Optimal Portfolio Selection with and without Risk-free Asset

Raymond Kan, Xiaolu Wang and Guofu Zhou*

JEL classification: G11; G12; C11

Keywords: Portfolio choice; estimation risk; global minimum-variance portfolio; $1/N$ rule.

First version: March, 2015
Current version: May, 2016

*Kan is from the University of Toronto, Wang is from Iowa State University and Zhou is from Washington University in St. Louis. 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. We thank Victor DeMiguel, Lorenzo Garlappi, Thomas Maurer, Thomas Tang, Raman Uppal, Michael Wolf and seminar participants at Australian National University, Hong Kong University of Science and Technology, Shanghai Advanced Institute of Finance, Southwestern University of Finance and Economics, Tsinghua University School of Economics and Management, and participants at the 2nd Financial Econometrics and Risk Management Conference at Western University for helpful comments. Kan gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada and the National Bank Financial Professorship in Capital Markets from the University of Toronto.

Optimal Portfolio Selection with and without Risk-free Asset

In this paper, we consider optimal portfolio problems with and without risk-free asset, taking into account estimation risk. For the case with a risk-free asset, we derive the exact distribution of out-of-sample returns of various optimal portfolio rules, including the two-fund and three-fund rules suggested by Kan and Zhou (2007), and compare their out-of-sample performance with the equally weighted portfolio (i.e., $1/N$ rule). We find that the dominance of the $1/N$ rule over various optimal portfolio rules as documented by DeMiguel, Garlappi, and Uppal (2009) was due in part to the exclusion of risk-free asset in their construction of optimal portfolios, even though those optimal portfolio rules were designed to include the risk-free asset. In order to have a direct comparison with the $1/N$ rule of risky assets only, we also consider an optimal portfolio problem without risk-free asset and develop a new portfolio rule that is designed to mitigate estimation risk in this case. We show that our new portfolio rule performs well relative to the $1/N$ rule in both calibrations and real datasets.

1. Introduction

Although many sophisticated portfolio selection models have been developed since Markowitz's (1952) seminal paper, the mean-variance framework is still the major model used in practice today in asset allocation and active portfolio management.[1] One main reason is that many implementation issues, such as factor exposures and trading constraints, can be easily accommodated within this framework which allows for analytical insights and fast numerical solutions. Another reason is that the intertemporal hedging demand is found typically small so that independent returns over time is a workable assumption in the real world. However, to apply the mean-variance framework in practice, the true parameters are unknown and have to be estimated from data. When estimated parameters instead of true parameters are used in an optimal portfolio rule, there can be a substantial deterioration of performance. Brown (1976), Bawa, Brown, and Klein (1979), and Jorion (1986) are examples of earlier work that provide sophisticated portfolio rules to mitigate the estimation risk. Recently, Kan and Zhou (2007), and Tu and Zhou (2011), among others, provide explicit portfolio rules that are designed to reduce the impact of estimation risk.

In a thought provoking paper, DeMiguel, Garlappi, and Uppal (2009, DGU hereafter) compare the equally weighted portfolio ($1/N$ rule) with the sample-based mean-variance portfolio rule as well as a host of more sophisticated rules. They find that

> *Based on parameters calibrated to the US equity market, our analytical results and simulations show that the estimation window needed for the sample-based mean-variance strategy and its extensions to outperform the* $1/N$ *benchmark is around 3000 months for a portfolio with 25 assets and about 6000 months for a portfolio with 50 assets. This suggests that there are still many "miles to go" before the gains promised by optimal portfolio choice can actually be realized out of sample.*

Note that there is an important condition for the above statement to hold, that is, the $1/N$ rule has a Sharpe ratio that is close to the true optimal. For example, the above statement was obtained under the assumption that the Sharpe ratio of the true optimal portfolio is 0.15, whereas the Sharpe ratio of the equally weighted portfolio is 0.12. In their simulations, the parameters were chosen such

[1]See Grinold and Kahn (1999), Litterman (2003), Meucci (2005), Qian, Hua, and Sorensen (2007) for practical applications of the mean-variance framework.

that the Sharpe ratio of the true optimal portfolio is 0.1477, whereas the Sharpe ratio of the equally weighted portfolio ranges from 0.1356 to 0.1466. However, such condition does not always hold in actual data. Therefore, it is surprising that DGU find outperformance of the $1/N$ rule even in some datasets where such condition does not seem to hold. It turns out that DGU are in fact comparing the $1/N$ rule with optimal portfolios that are normalized to be fully invested in the risky assets. This makes the comparison inappropriate because the optimal portfolio rules that they considered were derived under the assumption that the investor has access to a risk-free asset in addition to the risky assets. We show in this paper that the out-of-sample returns of these normalized version of optimal portfolios actually have no finite moments. This explains why the normalized optimal portfolios have poor out-of-sample performance. If the risk-free asset is allowed, we show that the optimal portfolios perform significantly better and they beat the $1/N$ rule in many cases.

However, there are situations where one is interested in fully investing in risky assets, especially in the context of delegated portfolio management because mutual funds and institutional equity funds are often required to be fully invested in the equity market. As a result, we are interested in comparing the out-of-sample performance of $1/N$ rule with optimal portfolios which are based on just risky assets. Instead of simply rescaling an optimal portfolio designed for the case with risk-free asset as done in DGU, we consider in this paper the optimal portfolio choice problem for the case without a risk-free asset, and provide a portfolio rule that mitigates the impact of estimation risk. Specifically, instead of plugging in the sample mean and covariance matrix into the optimal portfolio formula, we adjust the portfolio by a suitable function of the data designed in such a way to optimally account for the estimation risk. We derive the exact distribution of the out-of-sample return of an implementable version of this optimal portfolio rule. Under reasonable length of estimation window used in practice, we find that our new rule performs well. It dominates the plug-in rule which can perform quite poorly. In addition, we show that our new portfolio rule outperforms the $1/N$ rule in our calibrations. Moreover, for the same empirical datasets used by DGU, our new rule also outperforms the $1/N$ rule in most cases. While insights from DGU cast some doubt on the value of existing investment theory, our paper re-affirms the usefulness of a host of portfolio rules that take into account of the impact of estimation risk.

The remainder of the paper is organized as follows. Section 2 considers optimal portfolio rules for the case with a risk-free asset. Section 3 provides an analysis of the optimal portfolio rules for the case without a risk-free asset. Section 4 concludes.

2. Portfolio Rules with a Risk-free Asset

In this section, we discuss the portfolio choice problem for a mean-variance investor when both risky assets and a risk-free asset are available. In particular, we present the exact distribution of the out-of-sample returns and an analysis of the out-of-sample performance of various portfolio rules in the presence of estimation risk.

2.1 The Setup

Consider the standard portfolio choice problem in which an investor chooses his optimal portfolio among N risky assets and a risk-free asset. Denote the excess returns (in excess of risk-free rate) of the N risky assets at time t by r_t, with mean μ and covariance matrix Σ. Let w be the weights of a portfolio on the N risky assets, and $1 - 1_N' w$ is invested in the risk-free asset, where 1_N stands for the $N \times 1$ vector of ones. Under the standard mean-variance framework, the investor chooses an optimal portfolio to maximize the following utility function

$$U(w) = w'\mu - \frac{\gamma}{2} w'\Sigma w, \tag{1}$$

where γ is the coefficient of risk aversion. Practitioners (see, e.g., Qian, Hua, and Sorensen (2007)) often use this utility set-up because of its convenient interpretation that $U(w)$ is the risk-adjusted return (certainty equivalent).

When μ and Σ are known, it is straightforward to show that the optimal portfolio has the following weights on the risky assets

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

and the utility of holding this optimal portfolio is

$$U(w^*) = \frac{\theta^2}{2\gamma}, \tag{3}$$

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

2.2 Optimal Portfolio Rules with Estimation Risk

In reality, μ and Σ are unknown to investors, and they need to be estimated. We assume an investor estimates μ and Σ using an estimation window of h periods of historical return data, where $h > N$. For analytical tractability, we make the common assumption that r_t is independent and identically distributed over time, and has a multivariate normal distribution. Under this assumption, the maximum likelihood estimators of μ and Σ at time t are given by

$$\hat{\mu}_t = \frac{1}{h} \sum_{s=t-h+1}^{t} r_s, \tag{4}$$

$$\hat{\Sigma}_t = \frac{1}{h} \sum_{s=t-h+1}^{t} (r_s - \hat{\mu}_t)(r_s - \hat{\mu}_t)'. \tag{5}$$

The simplest way to estimate w^* is to replace μ and Σ in (2) by $\hat{\mu}_t$ and $\hat{\Sigma}_t$ to obtain

$$\hat{w}_t = \frac{1}{\gamma} \hat{\Sigma}_t^{-1} \hat{\mu}_t. \tag{6}$$

Since $\hat{w}_t$ is a maximum likelihood estimator of w^*, we call $\hat{w}_t$ the maximum likelihood (ML) rule. The expected out-of-sample utility of the ML rule is defined as

$$E[U(\hat{w}_t)] = E[\hat{w}_t' \mu] - \frac{\gamma}{2} E[\hat{w}_t' \Sigma \hat{w}_t]. \tag{7}$$

Since $\hat{w}_t \neq w^*$ with probability one, we have $E[U(\hat{w}_t)] < U(w^*)$. Specifically, Kan and Zhou (2007) show that when $h > N+4$,

$$E[U(\hat{w}_t)] = \frac{k_1 \theta^2}{2\gamma} - \frac{Nh(h-2)}{2\gamma(h-N-1)(h-N-2)(h-N-4)}, \tag{8}$$

where

$$k_1 = \left(\frac{h}{h-N-2} \right) \left[2 - \frac{h(h-2)}{(h-N-1)(h-N-4)} \right] < 1. \tag{9}$$

Although conditional on $\hat{w}_t$, the out-of-sample return of the ML rule at time $t+1$ (i.e., $\hat{w}_t' r_{t+1}$) is normally distributed, its unconditional distribution is not normally distributed. The exact unconditional distribution of the out-of-sample return of the ML rule is presented in Kan and Wang (2015).

Due to estimation errors in $\hat{\mu}_t$ and $\hat{\Sigma}_t$, the utility loss from the ML rule can be significant, especially when N is large relative to h. In order to mitigate the impact of estimation risk, Kan and Zhou (2007) propose two alternative portfolio rules. The first one is a two-fund rule that is optimal in the class of portfolio rules with weights $c\hat{\Sigma}_t^{-1}\hat{\mu}_t/\gamma$, where c is a scalar. They show that the optimal c is given by

$$c^* = \frac{k_3\theta^2}{\theta^2 + \frac{N}{h}}, \tag{10}$$

where

$$k_3 = \frac{(h-N-1)(h-N-4)}{h(h-2)}. \tag{11}$$

Since $c^* < 1$, the optimal two-fund rule calls for less investment in the sample tangency portfolio. This is because with estimation errors, the sample tangency portfolio involves more risk than the true tangency portfolio, making it less attractive. Note that θ^2 is unobservable to investors, so the optimal two-fund rule with weights $c^*\hat{\Sigma}_t^{-1}\hat{\mu}_t/\gamma$ is not attainable. Kan and Zhou (2007) suggest an implementable version of the optimal two-fund rule

$$\hat{w}_t^{\mathrm{II}} = \frac{\hat{c}_t}{\gamma}\hat{\Sigma}_t^{-1}\hat{\mu}_t, \tag{12}$$

where

$$\hat{c}_t = \frac{k_3\hat{\theta}_{a,t}^2}{\hat{\theta}_{a,t}^2 + \frac{N}{h}}, \tag{13}$$

and

$$\hat{\theta}_{a,t}^2 = \frac{(h-N-2)\hat{\theta}_t^2 - N}{h} + \frac{2(\hat{\theta}_t^2)^{\frac{N}{2}}(1+\hat{\theta}_t^2)^{-\frac{h-2}{2}}}{h\mathrm{B}_{\hat{\theta}_t^2/(1+\hat{\theta}_t^2)}(N/2,(h-N)/2)}, \tag{14}$$

with $\hat{\theta}_t^2 = \hat{\mu}_t'\hat{\Sigma}_t^{-1}\hat{\mu}_t$ being the maximum likelihood estimator of θ^2 at time t, and

$$\mathrm{B}_x(a,b) = \int_0^x y^{a-1}(1-y)^{b-1}\mathrm{d}y \tag{15}$$

being the incomplete beta function. Given that $\hat{\theta}_{a,t}^2$ is a function of $\hat{\theta}_t^2$, we can define a function of $\hat{\theta}_t^2$,

$$g_1(\hat{\theta}_t^2) = \frac{\hat{\theta}_{a,t}^2}{\hat{\theta}_{a,t}^2 + \frac{N}{h}}, \tag{16}$$

and then write the rule as

$$\hat{w}_t^{\mathrm{II}} = \frac{k_3 g_1(\hat{\theta}_t^2)}{\gamma}\hat{\Sigma}_t^{-1}\hat{\mu}_t. \tag{17}$$

