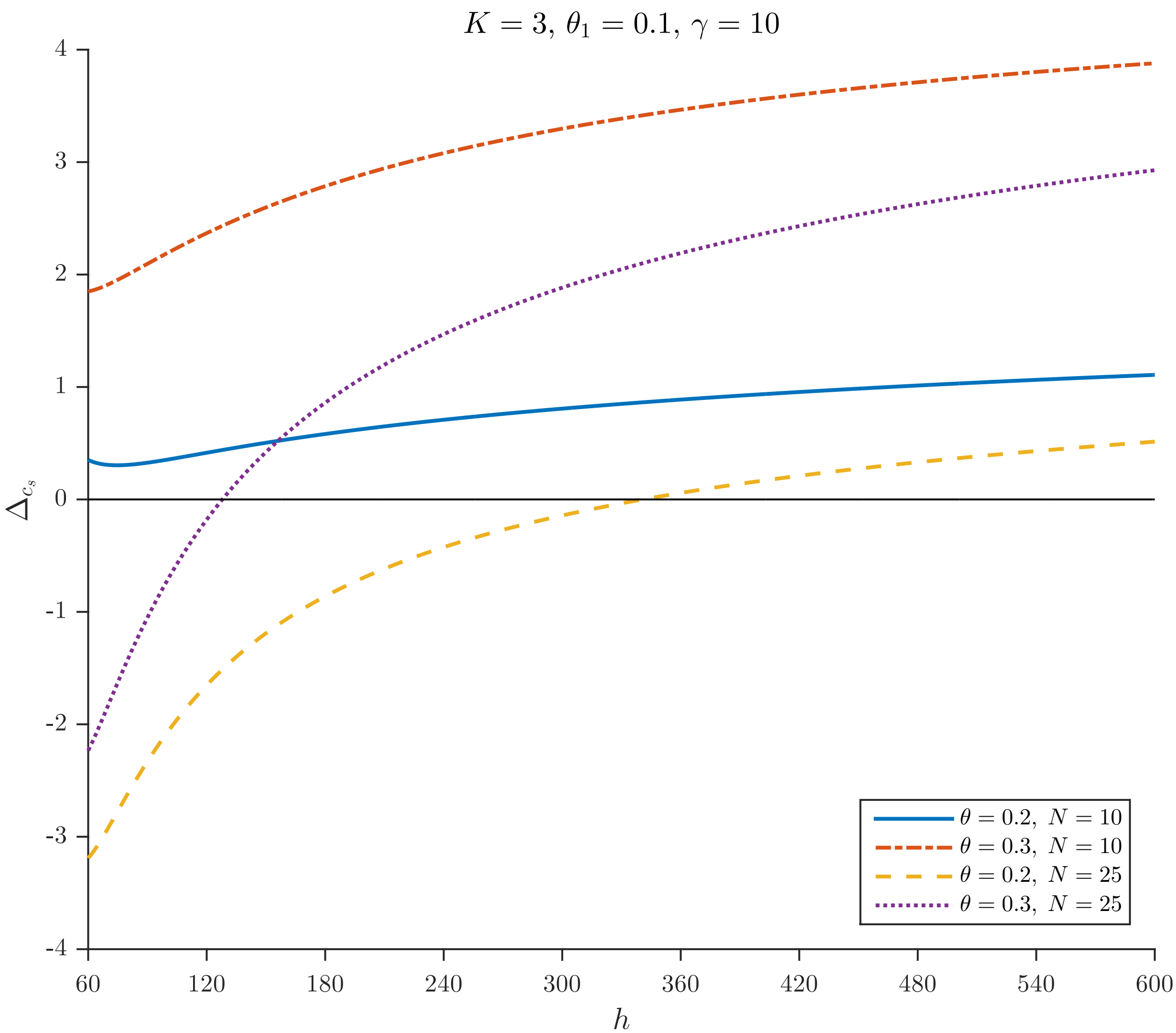

Figure 11
Difference in Expected Out-of-sample Performance between the Implementable Combining Portfolio with Sample Estimates and the Benchmark Sample Optimal Portfolio

The figure plots the difference in expected out-of-sample performance (annualized and in percentage points) of the implementable combining portfolio with sample estimates and the benchmark sample optimal portfolio ($\Delta_{c_s} = E[U_{c_s}] - E[U_s]$) as a function of the length of estimation window (h) for an investor with $\gamma = 10$, under the assumption that the excess returns are i.i.d. multivariate normal. The number of benchmark portfolios is assumed to be three ($K = 3$). The Sharpe ratio of the true benchmark optimal portfolio is assumed to be 0.1 ($\theta_1 = 0.1$). The four lines in the graph are for different combinations of θ (0.2 or 0.3) and N (10 or 25).