The second portfolio rule that Kan and Zhou (2007) consider is a three-fund rule. When the parameters are known, a mean-variance investor should only be interested in investing in the risk-free asset and the tangency portfolio. However, when the parameters are unknown, there are good reasons to move away from the sample tangency portfolio. The three-fun rule that Kan and Zhou (2007) consider takes the form of $(c\hat{\Sigma}_t^{-1}\hat{\mu}_t + d\hat{\Sigma}_t^{-1}1_N)/\gamma$, i.e., a combination of the risk-free asset, the sample tangency portfolio, and the sample global minimum-variance portfolio. The sample global minimum-variance portfolio is attractive because it does not require the estimation of μ and may have less estimation risk than the sample tangency portfolio. Kan and Zhou (2007) show the optimal choice of c and d for the three-fund rule are given by

$$c^{**} = \frac{k_3\psi^2}{\psi^2 + \frac{N}{h}}, \tag{18}$$

$$d^{**} = \frac{k_3(N/h)}{\psi^2 + \frac{N}{h}}\mu_g, \tag{19}$$

where $\mu_g = 1_N'\Sigma^{-1}\mu/(1_N'\Sigma^{-1}1_N)$ is the expected excess return of the global minimum-variance portfolio and $\psi^2 = \mu'\Sigma^{-1}\mu - (1_N'\Sigma^{-1}\mu)^2/(1_N'\Sigma^{-1}1_N)$ is the squared slope of the asymptote to the *ex ante* minimum-variance frontier, which is a measure of the cross-sectional difference of expected returns across the N assets. The optimal three-fund rule requires the knowledge of μ_g and ψ^2, but they are unobservable to investors. As a result, Kan and Zhou (2007) suggest the following implementable version of the optimal three-fund rule:

$$\hat{w}_t^{\text{III}} = \frac{\tilde{c}_t}{\gamma}\hat{\Sigma}_t^{-1}\hat{\mu}_t + \frac{\tilde{d}_t}{\gamma}\hat{\Sigma}_t^{-1}1_N, \tag{20}$$

where

$$\tilde{c}_t = \frac{k_3\hat{\psi}_{a,t}^2}{\hat{\psi}_{a,t}^2 + \frac{N}{h}}, \tag{21}$$

$$\tilde{d}_t = \frac{k_3(N/h)}{\hat{\psi}_{a,t}^2 + \frac{N}{h}}\hat{\mu}_{g,t}, \tag{22}$$

with $\hat{\mu}_{g,t} = 1_N'\hat{\Sigma}_t^{-1}\hat{\mu}_t/(1_N'\hat{\Sigma}_t^{-1}1_N)$ being the maximum likelihood estimator of μ_g and

$$\hat{\psi}_{a,t}^2 = \frac{(h-N-1)\hat{\psi}_t^2 - (N-1)}{h} + \frac{2(\hat{\psi}_t^2)^{\frac{N-1}{2}}(1+\hat{\psi}_t^2)^{-\frac{h-2}{2}}}{h\mathrm{B}_{\hat{\psi}_t^2/(1+\hat{\psi}_t^2)}((N-1)/2,(h-N+1)/2)}, \tag{23}$$

with $\hat{\psi}_t^2 = \hat{\mu}_t'\hat{\Sigma}_t^{-1}\hat{\mu}_t - (1_N'\hat{\Sigma}_t^{-1}\hat{\mu}_t)^2/(1_N'\hat{\Sigma}_t^{-1}1_N)$ being the maximum likelihood estimator of ψ^2. Given that $\hat{\psi}_{a,t}^2$ is a function of $\hat{\psi}_t^2$, we can define a function of $\hat{\psi}_t^2$,

$$g_2(\hat{\psi}_t^2) = \frac{\hat{\psi}_{a,t}^2}{\hat{\psi}_{a,t}^2 + \frac{N}{h}}, \tag{24}$$

and write

$$\hat{w}_t^{\mathrm{III}} = \frac{k_3}{\gamma}\left[g_2(\hat{\psi}_t^2)\hat{\Sigma}_t^{-1}(\hat{\mu}_t - 1_N\hat{\mu}_{g,t}) + \hat{\mu}_{g,t}\hat{\Sigma}_t^{-1}1_N\right]. \tag{25}$$

In order to evaluate the out-of-sample performance of the implementable version of the optimal two-fund and three-fund rules, we need to compute their expected out-of-sample utility, i.e., $E[U(\hat{w}_t^{\mathrm{II}})]$ and $E[U(\hat{w}_t^{\mathrm{III}})]$. Kan and Zhou (2007) rely on simulations to approximate the expected out-of-sample utilities of these two rules. Besides being time consuming, simulation does not tell us the parameters that determine the out-of-sample performance. To overcome these problems, we present below the exact unconditional distributions of the out-of-sample returns of these two rules as well as the explicit expressions of their out-of-sample utilities in terms of a one-dimensional integral.

To facilitate our presentation, we use $\mathscr{G}_{m,n}^{\delta}$ to stand for a random variable $y = x_1/x_2$ where $x_1 \sim \chi_m^2(\delta)$ and $x_2 \sim \chi_n^2$, independent of each other.[2]

Proposition 1: *Suppose $N > 1$. Let $b \sim \mathrm{Beta}((h-N+1)/2, (N-1)/2)$ and $u_1 \sim \chi_{h-N}^2$, independent of each other. Conditional on b, let $z_0 \sim \mathscr{N}(\sqrt{h}\theta\sqrt{b}, 1)$ and $\tilde{u} \sim \chi_{N-1}^2(h\theta^2(1-b))$, and they are independent of each other and u_1.*[3] *Then, the distribution of $\hat{\theta}_t^2$ is given by*

$$\hat{\theta}_t^2 = \frac{z_0^2 + \tilde{u}}{u_1}, \tag{26}$$

and the joint distribution of the conditional mean and variance of the out-of-sample return of the two-fund rule can be obtained using

$$\mu_{\mathrm{II},t} \equiv \hat{w}_t^{\mathrm{II}\prime}\mu = \frac{k_3 g_1(\hat{\theta}_t^2)\sqrt{h}\theta z_0}{\gamma u_1\sqrt{b}}, \tag{27}$$

[2]Note that $\mathscr{G}_{m,n}^{\delta} = (m/n)\mathscr{F}_{m,n}^{\delta}$, where $\mathscr{F}_{m,n}^{\delta}$ is a noncentral F-distribution with m and n degrees of freedom, and a noncentrality parameter of δ. The density function of $\mathscr{F}_{m,n}^{\delta}$ can be computed using the Matlab function `ncfpdf`.

[3]If we set $b = 1$ and $\tilde{u} = 0$ for $N = 1$, then the results in Proposition 1 also hold for the case of $N = 1$.

$$\sigma^2_{\mathrm{II},t} \equiv \hat{w}_t^{\mathrm{II}\prime}\Sigma\hat{w}_t^{\mathrm{II}} = \frac{k_3^2 g_1^2(\hat{\theta}_t^2)h\hat{\theta}_t^2}{\gamma^2 u_1 b}. \tag{28}$$

The expected out-of-sample utility of the two-fund rule is given by

$$E[U(\hat{w}_t^{\mathrm{II}})] = \frac{k_3 h\theta^2}{\gamma(h-N-2)}E[g_1(q_1)] - \frac{k_3(h-N-4)}{2\gamma(h-N-2)}E[g_1(q_2)^2 q_2] \tag{29}$$

when $h > N+4$, *where* $g_1(\cdot)$ *is the function defined in (16) and* $q_1 \sim \mathscr{G}^{h\theta^2}_{N+2,h-N-2}$, $q_2 \sim \mathscr{G}^{h\theta^2}_{N,h-N-2}$.

Proposition 2: *Suppose* $N > 3$. *Let* $z_1 \sim \mathscr{N}(\sqrt{h}\theta_g, 1)$, $z_2 \sim \mathscr{N}(\sqrt{h}\psi, 1)$, $u_0 \sim \chi^2_{N-2}$, $v_1 \sim \chi^2_{h-N}$, $v_2 \sim \chi^2_{h-N+1}$, $w_1 \sim \chi^2_{h-N+3}$, $w_2 \sim \chi^2_{h-N+2}$, $s_1 \sim \chi^2_{N-4}$, $s_2 \sim \chi^2_{N-3}$, $x_{11} \sim \mathscr{N}(0,1)$, $x_{21} \sim \mathscr{N}(0,1)$, $a \sim \mathscr{N}(0,1)$, $b \sim \mathscr{N}(0,1)$, $c \sim \mathscr{N}(0,1)$, *and they are independent of each other,*[4] *where* θ_g *is the Sharpe ratio of the global minimum-variance portfolio. Then, the distribution of* $\hat{\psi}_t^2$ *is given by*

$$\hat{\psi}_t^2 = \frac{z_2^2 + u_0}{v_2}. \tag{30}$$

Define

$$y_1 = \frac{x_{11}}{\sqrt{w_1}} + \frac{bx_{21}}{\sqrt{w_1 w_2}} + \frac{ax_{21}}{\sqrt{v_2 w_2}}, \tag{31}$$

$$y_2 = \frac{c}{\sqrt{w_1}} + \frac{b\sqrt{s_2}}{\sqrt{w_1 w_2}} + \frac{a\sqrt{s_2}}{\sqrt{v_2 w_2}}, \tag{32}$$

$$y_3 = \frac{a\hat{\psi}_t + z_1}{v_1}. \tag{33}$$

The joint distribution of the conditional mean and variance of the out-of-sample return of the three-fund rule can be obtained from

$$\mu_{\mathrm{III},t} \equiv \hat{w}_t^{\mathrm{III}\prime}\mu = \frac{\sqrt{h}k_3}{\gamma}\left[\frac{\psi g_2(\hat{\psi}_t^2)}{v_2}\left(\frac{x_{21}\sqrt{u_0}}{\sqrt{w_2}} + z_2\right) + \theta_g y_3 + \frac{\psi y_3}{\hat{\psi}_t}\left(\frac{\sqrt{u_0}y_1}{\sqrt{v_2}} + \frac{az_2}{v_2}\right)\right], \tag{34}$$

$$\sigma^2_{\mathrm{III},t} \equiv \hat{w}_t^{\mathrm{III}\prime}\Sigma\hat{w}_t^{\mathrm{III}} = \frac{hk_3^2}{\gamma^2}\left[\left(1 + \frac{s_1}{w_1}\right)y_3^2 + \left(\frac{ay_3 + g_2(\hat{\psi}_t^2)\hat{\psi}_t}{\sqrt{v_2}}\right)^2 + \left(y_1 y_3 + \frac{x_{21}g_2(\hat{\psi}_t^2)\hat{\psi}_t}{\sqrt{v_2 w_2}}\right)^2 + \left(y_2 y_3 + \frac{\sqrt{s_2}g_2(\hat{\psi}_t^2)\hat{\psi}_t}{\sqrt{v_2 w_2}}\right)^2\right]. \tag{35}$$

[4] If we set $u_0 = 0$, $x_{11} = 0$, $x_{21} = 0$, $s_1 = 0$, $s_2 = 0$, and $c = 0$ when $N = 2$, and set $s_1 = 0$, $s_2 = 0$, and $c = 0$ when $N = 3$, then the results in Proposition 2 also hold for the cases of $N = 2$ or $N = 3$.

The expected out-of-sample utility of the three-fund rule is given by

$$E[U(\hat{w}_t^{\rm III})] = \frac{k_3}{(h-N-2)\gamma}\left[\frac{h\theta_g^2}{2} + \frac{h\psi^2}{h-N-1} - \frac{h-4+h\psi^2}{2(h-N-3)}\right] + \frac{k_3 h\psi^2}{(h-N-1)\gamma}E[g_2(q_3)] - \frac{k_3(h-N-4)}{2(h-N)\gamma}E\left[\left(\frac{2g_2(q_4)}{h-N-2} + g_2(q_4)^2\right)q_4\right] \quad (36)$$

when $h > N+4$, where $g_2(\cdot)$ is the function defined in (24) and $q_3 \sim \mathcal{G}^{h\psi^2}_{N+1,h-N-1}$, $q_4 \sim \mathcal{G}^{h\psi^2}_{N-1,h-N-1}$.

Propositions 1 and 2 show that the conditional mean and variance of the two-fund and three-fund rules can be obtained by simulating a small number of random variables. Once we obtain the conditional mean and variance of the two-fund rule, we can simulate its out-of-sample return at time $t+1$ using $\hat{\mu}_{{\rm II},t} + \hat{\sigma}_{{\rm II},t}y$, where $y \sim \mathcal{N}(0,1)$ and it is independent of $\hat{\mu}_{{\rm II},t}$ and $\hat{\sigma}_{{\rm II},t}$. The out-of-sample return of the three-fund rule can also be simulated in a similar fashion. In addition, Propositions 1 and 2 show that besides N, h, and γ, the unconditional distribution of the out-of-sample return and the out-of-sample utility of the two-fund rule only depend on θ^2, whereas those of the three-fund rule only depend on θ_g^2 and θ^2 (because $\psi^2 = \theta^2 - \theta_g^2$). Thus, there is no need to specify μ and Σ for computing the return distributions and the utilities for these two rules. Moreover, the two Propositions express the out-of-sample portfolio returns in terms of a set of univariate random variables and the expected out-of-sample utilities in terms of a one-dimensional integral, providing a fast way of simulating the return distributions and computing the utilities of the two-fund and three-fund rules.

2.3 Combining the $1/N$ Rule with a Risk-free Asset

The $1/N$ rule refers to the portfolio strategy with equal weights in the N risky assets. However, when a risk-free asset is also available, how to allocate weights between the risk-free asset and the equally weighted portfolio of N risky assets is not entirely clear. One approach is to optimally allocate the weights between the risk-free asset and the equally weighted portfolio of risky assets, and DGU adopted this approach in their analytical section (i.e., Section 4).[5] Note that this is basically a portfolio choice problem with a risk-free asset and a single risky portfolio (i.e., the

[5]In their analysis based on empirical data (Section 3) and simulated data (Section 5), optimal portfolios are normalized to be fully invested in risky assets. The performance of these normalized optimal portfolios is compared to that of the portfolio with equal weights in the N risky assets.

equally weighted portfolio of risky assets). Denote the mean and variance of the excess return of the equally weighted portfolio of risky assets as $\mu_{ew} = 1_N'\mu/N$ and $\sigma_{ew}^2 = 1_N'\Sigma 1_N/N^2$. It is straightforward to show that the optimal weight allocated to the equally weighted portfolio of risky assets is $w_{ew}^* = \mu_{ew}/(\gamma\sigma_{ew}^2)$ and $1 - w_{ew}^*$ is invested in the risk-free asset. The resulting utility of this optimal choice is $\theta_{ew}^2/(2\gamma)$ where $\theta_{ew} = \mu_{ew}/\sigma_{ew}$ is the Sharpe ratio of the equally weighted portfolio of risky assets.

However, μ_{ew} and σ_{ew}^2 are unknown to investors in reality, so w_{ew}^* is also unknown.[6] An implementable version is to replace μ_{ew} and σ_{ew}^2 with their sample counterparts, $\hat{\mu}_{ew,t} = 1_N'\hat{\mu}_t/N$ and $\hat{\sigma}_{ew,t}^2 = 1_N'\hat{\Sigma}_t 1_N/N^2$, to obtain

$$\hat{w}_{ew,t} = \frac{1}{\gamma}\frac{\hat{\mu}_{ew,t}}{\hat{\sigma}_{ew,t}^2}, \tag{37}$$

which is basically the ML rule for the case with a risk-free asset and one risky asset. Using (8), we obtain the expected out-of-sample utility of this portfolio rule as

$$E[U(\hat{w}_{ew,t})] = \frac{h[(h-10)\theta_{ew}^2 - 1]}{2\gamma(h-3)(h-5)}. \tag{38}$$

Another approach is to combine the risk-free asset and the equally weighted portfolio of risky assets using the implementable optimal two-fund rule of Kan and Zhou (2007). Using (12) and (13), we have

$$\hat{w}_{ew,t}^{\mathrm{II}} = \frac{1}{\gamma}\left(\frac{h-5}{h}\right)\left(\frac{\hat{\theta}_{ew,a,t}^2}{\hat{\theta}_{ew,a,t}^2 + \frac{1}{h}}\right)\frac{\hat{\mu}_{ew,t}}{\hat{\sigma}_{ew,t}^2}, \tag{39}$$

where

$$\hat{\theta}_{ew,a,t}^2 = \frac{(h-3)\hat{\theta}_{ew,t}^2 - 1}{h} + \frac{2\hat{\theta}_{ew,t}(1+\hat{\theta}_{ew,t}^2)^{-\frac{h-2}{2}}}{hB_{\hat{\theta}_{ew,t}^2/(1+\hat{\theta}_{ew,t}^2)}\left(\frac{1}{2}, \frac{h-1}{2}\right)} \tag{40}$$

with $\hat{\theta}_{ew,t}^2 = \hat{\mu}_{ew,t}^2/\hat{\sigma}_{ew,t}^2$. Using the results of Proposition 1, the expected out-of-sample utility of portfolio $\hat{w}_{ew,t}^{\mathrm{II}}$ is given by

$$E[U(\hat{w}_{ew,t}^{\mathrm{II}})] = \frac{(h-5)\theta_{ew}^2}{\gamma(h-3)}E[\tilde{g}_1(\tilde{q}_1)] - \frac{(h-5)^2}{2\gamma h(h-3)}E[\tilde{g}_1(\tilde{q}_2)^2\tilde{q}_2] \tag{41}$$

[6]In their analysis, DGU assume that investors are able to hold w_{ew}^*, and compare the performance of w_{ew}^* with that of the sample-based mean-variance portfolio to obtain the number of estimation months required for the sample-based mean-variance portfolio to outperform the $1/N$ rule.

for $h > 5$, with $\tilde{q}_1 \sim \mathscr{G}_{3,h-3}^{h\theta_{ew}^2}$, $\tilde{q}_2 \sim \mathscr{G}_{1,h-3}^{h\theta_{ew}^2}$, and

$$\tilde{g}_1(\hat{\theta}_{ew,t}^2) = \frac{\hat{\theta}_{ew,a,t}^2}{\hat{\theta}_{ew,a,t}^2 + \frac{1}{h}}. \tag{42}$$

2.4 Comparison of Portfolio Rules: Analytical Results

The results from the previous subsection make it easy to compute the expected out-of-sample utility of the ML rule, the two-fund rule, the three-fund rule, and the $1/N$ rule. In this subsection, we first compare the expected out-of-sample utilities of these portfolio rules based on parameters calibrated to empirical data. We then examine the condition under which the $1/N$ rule outperforms various optimal portfolio rules.

In Figure 1, we plot the expected out-of-sample utilities of portfolios $\hat{w}_t$, $\hat{w}_t^{\mathrm{II}}$, $\hat{w}_t^{\mathrm{III}}$, $\hat{w}_{ew,t}$, and $\hat{w}_{ew,t}^{\mathrm{II}}$ as a function of the length of estimation window (i.e., h), with parameters calibrated using monthly excess returns of the 10 momentum portfolios from January 1927 to December 2014. Sample estimates based on this set of 10 risky portfolios give $\theta = 0.268$, $\psi = 0.176$, and $\theta_{ew} = 0.107$. To gauge the effect of estimation errors, we also include in Figure 1 the utilities of portolios w^* and w_{ew}^*. Note that the relative rankings of various portfolio rules do not depend on the value of γ, and we assume $\gamma = 3$ in the figure.

First, comparing the performance of the three versions of the $1/N$ rule (i.e., $\hat{w}_{ew,t}$, $\hat{w}_{ew,t}^{\mathrm{II}}$, and w_{ew}^*), we find that the performance of the two implementable versions (i.e., $\hat{w}_{ew,t}$ and $\hat{w}_{ew,t}^{\mathrm{II}}$) are almost identical except that $\hat{w}_{ew,t}^{\mathrm{II}}$ outperforms $\hat{w}_{ew,t}$ slightly when the estimation window is short (e.g., $h = 60$). This is because given only one risky asset, the estimation errors involved in $\hat{w}_{ew,t}$ are relatively small, and the two-fund rule (i.e., $\hat{w}_{ew,t}^{\mathrm{II}}$) does not generate significant improvement over the ML rule (i.e., $\hat{w}_{ew,t}$). In addition, we observe noticeable difference between the performance of the non-implementable version (i.e., w_{ew}^*) and that of the two implementable versions, especially for relatively short estimation window. This suggests that assuming investors know the true values of μ_{ew} and σ_{ew}^2 overstates the performance of the $1/N$ rule. Hence we focus below on $\hat{w}_{ew,t}^{\mathrm{II}}$ to understand the performance of the $1/N$ rule.

Next, comparing the performance of portfolios $\hat{w}_t$ (ML rule), $\hat{w}_t^{\mathrm{II}}$ (two-fund rule), and $\hat{w}_t^{\mathrm{III}}$ (three-fund rule) with that of the true optimal portfolio w^*, we can see that portfolio $\hat{w}_t$ performs poorly especially for short estimation window. This is due to the significant estimation errors involved in the ML rule. By taking into account of the estimation errors, we see that the two-fund and three-fund rules both outperform the ML rule, and the performance improvement is significant when h is small.

Finally, comparing the performance of the three optimal portfolio rules (i.e., $\hat{w}_t$, $\hat{w}_t^{\text{II}}$, and $\hat{w}_t^{\text{III}}$) with that of the $1/N$ rule (i.e., $\hat{w}_{ew,t}^{\text{II}}$), we can see that portfolio $\hat{w}_t$ needs an estimation window of at least $h = 198$ months to outperform the $1/N$ rule due to the significant estimation errors involved in this portfolio rule. Portfolios $\hat{w}_t^{\text{II}}$ and $\hat{w}_t^{\text{III}}$, taking into account the effect of estimation errors, both outperform the $1/N$ rule even with an estimation window as short as $h = 60$ months. These findings are inconsistent with the conclusion drawn in DGU that thousands of estimation months are needed for the sample-based mean-variance strategy and its extensions to outperform the $1/N$ benchmark.

Figure 2 presents similar plots with parameters calibrated to the monthly excess returns of the Fama-French 25 size and book-to-market ranked portfolios over January 1927 to December 2014. This choice gives $\theta = 0.301$, $\psi = 0.258$, and $\theta_{ew} = 0.128$. The pattern in Figure 2 is similar to that in Figure 1. However, given a larger number of risky assets $N = 25$, more estimation errors are involved in portfolios $\hat{w}_t$, $\hat{w}_t^{\text{II}}$, and $\hat{w}_t^{\text{III}}$. As a result, a longer estimation window is required for these portfolios to outperform the $1/N$ rule. The minimum length of estimation window for portfolios $\hat{w}_t$, $\hat{w}_t^{\text{II}}$, and $\hat{w}_t^{\text{III}}$ to outperform the $1/N$ rule are $h = 432$ months, $h = 94$ months, and $h = 93$ months, respectively.

To understand why DGU's conclusion is so different from the results in Figures 1 and 2, we next examine the required length of estimation window for portfolios $\hat{w}_t$, $\hat{w}_t^{\text{II}}$, and $\hat{w}_t^{\text{III}}$ to outperform the $1/N$ rule for different values of θ and θ_{ew}. DGU conducted similar analysis for portfolio $\hat{w}_t$ but did not provide an analysis for the two-fund and three-fund rules.[7] It turns out that the results of the two-fund and three-fund rules are very different from that of the ML rule.

We consider two levels of θ ($\theta = 0.4$, 0.2) and three levels of θ_{ew}/θ ($\theta_{ew}/\theta = 0.25$, 0.50, 0.75), similar to the choice in DGU. In addition, we set the value of $\theta_g = \theta/2$, which leads to $\psi = 0.346$ when $\theta = 0.4$ and $\psi = 0.173$ when $\theta = 0.2$. This choice of θ_g/θ is close to the sample estimate

[7]One minor difference between DGU's analysis and ours is that DGU uses portfolio w_{ew}^* to represent the performance of the $1/N$ rule, and we use portfolio $\hat{w}_{ew,t}^{\text{II}}$ instead.

in Figure 2 (0.515) but is lower than that in Figure 1 (0.754). Note that given the same level of θ, changing the value of θ_g only affects the performance of portfolio $\hat{w}_t^{\mathrm{III}}$. A higher θ_g leads to a better performance of the three-fund rule. Therefore, our choice of θ_g tends to underestimate the performance of the three-fund rule relative to the sample estimates.

Figure 3 plots the number of estimation months required as a function of the number of risky assets N, and the six panels are for different combinations of θ and θ_{ew}/θ.[8] First, it is obvious from Figure 3 that the required length of estimation window of the two-fund and three-fund rules are significantly shorter than that of the ML rule. For example, in panel (a) ($\theta = 0.4$ and $\theta_{ew} = 0.1$), the required length of estimation window when $N = 100$ is 1055 months for the ML rule, and that of the two-fund and three-fund rules are 162 and 153 months, respectively. This is because the ML rule contains significant estimation errors, and the two-fund and three-fund rules mitigate the effect of estimation risk.

Second, Figure 3 shows that all else equal, the higher θ_{ew}/θ, the longer estimation window is required. For example, given the level of $\theta = 0.4$, the required length of the estimation window when $N = 100$ are 1037 and 908 months for the two-fund and three-fund rules, respectively, when $\theta_{ew}/\theta = 0.75$ (i.e., panel (e)), but when $\theta_{ew}/\theta = 0.25$ (i.e., panel (a)), the numbers are only 162 and 153 months. Similar pattern can also be observed when $\theta = 0.2$ (i.e., panels (b), (d), (f)). These results are intuitive. A high value of θ_{ew} relative to θ indicates that the $1/N$ portfolio is close to the true optimal portfolio, and therefore, it is difficult for the optimal portfolio rules to beat the $1/N$ rule.

Third, we can see from Figure 3 that a longer estimation window is required when θ is lower, ceteris paribus. For example, given $\theta_{ew}/\theta = 0.25$, the required estimation window for $\hat{w}_t^{\mathrm{II}}$ and $\hat{w}_t^{\mathrm{III}}$ to outperform the $1/N$ rule when $N = 100$ are 343 and 281 months when $\theta = 0.2$ (i.e., panel (b)), and the numbers are 162 and 153 months when $\theta = 0.4$ (i.e., panel (a)). These results are due to the fact that when θ is low, the benefit of optimization is small relative to the cost of estimation errors. Therefore, a longer estimation window is needed for the optimal portfolio rules to outperform the $1/N$ rule.

In summary, Figure 3 suggests that the required length of estimation window for the optimal

[8] As the expected out-of-sample utility of all portfolio rules are proportional to $1/\gamma$, the required length of estimation window for the two-fund and three-fund rules to outperform the $1/N$ rule is independent of the choice of γ.

portfolio rules to outperform the $1/N$ rule crucially depends on the values of θ and θ_{ew}.[9] A low θ together with a high θ_{ew}/θ will lead to a long required estimation window. This insight helps us to understand the surprising findings in DGU. Their conclusion from the analytical section that 3000 (6000) months are required for the ML rule to outperform the $1/N$ rule when $N = 25$ ($N = 50$) is based on the assumption that $\theta = 0.15$ and $\theta_{ew} = 0.12$. In their simulation study, they set $\theta = 0.1477$ and $\theta_{ew} = 0.1356$, 0.1447, 0.1466 for $N = 10$, 25, 50 respectively. Given these parameter values, it is not surprising to find that extremely long estimation windows are required for the optimal portfolio rules to outperform the $1/N$ rule. However, such condition (i.e., low θ and high θ_{ew}/θ) does not always hold in real data. It is interesting that DGU also identify outperformance of the $1/N$ rule in some empirical data sets where the condition does not seem to hold. To further understand the issue, we examine portfolio performance using similar empirical data in the next subsection.

2.5 Comparison of Portfolio Rules: Empirical Results

In this subsection, we compare empirically the performance of various portfolio rules across six empirical datasets which are also used in DGU. The six datasets are: i) monthly excess returns of the standard 10 industry portfolios: Consumer-Discretionary, Consumer-Staples, Manufacturing, Energy, High-Tech, Telecommunication, Wholesale and Retail, Health, Utilities, and Others, plus the US equity market portfolio; ii) monthly excess returns on eight international equity indices: Canada, France, Germany, Italy, Japan, Switzerland, the UK, and the US, as well as the MSCI (Morgan Stanley Capital International) World index; iii) monthly excess returns of the Fama-French three factors, MKT, SMB and HML; iv) monthly excess returns of the Fama-French 20 portfolios sorted by size and book-to-market plus the market factor; v) monthly excess returns of the Fama-French 20 portfolios sorted by size and book-to-market plus the Fama-French three factors; and vi) monthly returns of the Fama-French 20 portfolios sorted by size and book-to-market plus the Fama-French three factors and the momentum factor. The 20 size and book-to-market ranked portfolios are the Fama-French 5×5 size and book-to-market ranked portfolios after removing the five portfolios with the largest size. Except for the international data which is obtained

[9] In addition, for a given set of N base assets, we can construct a new set of N assets based on N linear combinations of the original base assets. While the out-of-sample performance of the optimization based portfolio rules are invariant to repackaging of the base assets, the $1/N$ rule is sensitive to such repacking, and its out-of-sample performance critically depends on the choice of test assets.

from MSCI and is only available from January 1970 to December 2014, all other datasets are obtained from Kenneth French's Web site and cover the period from January 1927 to December 2014. The only dataset that DGU used but we do not is the one consisting of ten sector portfolios of the S&P 500 as this dataset is not available to us.