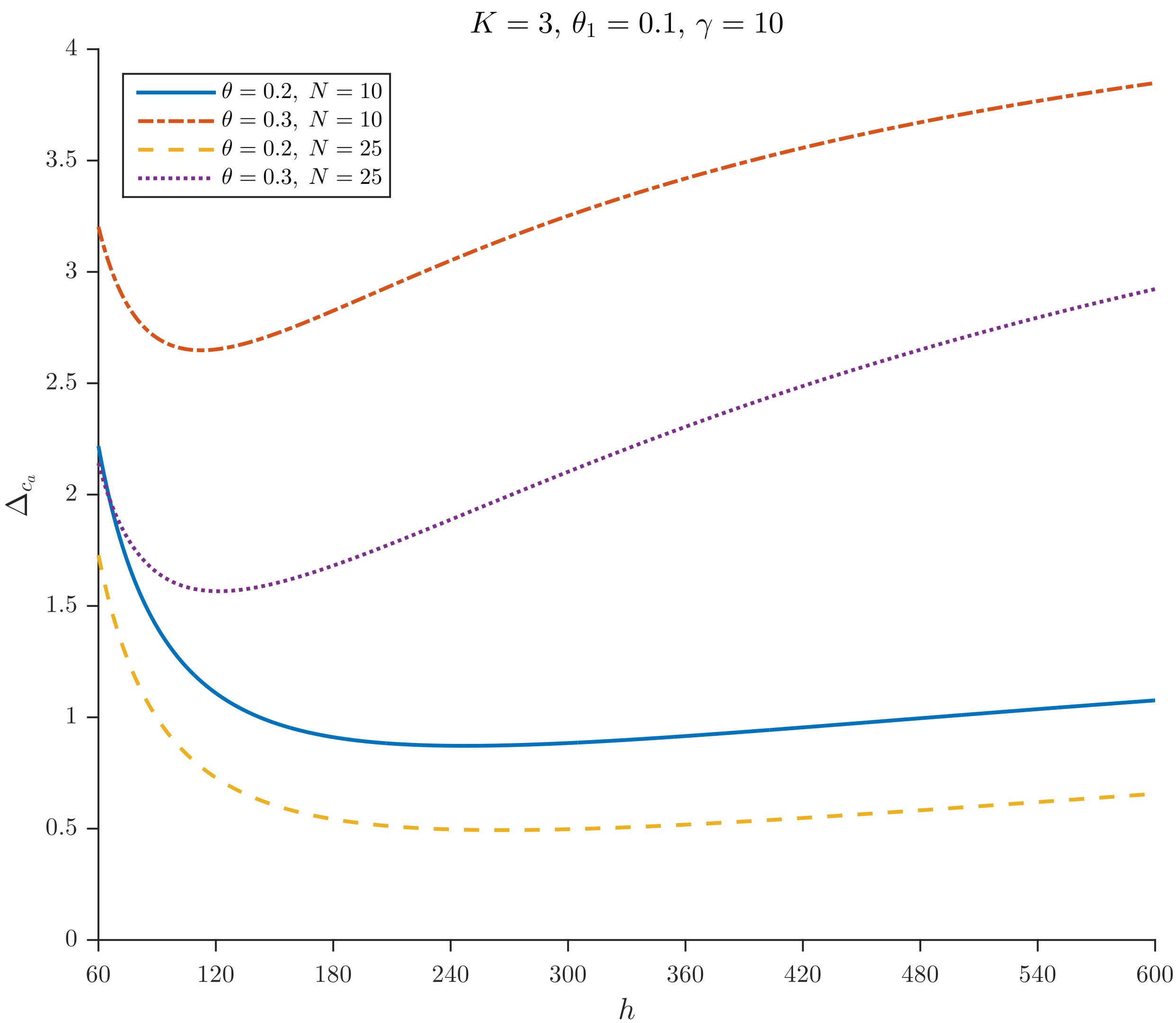

Figure 12
Difference in Expected Out-of-sample Performance between the Implementable Combining Portfolio with Adjusted Estimates and the Benchmark Sample Optimal Portfolio