Following DGU and Tu and Zhou (2011), we use a rolling estimation approach with an estimation window of h months. Specifically, for each month t, we use the data in the most recent h months up to month t to compute the weights of various portfolio rules, and obtain the associated out-of-sample portfolio excess returns in month $t+1$. This practice generates $T-h$ out-of-sample portfolio excess returns where T stands for the length of the sample period for the data. Based on these $T-h$ out-of-sample portfolio excess returns, we obtain the sample mean ($\hat{\mu}$) and the sample variance ($\hat{\sigma}^2$) for a portfolio rule, and compute its certainty-equivalent return ($CEQ = \hat{\mu} - \frac{\gamma}{2}\hat{\sigma}^2$) as well as its Sharpe ratio ($S.R. = \hat{\mu}/\hat{\sigma}$).

We notice that in their empirical section, DGU adopt a different approach to compare the performance of the optimal portfolio rules with that of the $1/N$ rule. Instead of optimally allocating weights between the risk-free asset and the equally weighted portfolio with only risky assets to construct the $1/N$ portfolio (such as in their analytical section), they choose to normalize those optimal portfolios such that the weights on the risky assets sum up to one. Specifically, the weights of the ML rule and the three-fund rule in DGU are given by[10]

$$\hat{w}_t^{DGU} = \frac{\hat{w}_t}{|1_N'\hat{w}_t|} = \frac{\hat{\Sigma}_t^{-1}\hat{\mu}_t}{\left|1_N'\hat{\Sigma}_t^{-1}\hat{\mu}_t\right|}, \tag{43}$$

$$\hat{w}_t^{DGU,\mathrm{III}} = \frac{\hat{w}_t^{\mathrm{III}}}{|1_N'\hat{w}_t^{\mathrm{III}}|} = \frac{\tilde{c}_t\hat{\Sigma}_t^{-1}\hat{\mu}_t + \tilde{d}_t\hat{\Sigma}_t^{-1}1_N}{\left|\tilde{c}_t 1_N'\hat{\Sigma}_t^{-1}\hat{\mu}_t + \tilde{d}_t 1_N'\hat{\Sigma}_t^{-1}1_N\right|}. \tag{44}$$

In Table 1, we use data in the same sample period as in the empirical section of DGU (i.e., 1963/7–2004/11), and construct portfolios following the normalization approach of DGU based on the same parameter values (i.e., $h = 120$ and $\gamma = 1$). Panel A presents the CEQ results, and Panel B presents the Sharpe ratio results. It should be emphasized that the rankings of the portfolio rules based on the CEQ results are not necessarily the same as the rankings based on the Sharpe ratio results. For performance comparison, the CEQ results are more relevant because it explicitly

[10]Note that with the normalization, the two-fund rule is identical to the ML rule, and as a result DGU did not present the results for the two-fund rule.

takes into account of the risk aversion coefficient of the investor, whereas the results of Sharpe ratio are independent of γ. The first row, "MV (in-sample)," is the performance of the normalized mean-variance optimal portfolio based on the in-sample estimates of mean and covariance matrix of excess returns, which is basically the *ex post* tangency portfolio. The second row, "$1/N$," is the performance of the equally weighted portfolio with only risky assets. The rest of the table reports the performance of the normalized ML and three-fund rules together with the one-sided p-values of the performance difference between these rules and the $1/N$ portfolio. Consistent with the findings in DGU, Table 1 shows dominance of the $1/N$ rule. Across all six datasets and based on either CEQ or Sharpe ratio, the $1/N$ rule outperforms the ML rule. The three-fund rule ("KZ3"), taking into account the effect of estimation risk, performs better than the ML rule, but still underperforms the $1/N$ rule in general. The only exceptions are the CEQ of "FF+3-factor" and the Sharpe ratios of "MKT/SMB/HML" and "FF+3-factor." It is important to note that while the *ex post* tangency portfolio maximizes the in-sample Sharpe ratio, it is in general not the portfolio that maximizes the CEQ. This is evident for the cases of "FF+3-factor" and "FF+4-factor," where we find the CEQ of MV (in-sample) to be negative.

In Tables 2 and 3, we repeat similar analysis using data in the extended sample period (i.e., 1927/1–2014/12) and for different combinations of h (120 or 240) and γ (1 or 3). Table 2 presents the CEQ results, and Table 3 reports the Sharpe ratio results. We continue to observe the dominance of the $1/N$ rule. When $h = 120$ months, the $1/N$ rule outperforms both the ML rule and the three-fund rule in all cases except for the three-fund rule in the dataset of "FF 3-factor." When $h = 240$ months, the longer estimation window reduces the estimation risk and increases the performance of the two optimal portfolio rules in most cases, but the optimal portfolio rules continue to underperform the $1/N$ rule in general. These results suggest that the finding in DGU is not due to specific sample period or parameter values.

However, the results in Tables 1-3 are inconsistent with the insight obtained from our analysis in Subsection 2.4. In particular, the poor performance of the three-fund rule in the last three datasets (i.e., FF+1-factor, FF+3-factor, and FF+4-factor) is surprising. For these datasets, we have a relatively high Sharpe ratio of the tangency portfolio, θ, (0.3379–0.4433 in the extended sample period) and a relatively low ratio of the Sharpe ratio of the equally weighted portfolio relative to that of the tangency portfolio, θ_{ew}/θ, (39%–44% in the extended sample period). Analytical results from Subsection 2.4 suggest that in these cases, the $1/N$ rule is less likely to outperform the three-fund rule.

We find that the inconsistency between the empirical results in Tables 1-3 and the analytical results in Subsection 2.4 is mostly due to the normalization on the optimal portfolios performed by DGU. This process basically excludes the risk-free asset from those optimal portfolios, but those optimal portfolio rules were derived under the assumption that the risk-free asset is available. Therefore, after the normalization, those rules are no longer "optimal." More importantly, it can be shown that the out-of-sample returns of those normalized optimal portfolios do not have integral moments, which contributes to their poor out-of-sample performance. To understand why the moments do not exist for those normalized portfolios, notice that the denominator of $\hat{w}_t^{DGU}$ and $\hat{w}_t^{DGU,\mathrm{III}}$ in (43) and (44) have non-negligible density at zero, and a zero denominator will lead to extreme positions in the risky assets, which results in very fat tails for the distribution of the out-of-sample returns of the normalized portfolios. Under the normality assumption on r_t, Okhrin and Schmid (2006) prove that the expectations of the weights of the sample tangency portfolio do not exist, so it is not surprising that its out-of-sample returns have no finite moments.[11] Therefore, to compare the performance of the normalized optimal portfolios with that of the $1/N$ rule does not seem to be a proper practice. Kirby and Ostdeik (2012) also point out that the use of normalized optimal portfolios is the main reason for the poor performance of the optimal portfolio rules in DGU.

In Tables 4 and 5, we report the performance of the portfolios without normalization using the same data as in Tables 2 and 3. Without normalization, the performance of the two-fund rule is different from that of the ML rule, so we report results for the ML rule, the two-fund rule (i.e., "KZ2"), and the three-fund rule (i.e., "KZ3"). In addition, the MV (in-sample) in Tables 4 and 5 report the performance of the *ex post* mean-variance optimal portfolio, which is not normalized. As a result, the MV (in-sample) portfolio in Tables 4-5 is not the same as the *ex post* tangency portfolio as in Tables 1–3. For the $1/N$ rule given risk-free asset, the weights are based on $\hat{w}_{ew,t}^{\mathrm{II}}$ in (39). The p-values are for the performance differences between the optimal portfolio rules and the $1/N$ rule.

[11]Note that the non-existence of integral moments of the out-of-sample return of a normalized portfolio does not depend on the normality assumption. Under a general continuous distribution of r_t, one can invoke a lemma due to Sargan (1976) to show the non-existence of the moments.

We can see from Tables 4 and 5 that the performance of the optimal portfolio rules improves in general without the normalization. In addition, the two-fund and three-fund rules outperform the $1/N$ rule in terms of both CEQ and Sharpe ratio for all the combinations of the parameter values in the three datasets with high θ and low θ_{ew}/θ (i.e., "FF+1-factor", "FF+3-factor", and "FF+4-factor"), consistent with the findings in Subsection 2.4. For the other three datasets, given a lower θ and a higher θ_{ew}/θ, it is more difficult for the two-fund and three-fund rules to outperform the $1/N$ rule. For the "Industry" dataset, the two-fund and three-fund rules underperform the $1/N$ rule when $h = 120$ months but outperform it when $h = 240$ months. For the "MKT/SMB/HML" dataset, the performance of the two-fund rule is slightly lower than that of the $1/N$ rule, but the three-fund rule either outperforms or has similar performance as the $1/N$ rule. For the "International" dataset, both the two-fund and three-fund rules underperform the $1/N$ rule.

In summary, results in Tables 4 and 5 support our argument that whether the $1/N$ rule can outperform the optimal portfolio rules critically depends on the parameter values. A higher Sharpe ratio of the tangency portfolio (i.e., θ) suggests more gains from optimization. A lower Sharpe ratio of the $1/N$ portfolio relative to the tangency portfolio (i.e., θ_{ew}/θ) indicates that the equally weighted portfolio of risky assets is further away from the true optimal portfolio. As a result, it is more difficult for the $1/N$ rule to beat the optimal portfolio rules for cases when θ is high and θ_{ew}/θ is low.

3. Portfolio Rules without Risk-free Asset

In this section, we consider the portfolio choice problem when the portfolio is restricted to just risky assets. We first present the optimal portfolio when μ and Σ are known, then study the properties of the portfolio obtained based on the ML rule to understand the estimation risk. Armed with this understanding, we proceed to derive a new optimal portfolio rule to mitigate the estimation risk. Finally, we compare the performance of the newly derived portfolio rule with that of the ML rule and the $1/N$ rule in the case without risk-free asset.

3.1 The Setup

Suppose an investor considers a portfolio of only risky assets. As before, he chooses his portfolio weights w to maximize the mean-variance utility function

$$U(w) = w'\mu - \frac{\gamma}{2} w'\Sigma w, \tag{45}$$

where γ is the coefficient of risk aversion, but now there is an additional constraint of $1_N' w = 1$. It is easy to show that, when both μ and Σ are known, the weights of the optimal portfolio p^* are

$$w^* = w_g + \frac{1}{\gamma} w_z, \tag{46}$$

where

$$w_g = \frac{\Sigma^{-1} 1_N}{1_N' \Sigma^{-1} 1_N}, \tag{47}$$

$$w_z = \Sigma^{-1} \left(\mu - 1_N \frac{1_N' \Sigma^{-1} \mu}{1_N' \Sigma^{-1} 1_N} \right) = \Sigma^{-1} (\mu - 1_N \mu_g). \tag{48}$$

In the familiar mean-variance frontier, w_g is the weights of the global minimum-variance portfolio, and w_z is the weights of a zero investment portfolio (i.e., $1_N' w_z = 0$). Equation (46) says that holding the optimal portfolio is the same as investing into two funds, w_g and w_z. Investors always hold 100% of w_g, and depending on their degrees of risk aversion, their exposures to w_z vary. It is clear from (46) that any portfolio on the minimum-variance frontier is a linear combination of w_g and w_z. As the risk aversion varies, the optimal portfolio from (46) will trace out the upper half of the minimum-variance frontier.

Let $r_{p^*,t+1} = w^{*\prime} r_{t+1}$ be the out-of-sample return of portfolio p^*. The mean and variance of $r_{p^*,t+1}$ are given by

$$\mu_{p^*} = \mu_g + \frac{\psi^2}{\gamma}, \tag{49}$$

$$\sigma_{p^*}^2 = \sigma_g^2 + \frac{\psi^2}{\gamma^2}, \tag{50}$$

where $\psi^2 = \mu' \Sigma^{-1} \mu - (1_N' \Sigma^{-1} \mu)^2 / (1_N' \Sigma^{-1} 1_N)$. It follows that the utility from holding the optimal portfolio is

$$U(w^*) = \mu_g - \frac{\gamma}{2} \sigma_g^2 + \frac{\psi^2}{2\gamma}. \tag{51}$$

This equation shows that w^* outperforms w_g by a certainty equivalent return of $\psi^2/(2\gamma)$, which is coming from the exposure to the zero investment portfolio w_z and is determined by the slope of the asymptote to the *ex ante* minimum-variance frontier (ψ) and the risk aversion coefficient (γ).

3.2 The ML Rule

In practice, however, the optimal portfolio weights, w^*, are not computable because μ and Σ are unknown, and they need to be estimated. Similar to the case with risk-free asset, the maximum likelihood estimator of w^* at time t is given by

$$\hat{w}_t = \hat{w}_{g,t} + \frac{1}{\gamma}\hat{w}_{z,t}, \tag{52}$$

where

$$\hat{w}_{g,t} = \frac{\hat{\Sigma}_t^{-1}1_N}{1_N'\hat{\Sigma}_t^{-1}1_N}, \tag{53}$$

$$\hat{w}_{z,t} = \hat{\Sigma}_t^{-1}(\hat{\mu}_t - 1_N\hat{\mu}_{g,t}), \tag{54}$$

with $\hat{\mu}_{g,t} = (1_N'\hat{\Sigma}_t^{-1}\hat{\mu}_t)/(1_N'\hat{\Sigma}_t^{-1}1_N)$. We call $\hat{w}_t$ the ML rule, and denote this portfolio as portfolio p and its out-of-sample portfolio return as $r_{p,t+1} = \hat{w}_t' r_{t+1}$. The following Proposition presents the exact distribution of $r_{p,t+1}$ and the expected out-of-sample utility of portfolio p.

Proposition 3: *Let z_2, u_0, v_2, w_1, w_2, s_1, s_2, x_{21}, a, y_1, y_2 and $\hat{\psi}_t^2$ be the set of random variables defined in Proposition 2. Then, the conditional mean and variance of the out-of-sample return of portfolio p are given by*

$$\mu_{p,t} = \mu_g + \frac{\sigma_g\psi}{\hat{\psi}_t}\left(\frac{\sqrt{u_0}y_1}{\sqrt{v_2}} + \frac{az_2}{v_2}\right) + \frac{\sqrt{h}\psi}{\gamma v_2}\left(\frac{x_{21}\sqrt{u_0}}{\sqrt{w_2}} + z_2\right), \tag{55}$$

$$\sigma_{p,t}^2 = \sigma_g^2\left(y_1^2 + y_2^2 + 1 + \frac{s_1}{w_1} + \frac{a^2}{v_2}\right) + \frac{h\hat{\psi}_t^2}{\gamma^2 v_2}\left(1 + \frac{x_{21}^2 + s_2}{w_2}\right) + \frac{2\sqrt{h}\sigma_g\hat{\psi}_t}{\gamma\sqrt{v_2}}\left(\frac{a}{\sqrt{v_2}} + \frac{x_{21}y_1}{\sqrt{w_2}} + \frac{\sqrt{s_2}y_2}{\sqrt{w_2}}\right). \tag{56}$$

When $h > N+3$, the expected out-of-sample utility of portfolio p is given by

$$E[U(\hat{w}_t)] = \mu_g - \frac{\gamma(h-2)\sigma_g^2}{2(h-N-1)} + \frac{h}{\gamma(h-N-1)}\left[\psi^2 - \frac{(h-2)(h\psi^2+N-1)}{2(h-N)(h-N-3)}\right]. \tag{57}$$

Proposition 3 suggests that both the unconditional distribution and the out-of-sample utility of portfolio p depend only on μ_g, σ_g^2, ψ^2, h, N, and γ. Therefore, there is no need to specify μ and Σ when computing the distribution and the expected out-of-sample utility of the ML rule.

In addition, Proposition 3 expresses the unconditional distribution of $r_{p,t+1}$ in terms of a set of univariate random variables, which provides a fast way of computing the distribution.

Figure 4 plots the density function of $r_{p,t+1}$ for estimation windows of $h = 60$ and 120 months, with parameter values calibrated using monthly excess returns of the ten momentum portfolios from January 1927 to December 2014. The risk aversion coefficient is set to be $\gamma = 3$. For comparison, we also include the density function of $r_{p^*,t+1}$, the return of the true optimal portfolio (i.e., $h = \infty$). Relative to $r_{p^*,t+1}$, Figure 4 shows that the return of the ML rule (i.e., $r_{p,t+1}$) is much more volatile than $r_{p^*,t+1}$, which can be explained by the significant amount of estimation errors involved in the ML rule. For a shorter estimation window (i.e., $h = 60$), there is more estimation risk, and therefore, $r_{p,t+1}$ becomes more volatile. Figure 5 presents similar plots but with parameters calibrated using monthly excess returns of the Fama-French 25 size and book-to-market ranked portfolios from January 1927 to December 2014. With more risky assets ($N = 25$), the ML rule involves more estimation errors. As a result, $r_{p,t+1}$ becomes more volatile in Figure 5 than in Figure 4.

Based on the results in Proposition 3, it is easy to obtain the unconditional mean and variance of $r_{p,t+1}$:

$$\mu_p = E[\mu_{p,t}] = \mu_g + \frac{h\psi^2}{\gamma(h-N-1)} \quad \text{for } h > N+1, \tag{58}$$

$$\begin{aligned} \sigma_p^2 &= E[\sigma_{p,t}^2] + E[\mu_{p,t}^2] - E[\mu_{p,t}]^2 \\ &= \frac{\sigma_g^2(h+\psi^2-2)}{h-N-1} + \frac{h(h-2)(h+1)\psi^2}{\gamma^2(h-N)(h-N-1)(h-N-3)} \\ &\quad + \frac{h(h-2)(N-1)}{\gamma^2(h-N)(h-N-1)(h-N-3)} + \frac{2h^2\psi^4}{\gamma^2(h-N-1)^2(h-N-3)} \quad \text{for } h > N+3. \end{aligned} \tag{59}$$

Comparing the unconditional mean and variance of portfolio p with those of the true optimal portfolio p^*, we can see that $\mu_p > \mu_{p^*}$ for $h > N+1$ and $\sigma_p^2 > \sigma_{p^*}^2$ for $h > N+3$.[12] As $h \to \infty$, μ_p converges to μ_{p^*}, and σ_p^2 converges to $\sigma_{p^*}^2$. Moreover, it is easy to verify that both μ_p and σ_p^2 are decreasing functions of h and increasing functions of N.

Comparing the expected out-of-sample utility of portfolio p with that of the true optimal port-

[12]When $h > N+1$, the coefficient of ψ^2/γ in (58), $h/(h-N-1) > 1$, which results in $\mu_p > \mu_{p^*}$. When $h > N+3$, the coefficients of the first two terms in (59), $(h+\psi^2-2)/(h-N-1) > 1$ and $\frac{h(h-2)(h+1)}{(h-N)(h-N-1)(h-N-3)} > 1$, and the last two terms in (59) are positive, which leads to $\sigma_p^2 > \sigma_{p^*}^2$.

folio p^*, we can figure out the utility loss due to the estimation errors involved in portfolio p:

$$\begin{aligned} L(\hat{w}_t, \hat{w}^*) &= U(w^*) - E[U(\hat{w}_t)] \\ &= \frac{\gamma\sigma_g^2(N-1)}{2(h-N-1)} + \frac{k_0\psi^2}{2\gamma} + \frac{(N-1)h(h-2)}{2\gamma(h-N)(h-N-1)(h-N-3)}, \end{aligned} \tag{60}$$

where

$$k_0 = \frac{(N+1)[h^2+(N+1)(h-N-3)]}{(h-N-1)^2(h-N-3)}. \tag{61}$$

The first term in (60) captures the utility loss due to estimation errors of $\hat{w}_{g,t}$, and the remaining two terms reflect the utility loss due to the estimation errors of $\hat{w}_{z,t}$. Both components decrease with h and increase with N. As $h \to \infty$, both components go to zero.

3.3 The QL Rule

The ML rule involves estimation errors which could lead to significant utility loss. Next, we derive an optimal portfolio rule that maximizes the expected out-of-sample utility, taking into account the estimation risk. Specifically, we limit our attention to the class of portfolio rules that have weights

$$\hat{w}_t(c) = \hat{w}_{g,t} + \frac{c}{\gamma}\hat{w}_{z,t}, \tag{62}$$

where c is a constant scalar. The ML rule is a special case of this class with $c = 1$. When there is estimation risk, it makes sense to allow for c to differ from one. We look for the optimal c to maximize the expected out-of-sample utility

$$E[U(\hat{w}_t(c))] = E[\hat{w}_t(c)'\mu] - \frac{\gamma}{2}E[\hat{w}_t(c)'\Sigma\hat{w}_t(c)]. \tag{63}$$

Interestingly, the search for this optimal c is equivalent to finding the c that minimizes the standard quadratic loss function on the estimated portfolio weights from statistical decision theory

$$L(\hat{w}_t(c), w^*) = E[(\hat{w}_t(c) - w^*)'\Sigma(\hat{w}_t(c) - w^*)]. \tag{64}$$

To see this, we use (46) and expand the quadratic loss function as

$$\begin{aligned} L(\hat{w}_t(c), w^*) &= E\left[\hat{w}_t(c)'\Sigma\hat{w}_t(c) - 2w^{*\prime}\Sigma\hat{w}_t(c) + w^{*\prime}\Sigma w^*\right] \\ &= E\left[\hat{w}_t(c)'\Sigma\hat{w}_t(c) - \frac{2 1_N'\hat{w}_t(c)}{1_N'\Sigma^{-1}1_N} - \frac{2}{\gamma}\hat{w}_t(c)'(\mu - 1_N\mu_g) + w^{*\prime}\Sigma w^*\right] \end{aligned}$$

$$= \frac{2}{\gamma} E\left[-\hat{w}_t(c)'\mu + \frac{\gamma}{2}\hat{w}_t(c)'\Sigma\hat{w}_t(c)\right] - 2\sigma_g^2 + \frac{2\mu_g}{\gamma} + \sigma_g^2 + \frac{\psi^2}{\gamma^2}$$

$$= \frac{2}{\gamma}(U(w^*) - E[U(\hat{w}_t(c)]). \tag{65}$$

The second last equality holds because $\hat{w}_t(c)'1_N = 1$ and $\sigma_g^2 = 1/(1_N'\Sigma^{-1}1_N)$, and the last equality follows from (51). It is then clear that maximizing $E[U(\hat{w}_t(c))]$ is the same as minimizing $L(\hat{w}_t(c), w^*)$.[13]

Using the results in the proof of Proposition 3, it can be readily shown that the optimal c is given by

$$c^* = \frac{\tilde{k}_3\psi^2}{\psi^2 + \frac{N-1}{h}} \tag{66}$$

with

$$\tilde{k}_3 = \frac{(h-N)(h-N-3)}{h(h-2)}. \tag{67}$$

Note that $c^* < 1$, so it is optimal to invest a smaller amount of investment in $\hat{w}_{z,t}$ when there is estimation risk. Equivalently, the investor chooses to hold a portfolio as if he had a higher risk aversion coefficient. The out-of-sample utility given c^* is

$$E[U(\hat{w}_t(c^*))] = \mu_g - \frac{\gamma}{2}\frac{(h-2)\sigma_g^2}{(h-N-1)} + \frac{h\psi^2\tilde{k}_3}{2\gamma(h-N-1)}\left(\frac{\psi^2}{\psi^2 + \frac{N-1}{h}}\right). \tag{68}$$

Note that c^* depends on ψ^2 which is unknown to investors in practice. Therefore, $\hat{w}_t(c^*)$ is not attainable. Following Kan and Zhou (2007), an implementable version of this optimal two-fund rule can be obtained as

$$\hat{w}_t(\hat{c}_t) = \hat{w}_{g,t} + \frac{\hat{c}_t}{\gamma}\hat{w}_{z,t}, \tag{69}$$

with

$$\hat{c}_t = \frac{\tilde{k}_3\hat{\psi}_{a,t}^2}{\hat{\psi}_{a,t}^2 + \frac{N-1}{h}} \tag{70}$$

[13] The equivalence between these two objective functions applies to more general portfolio rules as well as to the case with risk-free asset.

where $\hat{\psi}_{a,t}^2$ is defined in (23). Let g_3 be a function of $\hat{\psi}_t^2$ with

$$g_3(\hat{\psi}_t^2) = \frac{\hat{\psi}_{a,t}^2}{\hat{\psi}_{a,t}^2 + \frac{N-1}{h}}, \tag{71}$$

we have

$$\hat{w}_t(\hat{c}_t) = \hat{w}_{g,t} + \frac{\tilde{k}_3 g_3(\hat{\psi}_t^2)}{\gamma}\hat{w}_{z,t}. \tag{72}$$

We call this portfolio rule the QL rule because of its quadratic loss motivation, and denote it as portfolio q. Let $r_{q,t+1} = \hat{w}_t(\hat{c}_t)' r_{t+1}$ be the out-of-sample return of portfolio q at time $t+1$. The following Proposition presents the exact distribution of $r_{q,t+1}$ and the expected out-of-sample utility of portfolio q.

Proposition 4: *Let z_2, u_0, v_2, w_1, w_2, s_1, s_2, x_{21}, a, y_1, y_2, and $\hat{\psi}_t$ be the set of random variables defined in Proposition 2. Then, the conditional mean and variance of the out-of-sample return of portfolio q are given by*

$$\mu_{q,t} = \mu_g + \frac{\sigma_g \psi}{\hat{\psi}_t}\left(\frac{\sqrt{u_0}y_1}{\sqrt{v_2}} + \frac{a z_2}{v_2}\right) + \frac{\sqrt{h}\psi \tilde{k}_3 g_3(\hat{\psi}_t^2)}{\gamma v_2}\left(\frac{x_{21}\sqrt{u_0}}{\sqrt{w_2}} + z_2\right), \tag{73}$$

$$\begin{aligned}\sigma_{q,t}^2 = {} & \sigma_g^2\left(y_1^2 + y_2^2 + 1 + \frac{s_1}{w_1} + \frac{a^2}{v_2}\right) + \frac{h\tilde{k}_3^2 g_3^2(\hat{\psi}_t^2)\hat{\psi}_t^2}{\gamma^2 v_2}\left(1 + \frac{x_{21}^2 + s_2}{w_2}\right) \\ & + \frac{2\sqrt{h}\sigma_g \tilde{k}_3 g_3(\hat{\psi}_t^2)\hat{\psi}_t}{\gamma\sqrt{v_2}}\left(\frac{a}{\sqrt{v_2}} + \frac{x_{21}y_1}{\sqrt{w_2}} + \frac{\sqrt{s_2}y_2}{\sqrt{w_2}}\right).\end{aligned} \tag{74}$$

The expected out-of-sample utility of portfolio q is given by

$$E[U(\hat{w}_t(\hat{c}_t)] = \mu_g - \frac{\gamma(h-2)\sigma_g^2}{2(h-N-1)} + \frac{\tilde{k}_3 h \psi^2 E[g_3(q_3)]}{\gamma(h-N-1)} - \frac{\tilde{k}_3(h-N-3)E[g_3^2(q_4)q_4]}{2\gamma(h-N-1)} \tag{75}$$

when $h > N+3$, where $g_3(\cdot)$ is the function defined in (71), $q_3 \sim \mathscr{G}_{N+1,h-N-1}^{h\psi^2}$, $q_4 \sim \mathscr{G}_{N-1,h-N-1}^{h\psi^2}$.

Proposition 4 expresses the unconditional distribution of $r_{q,t+1}$ in terms of a set of univariate random variables and the out-of-sample utility of portfolio q in terms of a one-dimensional integral. Similar to the ML rule, Proposition 4 suggests that both the unconditional distribution and the expected out-of-sample utility of the QL rule depend only on μ_g, σ_g^2, ψ^2, h, N, and γ. Therefore, there is no need to specify μ and Σ when computing the distribution and the out-of-sample utility.