The figure plots the difference in expected out-of-sample performance (annualized and in percentage points) of the implementable combining portfolio with adjusted estimates and the benchmark sample optimal portfolio ($\Delta_{c_a} = E[U_{c_a}] - E[U_s]$) as a function of the length of estimation window (h) for an investor with $\gamma = 10$, under the assumption that the excess returns are i.i.d. multivariate normal. The number of benchmark portfolios is assumed to be three ($K = 3$). The Sharpe ratio of the true benchmark optimal portfolio is assumed to be 0.1 ($\theta_1 = 0.1$). The four lines in the graph are for different combinations of θ (0.2 or 0.3) and N (10 or 25). The figure is based on 1,000,000 simulations.

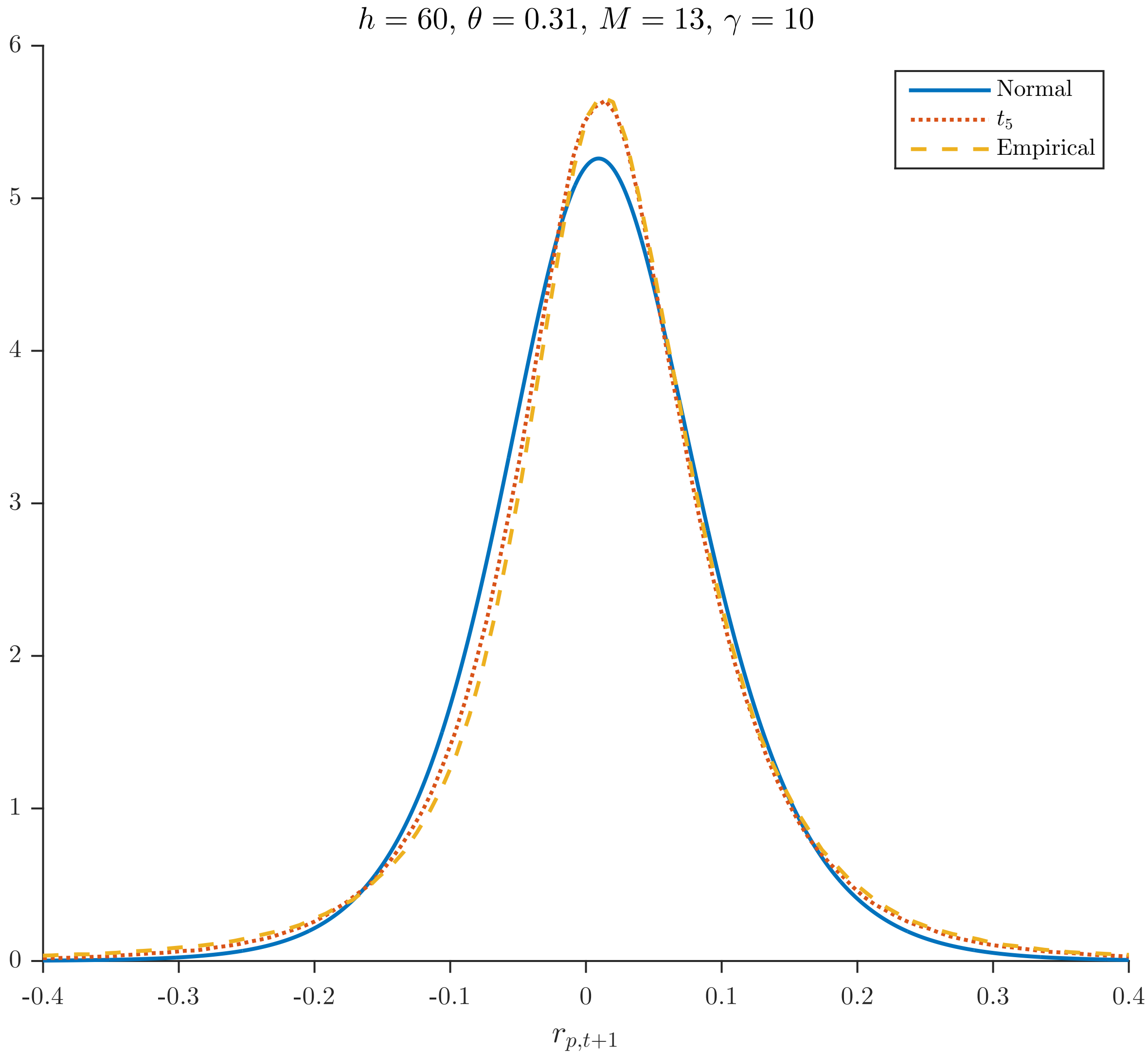

Figure 13
Distribution of Excess Return on the Sample Optimal Portfolio under Various Distributional Assumptions
The figure plots the distribution of excess return on the sample optimal portfolio with 13 risky assets ($M = 13$) for an investor with $\gamma = 10$ under three distributional assumptions on the excess returns of the risky assets: multivariate normal (solid line), multivariate t with five degrees of freedom (dotted line) and an empirical distribution resampled from the monthly excess returns on 10 momentum portfolios and the Fama-French three-factor portfolios over the period of 1931/7 to 2013/12 (dashed line). The mean and covariance matrix of multivariate normal and multivariate t are set equal to those from the empirical distribution. The estimation window is assumed to be $h = 60$. The figure is based on 1,000,000 simulations.