In Figures 6 and 7, we plot the density function of $r_{q,t+1}$ for estimation windows of $h = 60$ and 120 months with parameters calibrated using monthly excess returns of the ten momentum portfolios (Figure 6) or those of the Fama-French 25 size and book-to-market portfolios (Figure 7) in the period from January 1927 to December 2014. For comparison, we also include the distribution of $r_{p^*,t+1}$, the return of the true optimal portfolio (i.e., $h = \infty$). The risk aversion coefficient is set to be $\gamma = 3$.

Relative to the distribution of $r_{p,t+1}$ in Figures 4 and 5, Figures 6 and 7 show that the distribution of $r_{q,t+1}$ is much less dispersed. Since $\hat{c}_t < 1$, the QL rule invests less in $\hat{w}_{z,t}$ than the ML rule. Because of the estimation errors involved in $\hat{w}_{z,t}$, a lower exposure to $\hat{w}_{z,t}$ makes $r_{q,t+1}$ less volatile than $r_{p,t+1}$. In addition, we can see from Figures 6 and 7 that unlike the ML rule, the mean of $r_{q,t+1}$ is lower than that of the true optimal portfolio. This is also due to the fact that $\hat{c}_t < 1$ and the QL rule has less exposure to $\hat{w}_{z,t}$, which leads to a lower unconditional mean.

Based on the results in Proposition 4, we can show that the unconditional mean and variance of $r_{q,t+1}$ are given by:

$$\mu_q = E[\mu_{q,t}] = \mu_g + \frac{\tilde{k}_3 h\psi^2}{\gamma(h-N-1)} E[g_3(q_3)] \quad \text{for } h > N+1, \tag{76}$$

$$\begin{aligned} \sigma_q^2 &= E[\sigma_{q,t}^2] + E[\mu_{q,t}^2] - E[\mu_{q,t}]^2 \\ &= \frac{(h-2+\psi^2)\sigma_g^2}{h-N-1} + \frac{h\tilde{k}_3^2(h-2+\psi^2)E[g_3^2(q_4)q_4]}{\gamma^2(h-N)(h-N-1)} - \frac{h^2\tilde{k}_3^2\psi^4E^2[g_3(q_3)]}{\gamma^2(h-N-1)^2} \\ &\quad + \frac{h\tilde{k}_3^2\psi^2\left(E[g_3^2(q_5)] + h\psi^2E[g_3^2(q_6)]\right)}{\gamma^2(h-N)(h-N-3)} \quad \text{for } h > N+3, \end{aligned} \tag{77}$$

where q_3 and q_4 are defined in Proposition 2, $q_5 \sim \mathscr{G}_{N+1,h-N-3}^{h\psi^2}$ and $q_6 \sim \mathscr{G}_{N+3,h-N-3}^{h\psi^2}$.

Note that when $N > 1$, $\tilde{k}_3 h/(h-N-1) < 1$ and $g_3(\cdot) < 1$, and therefore $\mu_q < \mu_{p^*}$. As $h \to \infty$, μ_q converges to μ_{p^*}. The relation between σ_q^2 and $\sigma_{p^*}^2$ is not straightforward from (77), but it can be shown that as $h \to \infty$, σ_q^2 converges to $\sigma_{p^*}^2$.

The utility loss of portfolio q relative to the optimal portfolio p^* is

$$\begin{aligned} L(\hat{w}_t(\hat{c}_t), w^*) &= U(w^*) - E[U(\hat{w}_t(\hat{c}_t))] \\ &= \frac{\gamma\sigma_g^2(N-1)}{2(h-N-1)} + \frac{\psi^2}{2\gamma}\left(1 - \frac{2h\tilde{k}_3E[g_3(q_3)]}{h-N-1}\right) + \frac{\tilde{k}_3(h-N-3)E[g_3^2(q_4)q_4]}{2\gamma(h-N-1)}. \end{aligned} \tag{78}$$

The first term in (78) is the same as that in (60), which captures the utility loss due to estimation errors of $\hat{w}_{g,t}$. The remaining two terms are the utility loss coming from estimation errors of $\hat{c}_t \hat{w}_{z,t}$.

3.4 Comparison of Portfolio Rules: Analytical Results

In this subsection, we compare analytically the performance of the ML rule (i.e., portfolio p), the QL rule (i.e., portfolio q), and that of the $1/N$ rule. Figure 8 plots the expected utility of these portfolios using parameters calibrated to the monthly excess returns of the ten momentum portfolio over the period 1927/1 – 2014/12. For comparison, we also include the utility of the true optimal portfolio p^*. The risk aversion coefficient is set to $\gamma = 3$.

Figure 8 shows that as the estimation window increases, the out-of-sample utilities of both the ML rule and the QL rule go toward the utility level of the optimal portfolio. However, due to the significant estimation errors involved in the ML rule, the out-of-sample utility of the ML rule is negative when h is small. A minimum estimation window of $h = 168$ month is required for the ML rule to generate positive out-of-sample utility given the parameter values. The QL rule, taking into account the estimation errors, performs much better than the ML rule, especially when the estimation window is short. The QL rule generates positive expected out-of-sample utility even with $h = 60$ months.

In the case without risk-free asset, the out-of-sample utility of the $1/N$ rule is constant and it is equal to

$$U(w_{ew}) = \mu_{ew} - \frac{\gamma}{2}\sigma_{ew}^2. \tag{79}$$

Let $\tilde{\sigma}_{ew}^2$ be the variance of the minimum-variance portfolio with mean μ_{ew}, we can use the fact that (see, for example, Eq.(29) in Kan and Smith (2008))

$$\tilde{\sigma}_{ew}^2 = \sigma_g^2 + \frac{(\mu_{ew} - \mu_g)^2}{\psi^2}, \tag{80}$$

to decompose the utility loss of the $1/N$ rule into two components as

$$\begin{aligned} L(w_{ew}, w^*) &= U(w^*) - U(w_{ew}) \\ &= \mu_g - \mu_{ew} + \frac{\gamma}{2}(\sigma_{ew}^2 - \sigma_g^2) + \frac{\psi^2}{2\gamma} \\ &= \frac{\gamma}{2}(\sigma_{ew}^2 - \tilde{\sigma}_{ew}^2) + \frac{\gamma}{2\psi^2}\left(\mu_{ew} - \mu_g - \frac{\psi^2}{\gamma}\right)^2 \end{aligned}$$

$$= \frac{\gamma}{2}(\sigma_{ew}^2 - \tilde{\sigma}_{ew}^2) + \frac{\gamma}{2\psi^2}(\mu_{ew} - \mu_{p^*})^2, \tag{81}$$

where the last equation follows from (49). The first term in (81) captures the utility loss due to the inefficiency of the $1/N$ rule. In other words, to what extent the variance of the $1/N$ rule (i.e., σ_{ew}^2) is larger than the minimum variance (i.e., $\tilde{\sigma}_{ew}^2$) for a portfolio with expected return of μ_{ew}. The second term in (81) captures the utility loss coming from ignoring the risk aversion of the investor. Note that when taking into account the risk aversion of the investor, the optimal level of expected return is μ_{p^*} instead of μ_{ew}, so not adjusting the portfolio based on the risk aversion coefficient imposes a utility loss to the investor. Given the parameter values in Figure 8, the utility loss due to the first component is 0.1513 and that due to the second component is 0.8985.

Whether the ML and the QL rules can outperform the $1/N$ rule with a reasonable length of estimation window depends on how close the utility level of the $1/N$ rule is to that of the true optimal portfolio, which is ultimately determined by the two components in (81). The closer the utility of the $1/N$ rule to the true optimal (i.e., smaller $L(w_{ew}, w^*)$), the longer the estimation window is required for the ML and the QL rules to beat the $1/N$ rule. Given the parameter values in Figure 8, the QL rule outperforms the $1/N$ rule even with an estimation window of $h = 60$ months. The ML rule, with more estimation errors, needs a longer estimation window $h = 182$ months to beat the $1/N$ rule.

Figure 9 presents similar plots with parameters calibrated to the monthly excess returns of the 25 Fama-French size and book-to-market portfolios in the period of 1927/1–2014/12. Given more risky assets $N = 25$, there are more estimation errors involved in the ML and the QL rules. The estimation window required for the ML rule to generate positive utility is now $h = 382$ months, and that of the QL rule is $h = 62$ months. With the parameter values in Figure 9, the two components in (81) are 0.4254 and 0.9157 respectively. Ignoring the risk aversion of the investor continues to be the major determinant of the utility loss of the $1/N$ rule. Now, the QL rule needs an estimation window of at least $h = 77$ months to outperform the $1/N$ rule, and that of the ML rule is $h = 398$ months.

In the case without risk-free asset, the relative rankings of the ML rule, the QL rule, and the $1/N$ rule are no longer invariant to the value of γ.[14] In Figures 10 and 11, we plot the results

[14]The relative ranking of the ML rule and the QL rule is invariant to the value of γ because the first two terms in (57) and (75) are the same. However, it is not the case when the $1/N$ rule is also considered.

with a different risk aversion coefficient $\gamma = 1$. With parameters calibrated to the returns of the ten momentum portfolios in Figure 10, the required estimation window for the QL and the ML rules to outperform the $1/N$ rule are $h = 63$ months and $h = 272$ months, respectively, when $\gamma = 1$. In Figure 11, with parameters calibrated to the returns of the 25 size and book-to-market portfolios and $\gamma = 1$, the required estimation windows for the QL and the ML rules are $h = 101$ months and $h = 467$ months, respectively.

To get a better understanding of the performance of the optimal rules (i.e., the ML and the QL rule) relative to the $1/N$ rule, in Figure 12, we conduct analysis similar to Figure 3 in the case without risk-free asset. Specifically, we present the required length of the estimation window for the optimal rules to outperform the $1/N$ rule as a function of the number of assets (N). As the relative performance of the optimal rules and the $1/N$ rule is not invariant to the value of γ when a risk-free asset is not available, we report results for both $\gamma = 1$ and $\gamma = 3$. The parameters in the six panels of Figure 12 are set in the same way as those in Figure 3, i.e., $\theta = 0.4$ or 0.2, $\theta_{ew}/\theta = 0.25$, 0.50, or 0.75, and $\theta_g = \theta/2$. To obtain the expected utilities of the ML and the QL rules, in addition to θ and θ_g, we also need to specify μ_g or σ_g. We set $\sigma_g = 0.05$, a value close to the sample estimates as shown in Figures 8–11. We choose to specify σ_g instead of μ_g for a given θ_g because the sample estimate of σ_g involves less estimation error compared to that of μ_g. Similarly, we need to specify μ_{ew} or σ_{ew} to compute the expected utility of the $1/N$ rule. Based on the sample estimates in Figures 8–11, we set $\sigma_{ew} = 0.065$.

It can be seen that the results in Figure 12 are qualitatively very similar to those in Figure 3, suggesting that the insights we obtain in the case with a risk-free asset also apply to the case without a risk-free asset. First, Figure 12 shows that the QL rule significantly outperforms the ML rule in all cases, and the required length of the QL rule is reasonable in general even for the case with large N. For example, with $\theta = 0.4$ and $\theta_{ew} = 0.1$ (i.e., panel (a)), even for the case of $N = 100$, the required length for the QL rule is 147 months with $\gamma = 1$ and 163 months with $\gamma = 3$, while the ML rule needs an estimation window of 1149 ($\gamma = 1$) and 1001 ($\gamma = 3$) months to beat the $1/N$ rule. The long estimation window required for the ML rule is due to the severe estimation errors involved. The QL rule, taking into account the estimation errors, improves the expected out-of-sample utility significantly, and therefore, needs a much shorter estimation window to outperform the $1/N$ rule.

Second, Figure 12 suggests that a longer estimation window is required with a higher θ_{ew}, all else equal. For example, for the case of $N = 100$, the results in panels (a), (c), and (e) show that the required length for the QL rule are 147 and 163 months with $\gamma = 1$ and $\gamma = 3$ when $\theta_{ew} = 0.1$, and the numbers are 208 and 281 respectively when $\theta_{ew} = 0.2$ and 317 and 704 respectively when $\theta_{ew} = 0.3$. Given the same σ_{ew}, a higher θ_{ew} means a higher out-of-sample utility of the $1/N$ rule, and therefore, it is more difficult for the optimal rules to beat the $1/N$ rule.

Third, we can see from Figure 12 that a longer estimation window is required when θ is lower, all else equal. For example, given the same $\theta_{ew}/\theta = 0.25$, with a lower $\theta = 0.2$ in panel (b), the required length of estimation window for the QL rule are 251 and 209 months with $\gamma = 1$ and $\gamma = 3$ for $N = 100$, compared to 147 and 163 months in panel (a). A lower θ reduces the benefit of optimization relative to the cost of estimation errors, and therefore, a longer estimation window is needed for the optimal rules to outperform the $1/N$ rule.

Finally, we explore the effect of γ on the required length of estimation window for the optimal rules to outperform the $1/N$ rule. Results in Figures 12 indicates that whether the required h is larger for $\gamma = 1$ or $\gamma = 3$ varies for different parameter values. But from Figure 12 we can see that as θ_{ew} increases, it is more likely for the case with $\gamma = 3$ to have a larger required h. This holds for both the ML and the QL rule. For example, for the case with $\theta = 0.4$ and $N = 10$, the required h are 110 ($\gamma = 1$) and 96 ($\gamma = 3$) months for the ML rule given $\theta_{ew} = 0.1$. When $\theta_{ew} = 0.2$, the required h are both 119 months, and when $\theta_{ew} = 0.3$, the required h for $\gamma = 1$ and $\gamma = 3$ are 131 and 164 months, respectively. In the case of the QL rule, the required h are 30 months ($\gamma = 1$) and 25 months ($\gamma = 3$) when $\theta_{ew} = 0.1$, 37 months ($\gamma = 1$) and 40 months ($\gamma = 3$) when $\theta_{ew} = 0.2$, and 47 months ($\gamma = 1$) and 83 months ($\gamma = 3$) when $\theta_{ew} = 0.3$. From (57), the difference between the performance of the ML and the $1/N$ rules is given by

$$E[U(\hat{w}_t)] - U(w_{ew}) = \mu_g - \mu_{ew} + \frac{\gamma}{2}\left(\sigma_{ew}^2 - \frac{(h-2)\sigma_g^2}{h-N-1}\right) + \frac{h}{\gamma(h-N-1)}\left[\psi^2 - \frac{(h-2)(h\psi^2+N-1)}{2(h-N)(h-N-3)}\right]. \quad (82)$$

We find that the terms involving γ are positive evaluating at the required h, but the terms involving $1/\gamma$ can be negative for small h and become positive for large h. In addition, the magnitude of the terms involving $1/\gamma$ is larger than those involving γ. Notice that the terms involving $1/\gamma$ reflects the utility due to w_z. Estimation errors can significantly deteriorate its utility when h is small. Given a low level of θ_{ew}, a small h is required for the optimal rules to outperform the $1/N$ rule. A small h tends to make the terms involving $1/\gamma$ negative, and a higher γ will lead to a higher relative performance for the ML rule, holding all other parameter values unchanged. Therefore, a lower h is required for a higher value of γ when θ_{ew} is low. As θ_{ew} increases, a larger h is required, making the terms involving $1/\gamma$ positive. In that case, a higher γ results in a lower relative performance for the ML rule, and a higher h is required for the optimal rules to outperform the $1/N$ rule. Similar logic also applies to the QL rule.

In summary, the results in this subsection suggest that in the case without risk-free asset, the QL rule successfully mitigates the estimation errors involved in the ML rule, and significantly improves the expected out-of-sample utility. In addition, we find that the QL rule outperforms the $1/N$ rule with reasonable length of the estimation window based on the parameters calibrated to the real data.

3.5 Comparison of Portfolio Rules: Empirical Results

In this subsection, we compare empirically the performance of the ML rule, the QL rule, and the $1/N$ rule using the same datasets as in Tables 4 and 5 for different combinations of h (120 or 240 months) and γ (1 or 3). Note that in the case without risk-free asset, the optimal portfolio p^* is not the one with highest Sharpe ratio, and the Sharpe ratio of the optimal portfolio p^* depends on γ. Therefore, Sharpe ratio is not an appropriate measure of performance in the case without risk-free asset, and we focus our attention to the certainty-equivalent returns.[15]

In Table 6, we report portfolio performance based on the certainty-equivalent returns. Results in Table 6 are consistent with our findings from the analytical results from the previous subsection. First, due to estimation errors, we can see that the ML rule has poor performance. Except for the set "MKT/SMB/HML" with $\gamma = 3$, the ML performance is negative in all other cases. In addition, we find that the ML performance improves, in general, with a higher risk aversion coefficient $\gamma = 3$ and a longer estimation window $h = 240$ months. A higher risk aversion coefficient leads to a smaller exposure to $\hat{w}_{z,t}$, which lowers the effect of estimation errors. A longer estimation window

[15]Nevertheless, we also examine Sharpe ratios in untabulated results, and find similar pattern as that of the CEQ results.

helps to provide more accurate estimates of the parameters, which also reduces the influence of estimation risk. As a result, the out-of-sample utility of the ML rule increases. Second, Table 6 shows that the QL rule outperforms the ML rule in all cases. This is because the QL rule takes into account the estimation error.

Finally, comparing the performance of the optimal portfolio rules, in particular that of the QL rule, with the performance of the $1/N$ rule, we find that the QL rule outperforms the $1/N$ rule in most cases. In datasets "FF 1-factor", "FF 3-factor", and "FF 4-factor", the QL rule outperforms the $1/N$ rule for all combinations of h and γ. Note that in these three sets, the utility of the *ex post* optimal portfolio (i.e., MV in-sample) is significantly higher than that of the $1/N$ rule, which explains the dominance of the QL rule. In the other three datasets, the utility level of the $1/N$ rule is closer to that of the *ex post* optimal portfolio, which makes it more difficult for the QL rule to outperform the $1/N$ rule. In dataset "Industry," the QL rule underperforms the $1/N$ rule when $h = 120$ months, and has similar or higher level of utility than the $1/N$ rule when $h = 240$ months. Similar situation can be observed for the dataset "International." For the dataset "MKT/SMB/HML," the QL rule outperforms the $1/N$ rule when $h = 120$ months but underperforms it when $h = 240$ months.

In summary, Tables 6, based on real datasets, confirms the findings from the analytical results of the previous subsection. In the case without a risk-free asset, the newly proposed QL rule can reduce the effect of estimation risk of the ML rule and significantly improve the portfolio performance. In addition, the QL rule performs well relative to the $1/N$ rule.

4. Conclusion

In this paper, we analyze optimal portfolio problems with and without risk-free asset in the presence of estimation risk. Our analysis explicitly takes into account the estimation errors in expected returns and covariance matrix of the risky assets. Instead of the usual plug-in portfolio rule that replaces the population parameters by the sample estimates, we consider optimal portfolios that are designed to mitigate the impact of estimation risk. Unlike earlier studies, such as Kan and Zhou (2007) and Tu and Zhou (2011), which rely on simulations to compute the expected out-of-sample utility of various sample optimal portfolio rules, we derive their out-of-sample utility explicitly in terms of 1-dimensional integrals. Besides allowing for speedy computation, our approach provides analytical insights to the problem and an understanding of what are the parameters that are important in determining the out-of-sample performance of these portfolio rules.

We also conduct comparisons of various sample optimal portfolio rules with equally weighted portfolio ($1/N$) rule. Specifically, we point out that the horse race conducted by DGU is somewhat unfair to various sample optimal portfolio rules that are designed to invest in both risk-free asset and risky assets. DGU normalize these optimal portfolio rules so that they invest in just risky assets. Due to the non-existence of moments for the out-of-sample return of such normalized portfolio rules, they have very poor out-of-sample performance and can be easily dominated by the $1/N$ rule. In this paper, we conduct two comparisons of various optimal portfolio rules with the $1/N$ rule. For the case with risk-free asset, we compare the optimal portfolio rules with the one that optimally combines the risk-free asset and $1/N$ rule (on risky assets). For the case without risk-free asset, we compare the $1/N$ rule with the optimal rules that are specifically designed for the case without risk-free asset. In both cases and by using both analytical comparison as well as actual data (using the same datasets in DGU), we find that the optimal portfolio rules that explicitly take into account of estimation errors can in general do quite well relative to the $1/N$ rule.

While insights from DGU cast some doubt on the value of existing investment theory, our paper re-affirms the value of the theory when used properly. Estimation risk is certainly a serious issue in portfolio analysis, but the problem is not as hopeless as suggested by DGU, and further research on this problem could generate portfolio rules with even better out-of-sample performance.

Appendix A: Proofs

We first cite two simple lemmas from Kan and Wang (2015). Suppose $z \sim \mathcal{N}(\mu, 1)$, $w \sim \chi^2_{m-1}$, $u \sim \chi^2_n$, and they are independent of each other. It follows that $v = z^2 + w \sim \chi^2_m(\delta)$, where $\delta = \mu^2$.

Lemma 1: Let $g(v)$ be a function of v. When the expectations exist, we have

$$E[g(v)z] = \mu E[g(v_1)], \tag{A1}$$

$$E[g(v)z^2] = E[g(v_1)] + \delta E[g(v_2)], \tag{A2}$$

where $v_1 \sim \chi^2_{m+2}(\delta)$, $v_2 \sim \chi^2_{m+4}(\delta)$.

Lemma 2: Let $g(y)$ be a function of $y = v/u \sim \mathcal{G}^{\delta}_{m,n}$. When the expectations exist, we have

$$E\left[\frac{g(y)}{u^k}\right] = \frac{E[g(y_1)]}{2^k \left(\frac{n}{2} - k\right)_k} \quad \text{for } k < \frac{n}{2}, \tag{A3}$$

where $y_1 \sim \mathcal{G}^{\delta}_{m,n-2k}$.

Proof of Proposition 1

Under the multivariate normality assumption on r_t, it is well known that $\hat{\mu}_t$ and $\hat{\Sigma}_t$ are independent of each other, and they have the following distributions:

$$\hat{\mu}_t \sim \mathcal{N}(\mu, \Sigma/h), \tag{A4}$$

$$\hat{\Sigma}_t \sim \mathcal{W}_N(h-1, \Sigma/h), \tag{A5}$$

where $\mathcal{W}_N(h-1, \Sigma/h)$ is a Wishart distribution with $h-1$ degrees of freedom and covariance matrix Σ/h. Let P be an $N \times N$ orthonormal matrix with its first column equals to η, $P = [\eta,\ P_1]$, where

$$\eta = \frac{\Sigma^{-\frac{1}{2}}\mu}{(\mu'\Sigma^{-1}\mu)^{\frac{1}{2}}} = \frac{\Sigma^{-\frac{1}{2}}\mu}{\theta}. \tag{A6}$$

Define

$$z = \sqrt{h}P'\Sigma^{-\frac{1}{2}}\hat{\mu}_t \sim \mathcal{N}\left(\left[\begin{array}{c} \sqrt{h}\theta \\ 0_{N-1} \end{array}\right], I_N\right), \tag{A7}$$

$$W = hP'\Sigma^{-\frac{1}{2}}\hat{\Sigma}_t\Sigma^{-\frac{1}{2}}P \sim \mathcal{W}_N(h-1, I_N), \tag{A8}$$

and they are independent of each other. Let $z_1 \sim \mathcal{N}(\sqrt{h}\theta, 1)$ be the first element of z, and $z'z = z_1^2 + u_0$, where $u_0 \sim \chi^2_{N-1}$ and it is independent of z_1. With the definition of z and W, we can write

$$\hat{\theta}_t^2 = \hat{\mu}_t'\hat{\Sigma}_t^{-1}\hat{\mu}_t = z'W^{-1}z, \tag{A9}$$

$$\mu_{\mathrm{II},t} = \frac{k_3 g_1(\hat{\theta}_t^2)}{\gamma}\sqrt{h}\theta e_1'W^{-1}z, \tag{A10}$$

$$\sigma^2_{\mathrm{II},t} = \frac{k_3^2 g_1^2(\hat{\theta}_t^2)}{\gamma^2} hz'W^{-2}z, \tag{A11}$$

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

$$A = (Q'W^{-1}Q)^{-1} = \begin{bmatrix} \tilde{z}'W^{-1}\tilde{z} & \tilde{z}'W^{-1}Q_1 \\ Q_1'W^{-1}\tilde{z} & Q_1'W^{-1}Q_1 \end{bmatrix}^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \sim \mathscr{W}_N(h-1, I_N), \tag{A12}$$

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

$$u_1 \equiv A_{11\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21} \sim \chi^2_{h-N}, \tag{A13}$$

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

$$-A_{22}^{-1}A_{21} \sim \frac{x}{\sqrt{u_2}}, \tag{A14}$$

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

$$\tilde{z}'W^{-1}\tilde{z} = A_{11\cdot 2}^{-1} = \frac{1}{u_1}, \tag{A15}$$

$$Q_1'W^{-1}\tilde{z} = -A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1} = \frac{x}{u_1\sqrt{u_2}}. \tag{A16}$$

With these identities, we can write

$$z'W^{-1}z = \frac{z_1^2 + u_0}{u_1}, \tag{A17}$$

$$z'W^{-2}z = z'W^{-1}(\tilde{z}\tilde{z}' + Q_1Q_1')W^{-1}z = \frac{z_1^2 + u_0}{u_1^2}\left(1 + \frac{x'x}{u_2}\right). \tag{A18}$$

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

$$\xi = \frac{(I_N - \tilde{z}\tilde{z}')e_1}{\left[e_1'(I_N - \tilde{z}\tilde{z}')e_1\right]^{\frac{1}{2}}} = \frac{(I_N - \tilde{z}\tilde{z}')e_1}{\sqrt{u_0/z'z}}. \tag{A19}$$

From (A16), we get that

$$\frac{x_1}{u_1\sqrt{u_2}} = \xi' W^{-1}\tilde{z} = \frac{e_1'W^{-1}z - (z_1/u_1)}{\sqrt{u_0}}, \tag{A20}$$

where x_1 is the first element of x, and hence

$$e_1'W^{-1}z = \frac{z_1}{u_1} + \frac{x_1\sqrt{u_0}}{u_1\sqrt{u_2}}. \tag{A21}$$

Using (A17), (A18), and (A21), we have

$$\hat{\theta}_t^2 = \frac{z_1^2 + u_0}{u_1}, \tag{A22}$$

$$\mu_{\mathrm{II},t} = \frac{k_3 g_1(\hat{\theta}_t^2)\sqrt{h}\theta}{\gamma u_1}\left(z_1 + \frac{x_1\sqrt{u_0}}{\sqrt{u_2}}\right), \tag{A23}$$

$$\sigma_{\mathrm{II},t}^2 = \frac{k_3^2 g_1^2(\hat{\theta}_t^2)h}{\gamma^2}\left(\frac{z_1^2 + u_0}{u_1^2}\right)\left(1 + \frac{x'x}{u_2}\right) = \frac{k_3^2 g_1^2(\hat{\theta}_t^2)h\hat{\theta}_t^2}{\gamma^2 u_1}\left(1 + \frac{x'x}{u_2}\right). \tag{A24}$$