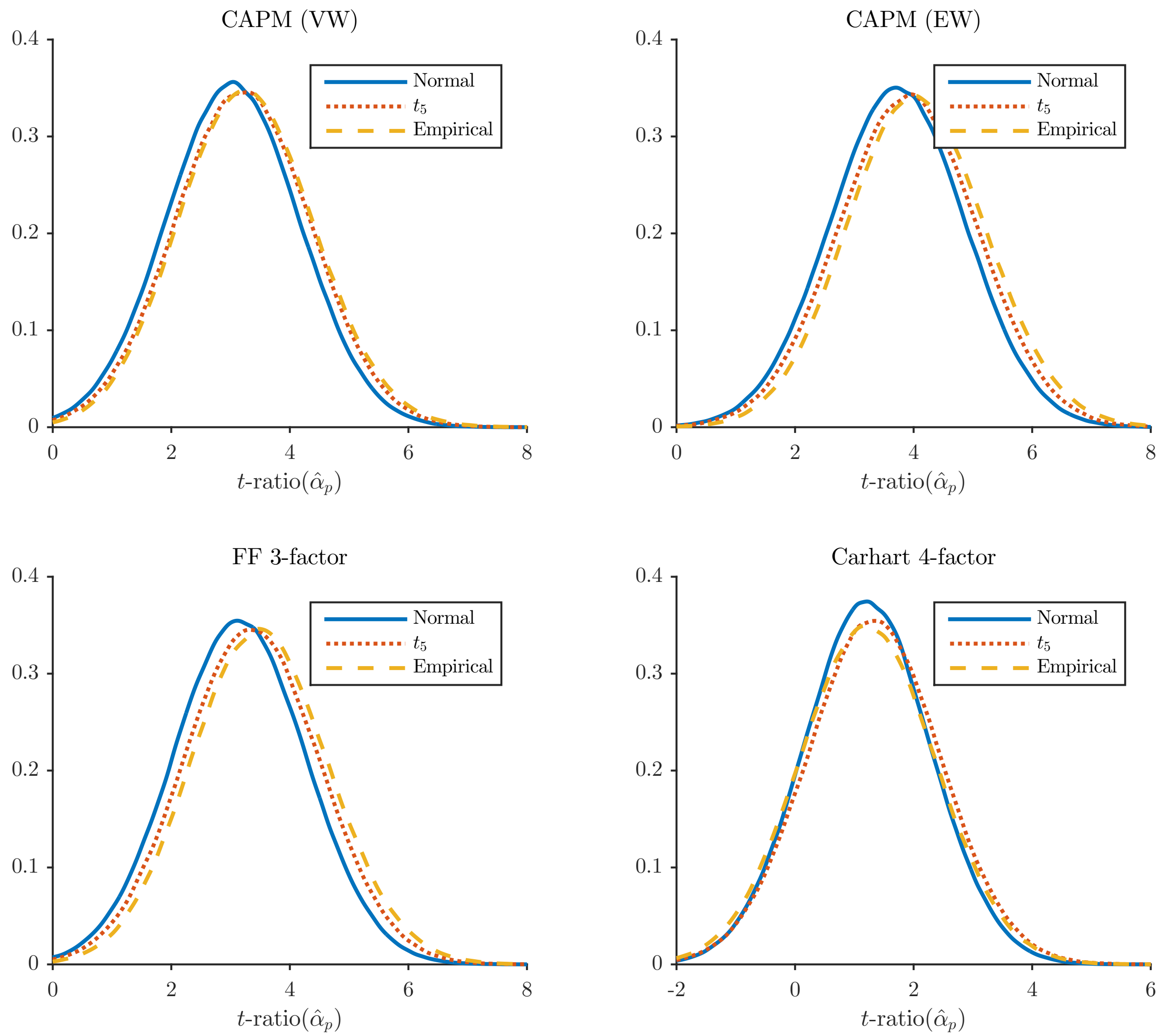

Figure 14
Distribution of OLS t-ratio of Sample Alpha of the Sample Optimal Portfolio under Various Distributional Assumptions
The figure plots the distribution of OLS t-ratio of sample alpha of the sample optimal portfolio with $N = 10$ under three distributional assumptions on the excess returns on the risky assets and the benchmark portfolios: multivariate normal (solid line), multivariate t with five degree of freedom (dotted line) and an empirical distribution resampled from the monthly excess returns on 10 momentum portfolios and the benchmark portfolios over the period of 1931/7 to 2013/12 (dashed line). The mean and covariance matrix of multivariate normal and multivariate t are set equal to those from the empirical distribution. The estimation window is assumed to be $h = 60$. The figure is based on 1,000,000 simulations.

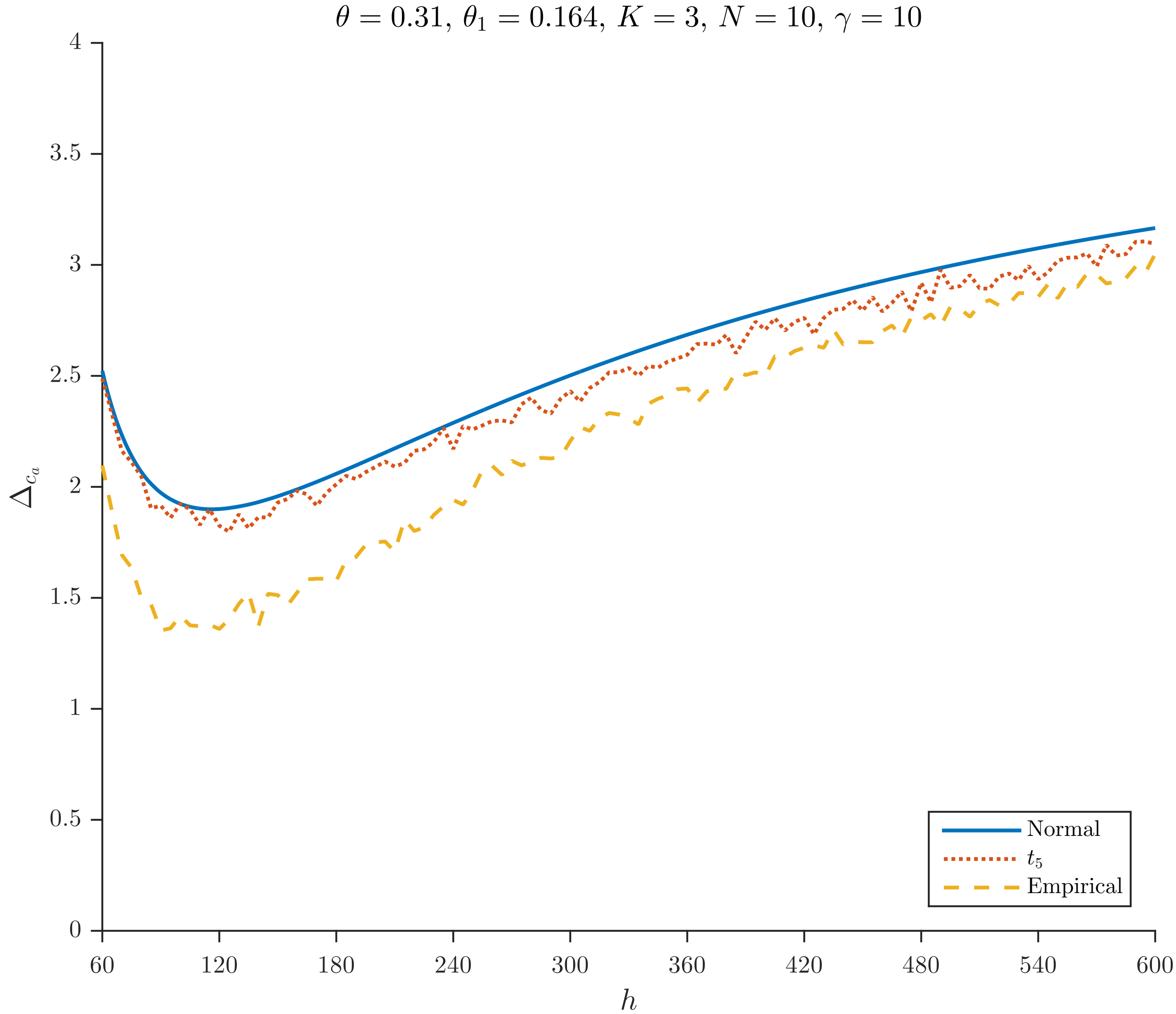

Figure 15
Difference in Expected Out-of-sample Performance between the Implementable Combining Portfolio with Adjusted Estimates and the Benchmark Sample Optimal Portfolio under Various Distributional Assumptions
The figure plots the difference in the expected out-of-sample performance (annualized and in percentage points) of the implementable combining portfolio with adjusted estimates and the benchmark sample optimal portfolio with three benchmark portfolios ($K = 3$) and 10 risky assets ($N = 10$) as a function of the length of estimation window (h) for an investor with $\gamma = 10$ under three distributional assumptions on the excess returns of the risky assets: multivariate normal (solid line), multivariate t with five degrees of freedom (dotted line) and an empirical distribution resampled from the monthly excess returns on 10 momentum portfolios and the Fama-French three-factor portfolios over the period 1931/7 to 2013/12 (dashed line). The mean and covariance matrix of multivariate normal and multivariate t are set equal to those from the empirical distribution. The figure is based on 1,000,000 simulations.

Table I: Alpha of Sample Optimal Portfolios (with $h = 120$ months)

The table presents alpha (in percentage points) and its t-ratios (OLS and White) of the sample optimal portfolio strategy with respect to four asset pricing models: the CAPM with value-weighted market index, the CAPM with equal-weighted market index, the Fama-French three-factor model, and the Carhart four-factor model, assuming $\gamma = 10$. From 1941/7 to 2013/12, a sample optimal portfolio is constructed at the beginning of each month using historical excess returns from the previous 120 months and held for one month. Results are reported for sample optimal portfolios constructed from five different sets of test assets. They are (1) 10 momentum portfolios, (2) 10 volatility portfolios, (3) 10 idiosyncratic volatility portfolios, (4) 10 short-term reversal portfolios, and (5) 5×5 size and B/M portfolios. All portfolios are value-weighted. Sample Sharpe ratio ($\hat{\theta}$) of the sample optimal portfolios are also reported in the table.

	VW	EW	FF3	Carhart4
		10 Momentum Portfolios		
$\hat{\theta}_p$	0.270	0.306	0.310	0.344
$\hat{\alpha}_p$	1.512	1.794	1.835	1.540
$t(\hat{\alpha}_p)$-OLS	7.04	7.93	7.04	5.54
$t(\hat{\alpha}_p)$-White	7.21	8.17	6.98	5.11
		10 Volatility Portfolios		
$\hat{\theta}_p$	0.274	0.411	0.410	0.468
$\hat{\alpha}_p$	1.765	3.938	4.018	4.151
$t(\hat{\alpha}_p)$-OLS	7.16	11.34	10.57	10.44
$t(\hat{\alpha}_p)$-White	7.40	11.71	10.59	10.25
		10 Idiosyncratic Volatility Portfolios		
$\hat{\theta}_p$	0.286	0.480	0.449	0.488
$\hat{\alpha}_p$	1.895	5.040	4.521	4.695
$t(\hat{\alpha}_p)$-OLS	7.48	13.33	11.80	11.30
$t(\hat{\alpha}_p)$-White	7.66	13.52	11.89	10.90
		10 Short-term Reversal Portfolios		
$\hat{\theta}_p$	0.168	0.197	0.204	0.332
$\hat{\alpha}_p$	0.674	0.874	0.638	1.195
$t(\hat{\alpha}_p)$-OLS	3.77	4.61	3.15	4.84
$t(\hat{\alpha}_p)$-White	3.83	4.63	3.15	4.43
		5×5 Size and B/M Portfolios		
$\hat{\theta}_p$	0.261	0.291	0.250	0.298
$\hat{\alpha}_p$	2.525	3.047	2.369	2.048
$t(\hat{\alpha}_p)$-OLS	7.06	7.91	6.02	5.01
$t(\hat{\alpha}_p)$-White	6.88	7.77	5.99	5.00

Table II: Alpha of Sample Optimal Portfolios (with $h = 60$ months)

The table presents alpha (in percentage points) and its t-ratios (OLS and White) of the sample optimal portfolio strategy with respect to four asset pricing models: the CAPM with value-weighted market index, the CAPM with equal-weighted market index, the Fama-French three-factor model, and the Carhart four-factor model, assuming $\gamma = 10$. From 1936/7 to 2013/12, a sample optimal portfolio is constructed at the beginning of each month using historical excess returns from the previous 60 months and held for one month. Results are reported for sample optimal portfolios constructed from five different sets of test assets. They are (1) 10 momentum portfolios, (2) 10 volatility portfolios, (3) 10 idiosyncratic volatility portfolios, (4) 10 short-term reversal portfolios, and (5) 5×5 size and B/M portfolios. All portfolios are value-weighted. Sample Sharpe ratio ($\hat{\theta}$) of the sample optimal portfolios are also reported in the table.

	VW	EW	FF3	Carhart4
		10 Momentum Portfolios		
$\hat{\theta}_p$	0.209	0.262	0.240	0.290
$\hat{\alpha}_p$	1.780	2.317	2.408	2.473
$t(\hat{\alpha}_p)$-OLS	5.60	7.23	6.12	5.51
$t(\hat{\alpha}_p)$-White	5.83	7.39	6.10	5.22
		10 Volatility Portfolios		
$\hat{\theta}_p$	0.256	0.384	0.387	0.429
$\hat{\alpha}_p$	2.518	5.199	5.488	5.517
$t(\hat{\alpha}_p)$-OLS	7.11	11.19	10.86	10.36
$t(\hat{\alpha}_p)$-White	7.35	11.53	10.81	10.45
		10 Idiosyncratic Volatility Portfolios		
$\hat{\theta}_p$	0.285	0.472	0.426	0.463
$\hat{\alpha}_p$	2.832	6.864	6.061	6.383
$t(\hat{\alpha}_p)$-OLS	7.92	13.82	11.96	11.54
$t(\hat{\alpha}_p)$-White	8.09	14.05	12.14	11.89
		10 Short-term Reversal Portfolios		
$\hat{\theta}_p$	0.179	0.224	0.214	0.305
$\hat{\alpha}_p$	1.295	1.759	1.736	2.206
$t(\hat{\alpha}_p)$-OLS	4.58	6.07	5.01	5.58
$t(\hat{\alpha}_p)$-White	4.53	6.12	4.89	5.52
		5×5 Size and B/M Portfolios		
$\hat{\theta}_p$	0.170	0.215	0.151	0.195
$\hat{\alpha}_p$	3.536	4.945	3.643	3.429
$t(\hat{\alpha}_p)$-OLS	4.60	6.09	4.11	3.53
$t(\hat{\alpha}_p)$-White	4.41	5.94	3.86	3.31

Table III: Performance Improvement over the Sample Benchmark Portfolio for Various Portfolio Rules (with $h = 120$ months)

The table presents improvement in out-of-sample performance (annualized and in percentage points) of various portfolio rules over the benchmark sample optimal portfolio (s) for an investor with $\gamma = 10$. The three portfolio rules are (1) sample optimal portfolio (p), (2) combining portfolio of p and s with sample estimates of weights (c_s), and (3) combining portfolio of p and s with adjusted estimates of weights (c_a). Five different sets of test assets are examined against four different benchmarks. The benchmark portfolios are (1) the CAPM with value-weighted market index, (2) the CAPM with equal-weighted market index, (3) the Fama-French three-factor model, and (4) the Carhart four-factor model. The test assets considered are (1) 10 momentum portfolios, (2) 10 volatility portfolios, (3) 10 idiosyncratic volatility portfolios, (4) 10 short-term reversal portfolios, and (5) 5×5 size and B/M portfolios. At the beginning of each month from 1941/7 to 2013/12, portfolios are constructed using historical excess returns from previous 120 months and held for one month. Out-of-sample performance of a given portfolio q is calculated as $\hat{\mu}_q - (\gamma/2)\hat{\sigma}_q^2$, where $\hat{\mu}_q$ and $\hat{\sigma}_q^2$ are the sample mean and variance computed using the 870 monthly excess returns. Δ_q is the performance improvement of a portfolio q (which can be either p, c_s, or c_a) over portfolio s.

	VW	EW	FF3	Carhart4
		10 Momentum Portfolios		
Δ_p	−4.830	−3.502	−8.639	−16.164
Δ_{c_s}	2.168	3.289	2.584	0.310
Δ_{c_a}	2.462	3.523	3.042	1.156
		10 Volatility Portfolios		
Δ_p	−9.259	−12.937	−20.107	−21.850
Δ_{c_s}	0.381	2.650	1.145	2.694
Δ_{c_a}	0.746	3.038	1.605	3.246
		10 Idiosyncratic Volatility Portfolios		
Δ_p	−9.424	−10.988	−15.324	−22.097
Δ_{c_s}	1.198	5.936	5.012	4.172
Δ_{c_a}	1.887	6.334	5.666	5.152
		10 Short-term Reversal Portfolios		
Δ_p	−7.645	−6.923	−10.003	−12.152
Δ_{c_s}	−0.603	−0.115	−0.247	1.328
Δ_{c_a}	0.064	0.400	0.542	2.003
		5×5 Size and B/M Portfolios		
Δ_p	−34.357	−37.806	−44.979	−55.068
Δ_{c_s}	0.944	1.127	−0.202	−0.944
Δ_{c_a}	2.987	2.972	2.207	2.070

Table IV: Performance Improvement over the Sample Benchmark Portfolio for Various Portfolio Rules (with $h = 60$ months)

The table presents improvement in out-of-sample performance (annualized and in percentage points) of various portfolio rules over the benchmark sample optimal portfolio (s) for an investor with $\gamma = 10$. The three portfolio rules are (1) sample optimal portfolio (p), (2) combining portfolio of p and s with sample estimates of weights (c_s), and (3) combining portfolio of p and s with adjusted estimates of weights (c_a). Five different sets of test assets are examined against four different benchmarks. The benchmark portfolios are (1) the CAPM with value-weighted market index, (2) the CAPM with equal-weighted market index, (3) the Fama-French three-factor model, and (4) the Carhart four-factor model. The test assets considered are (1) 10 momentum portfolios, (2) 10 volatility portfolios, (3) 10 idiosyncratic volatility portfolios, (4) 10 short-term reversal portfolios, and (5) 5×5 size and B/M portfolios. At the beginning of each month from 1936/7 to 2013/12, portfolios are constructed using historical excess returns from previous 60 months and held for one month. Out-of-sample performance of a given portfolio q, is calculated as $\hat{\mu}_q - (\gamma/2)\hat{\sigma}_q^2$, where $\hat{\mu}_q$ and $\hat{\sigma}_q^2$ are the sample mean and variance computed using the 930 monthly excess returns. Δ_q is the performance improvement of a portfolio q (which can be either p, c_s, or c_a) over portfolio s.

	VW	EW	FF3	Carhart4
		10 Momentum Portfolios		
Δ_p	−32.354	−26.541	−48.138	−68.535
Δ_{c_s}	−0.787	1.390	1.813	2.160
Δ_{c_a}	1.018	2.541	4.253	5.285
		10 Volatility Portfolios		
Δ_p	−36.841	−53.550	−64.870	−72.036
Δ_{c_s}	−0.240	0.381	4.360	8.132
Δ_{c_a}	1.632	2.072	6.205	10.011
		10 Idiosyncratic Volatility Portfolios		
Δ_p	−34.611	−50.386	−58.941	−72.085
Δ_{c_s}	1.255	5.050	7.325	10.087
Δ_{c_a}	3.062	6.013	8.768	12.032
		10 Short-term Reversal Portfolios		
Δ_p	−26.561	−23.014	−38.190	−48.599
Δ_{c_s}	−0.695	0.757	1.906	5.570
Δ_{c_a}	0.852	1.775	3.970	7.676
		5×5 Size and B/M Portfolios		
Δ_p	−280.717	−298.117	−374.518	−452.742
Δ_{c_s}	−5.789	−4.212	−4.558	−2.006
Δ_{c_a}	−0.162	1.014	1.842	5.182