We now provide a further simplification of the joint distribution of $\mu_{\mathrm{II},t}$ and $\sigma_{\mathrm{II},t}^2$ when $N > 1$. Define $p_1 = x_1/\sqrt{x'x}$ and $p_2 = z_2/\sqrt{u_0}$, where $z_2 \sim \mathcal{N}(0,1)$ is the second element of z. It is well known that p_1 is independent of $x'x$ and p_2 is independent of u_0 (see for example, Theorem 1.5.6 of Muirhead (1982)). As a result, p_1, p_2, u_0 and $x'x$ are independent of each other. Using the fact that p_1 and p_2 have the same distribution, we have

$$x_1\sqrt{u_0} = p_1\sqrt{x'x}\sqrt{u_0} \stackrel{d}{=} p_2\sqrt{x'x}\sqrt{u_0} = z_2\sqrt{x'x}, \tag{A25}$$

so $x_1\sqrt{u_0}$ has the same distribution as $z_2\sqrt{x'x}$. Note that replacing $x_1\sqrt{u_0}$ in $\mu_{\mathrm{II},t}$ by $z_2\sqrt{x'x}$ does not affect the joint distribution of $\mu_{\mathrm{II},t}$ and $\sigma_{\mathrm{II},t}^2$ because all the other terms in $\mu_{\mathrm{II},t}$ and $\sigma_{\mathrm{II},t}^2$ (i.e., z_1, $x'x$, u_0, u_1, and u_2) are independent of p_1 and p_2. As a result, we can write

$$\mu_{\mathrm{II},t} = \frac{k_3 g_1(\hat{\theta}_t^2)\sqrt{h}\theta}{\gamma u_1}\left(z_1 + \frac{z_2\sqrt{x'x}}{\sqrt{u_2}}\right), \tag{A26}$$

$$\sigma_{\mathrm{II},t}^2 = \frac{k_3^2 g_1^2(\hat{\theta}_t^2)h\hat{\theta}_t^2}{\gamma^2 u_1}\left(1 + \frac{x'x}{u_2}\right). \tag{A27}$$

Let

$$b=\frac{u_2}{x'x+u_2}\sim \text{Beta}\left(\frac{h-N+1}{2},\frac{N-1}{2}\right). \tag{A28}$$

Conditional on b, we have

$$z_0\equiv\frac{\sqrt{u_2}z_1+\sqrt{x'x}z_2}{\sqrt{x'x+u_2}}=\sqrt{b}z_1+\sqrt{1-b}z_2\sim\mathscr{N}(\sqrt{h}\theta\sqrt{b},1), \tag{A29}$$

and we can decompose $z'z$ as

$$z'z=z_0^2+\tilde{u}, \tag{A30}$$

where

$$\tilde{u}\sim\chi^2_{N-1}(h\theta^2(1-b)), \tag{A31}$$

and it is independent of z_0. Using these results, we can write

$$\hat{\theta}_t^2=\frac{z_0^2+\tilde{u}}{u_1},\qquad \mu_{\text{II},t}=\frac{k_3g_1(\hat{\theta}_t^2)\sqrt{h}\theta z_0}{\gamma u_1\sqrt{b}},\qquad \sigma^2_{\text{III},t}=\frac{k_3^2g_1^2(\hat{\theta}_t^2)h\hat{\theta}_t^2}{\gamma^2u_1b}, \tag{A32}$$

which are the expressions given in the Proposition. Applying Lemmas 1 and 2 to (A23) and using the fact that x_1 has zero mean and it is independent of z_1, u_0, u_1, u_2, we obtain

$$E[\mu_{\text{II},t}]=\frac{k_3\sqrt{h}\theta}{\gamma}E\left[g_1(\hat{\theta}_t^2)\frac{z_1}{u_1}\right]=\frac{k_3h\theta^2}{\gamma(h-N-2)}E[g_1(q_1)], \tag{A33}$$

where $q_1\sim\mathscr{G}^{h\theta^2}_{N+2,h-N-2}$. Given that x, z_1, u_1 and u_2 are independent of each other, $\hat{\theta}_t^2\sim\mathscr{G}^{h\theta^2}_{N,h-N}$, and applying Lemma 2 to (A24), we have

$$\begin{aligned}E[\sigma^2_{\text{II},t}]&=\frac{k_3^2h}{\gamma^2}E\left[g_1^2(\hat{\theta}_t^2)\frac{\hat{\theta}_t^2}{u_1}\right]E\left[1+\frac{x'x}{u_2}\right]\\&=\frac{k_3^2h}{\gamma^2}E\left[g_1^2(\hat{\theta}_t^2)\frac{\hat{\theta}_t^2}{u_1}\right]\frac{h-2}{h-N-1}\\&=\frac{k_3(h-N-4)}{\gamma^2(h-N-2)}E[g_1^2(q_2)q_2],\end{aligned} \tag{A34}$$

where $q_2\sim\mathscr{G}^{h\theta^2}_{N,h-N-2}$. With the above expressions of $E[\mu_{\text{II},t}]$ and $E[\sigma^2_{\text{II},t}]$, we obtain

$$E[U(\hat{w}_t^{\text{II}})]=\frac{k_3h\theta^2}{\gamma(h-N-2)}E[g_1(q_1)]-\frac{k_3(h-N-4)}{2\gamma(h-N-2)}E[g_1^2(q_2)q_2]. \tag{A35}$$

This completes the proof.

Proof of Proposition 2

Let $\hat{w}_{z,t}$ be the weights of a zero-investment portfolio

$$\hat{w}_{z,t} = \hat{\Sigma}_t^{-1}(\hat{\mu}_t - 1_N\hat{\mu}_{g,t}), \tag{A36}$$

and $\hat{w}_{g,t}$ be the weights of the sample global minimum-variance portfolio

$$\hat{w}_{g,t} = \frac{\hat{\Sigma}_t^{-1}1_N}{1_N'\hat{\Sigma}_t^{-1}1_N}. \tag{A37}$$

We can write

$$\hat{w}_t^{\mathrm{III}} = \frac{k_3}{\gamma}[g_2(\hat{\psi}_t^2)\hat{w}_{z,t} + 1_N'\hat{\Sigma}_t^{-1}\hat{\mu}_t\hat{w}_{g,t}]. \tag{A38}$$

Then the conditional mean and variance of portfolio $\hat{w}_t^{\mathrm{III}}$ can be written as

$$\mu_{\mathrm{III},t} = \frac{k_3}{\gamma}\left[g_2(\hat{\psi}_t^2)\mu_{z,t} + \left(1_N'\hat{\Sigma}_t^{-1}\hat{\mu}_t\right)\mu_{g,t}\right], \tag{A39}$$

$$\sigma^2_{\mathrm{III},t} = \frac{k_3^2}{\gamma^2}\left[g_2^2(\hat{\psi}_t^2)\sigma^2_{z,t} + \left(1_N'\hat{\Sigma}_t^{-1}\hat{\mu}_t\right)^2\sigma^2_{g,t} + 2g_2(\hat{\psi}_t^2)\left(1_N'\hat{\Sigma}_t^{-1}\hat{\mu}_t\right)\sigma_{gz,t}\right], \tag{A40}$$

where $\mu_{z,t} = \hat{w}_{z,t}'\mu$, $\mu_{g,t} = \hat{w}_{g,t}'\mu$, $\sigma^2_{z,t} = \hat{w}_{z,t}'\Sigma\hat{w}_{z,t}$, $\sigma^2_{g,t} = \hat{w}_{g,t}'\Sigma\hat{w}_{g,t}$ are the conditional means and variances of portfolios $\hat{w}_{z,t}$ and $\hat{w}_{g,t}$, and $\sigma_{gz,t} = \hat{w}_{g,t}'\Sigma\hat{w}_{z,t}$ is the conditional covariance between these two portfolios.

Let $P = [\nu,\ \eta,\ P_1]$ be an $N\times N$ orthonormal matrix with its first two columns as

$$\nu = \frac{\Sigma^{-\frac{1}{2}}1_N}{\left(1_N'\Sigma^{-1}1_N\right)^{\frac{1}{2}}} = \sigma_g\Sigma^{-\frac{1}{2}}1_N, \tag{A41}$$

$$\eta = \frac{(I_N - \nu\nu')\Sigma^{-\frac{1}{2}}\mu}{\left[\mu'\Sigma^{-\frac{1}{2}}(I_N - \nu\nu')\Sigma^{-\frac{1}{2}}\mu\right]^{\frac{1}{2}}} = \frac{\Sigma^{-\frac{1}{2}}(\mu - 1_N\mu_g)}{\psi}, \tag{A42}$$

where $\mu_g = 1_N'\Sigma^{-1}\mu/1_N'\Sigma^{-1}1_N$ and $\sigma_g^2 = 1/1_N'\Sigma^{-1}1_N$ are the mean and variance of the global minimum-variance portfolio. Define

$$z = \sqrt{h}P'\Sigma^{-\frac{1}{2}}\hat{\mu}_t \sim \mathcal{N}\left(\begin{bmatrix}\sqrt{h}\theta_g \\ \sqrt{h}\psi \\ 0_{N-2}\end{bmatrix}, I_N\right), \tag{A43}$$

$$W = hP'\Sigma^{-\frac{1}{2}}\hat{\Sigma}_t\Sigma^{-\frac{1}{2}}P \sim \mathscr{W}_N(h-1, I_N), \tag{A44}$$

and they are independent of each other. Let $z_1 \sim \mathcal{N}(\sqrt{h}\theta_g, 1)$ and $z_2 \sim \mathcal{N}(\sqrt{h}\psi, 1)$ be the first two elements of z, we can then write $z'z = z_1^2 + z_2^2 + u_0$, where $u_0 \sim \chi^2_{N-2}$ and it is independent of z_1 and z_2. With the definition of z and W, we can write

$$\hat{\psi}_t^2 = z'W^{-1}z - \frac{(e_1'W^{-1}z)^2}{e_1'W^{-1}e_1}, \tag{A45}$$

$$1_N'\hat{\Sigma}_t^{-1}\hat{\mu}_t = \frac{\sqrt{h}}{\sigma_g} e_1'W^{-1}z, \tag{A46}$$

$$\mu_{z,t} = \sqrt{h}\psi\left(e_2'W^{-1}z - \frac{e_1'W^{-1}e_2e_1'W^{-1}z}{e_1'W^{-1}e_1}\right), \tag{A47}$$

$$\mu_{g,t} = \mu_g + \sigma_g\psi\frac{e_1'W^{-1}e_2}{e_1'W^{-1}e_1}, \tag{A48}$$

$$\sigma_{z,t}^2 = hz'W^{-2}z + \frac{h(e_1'W^{-1}z)^2e_1'W^{-2}e_1}{(e_1'W^{-1}e_1)^2} - \frac{2h(e_1'W^{-1}z)(e_1'W^{-2}z)}{e_1'W^{-1}e_1}, \tag{A49}$$

$$\sigma_{g,t}^2 = \frac{\sigma_g^2 e_1'W^{-2}e_1}{(e_1'W^{-1}e_1)^2}, \tag{A50}$$

$$\sigma_{gz,t} = \sqrt{h}\sigma_g\left[\frac{e_1'W^{-2}z}{e_1'W^{-1}e_1} - \frac{e_1'W^{-2}e_1e_1'W^{-1}z}{(e_1'W^{-1}e_1)^2}\right]. \tag{A51}$$

where $e_1 = [1,\ 0_{N-1}']'$ and $e_2 = [0,\ 1,\ 0_{N-2}']'$. Define an $N \times N$ orthonormal matrix $Q = [e_1,\ \xi,\ Q_1]$ with its first two columns being e_1 and ξ, where

$$\xi = \frac{(I_N - e_1e_1')z}{\left[z'(I_N - e_1e_1')z\right]^{\frac{1}{2}}} = \frac{(I_N - e_1e_1')z}{\sqrt{z_2^2 + u_0}}. \tag{A52}$$

Let

$$A = (Q'W^{-1}Q)^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \sim \mathcal{W}_N(h-1, I_N), \tag{A53}$$

where A_{11} is the upper left 2×2 submatrix of A. Using Theorem 3.2.10 of Muirhead (1982), we have

$$A_{11\cdot 2} \equiv A_{11} - A_{12}A_{22}^{-1}A_{21} \sim \mathcal{W}_2(h-N+1, I_2), \tag{A54}$$

$$\mathrm{vec}(y) \equiv \mathrm{vec}(-A_{22}^{-\frac{1}{2}}A_{21}) \sim \mathcal{N}(0_{2N-4}, I_{2N-4}), \tag{A55}$$

$$A_{22} \sim \mathcal{W}_{N-2}(h-1, I_{N-2}), \tag{A56}$$

and they are independent of each other. Based on the Bartlett decomposition, we can write

$$A_{11\cdot 2} = \begin{bmatrix} v_1 + a^2 & -a\sqrt{v_2} \\ -a\sqrt{v_2} & v_2 \end{bmatrix}, \tag{A57}$$

where $v_1 \sim \chi^2_{h-N}$, $v_2 \sim \chi^2_{h-N+1}$, and $a \sim \mathcal{N}(0,1)$, all of which are independent of each other. Taking the inverse of $A_{11\cdot 2}$, we obtain

$$A^{-1}_{11\cdot 2} = \begin{bmatrix} \frac{1}{v_1} & \frac{a}{v_1\sqrt{v_2}} \\ \frac{a}{v_1\sqrt{v_2}} & \frac{1}{v_2} + \frac{a^2}{v_1 v_2} \end{bmatrix}. \tag{A58}$$

It follows that

$$e_1' W^{-1} e_1 = \frac{1}{v_1}, \tag{A59}$$

$$e_1' W^{-1} \xi = \frac{a}{v_1\sqrt{v_2}}, \tag{A60}$$

$$\xi' W^{-1} \xi = \frac{1}{v_2} + \frac{a^2}{v_1 v_2}. \tag{A61}$$

Using the definition of ξ, we obtain

$$e_1' W^{-1} \xi = \frac{e_1' W^{-1} z - e_1' W^{-1} e_1 z_1}{\sqrt{z_2^2 + u_0}} = \frac{a}{v_1\sqrt{v_2}}$$

$$\Rightarrow \quad e_1' W^{-1} z = \frac{a\sqrt{z_2^2 + u_0}}{v_1\sqrt{v_2}} + \frac{z_1}{v_1}, \tag{A62}$$

and

$$\xi' W^{-1} \xi = \frac{z'(I_N - e_1 e_1')W^{-1}(I_N - e_1 e_1')z}{z_2^2 + u_0} = \frac{1}{v_2} + \frac{a^2}{v_1 v_2}$$

$$\Rightarrow \quad z' W^{-1} z = \frac{z_2^2 + u_0}{v_2} + \frac{1}{v_1}\left(\frac{a\sqrt{z_2^2 + u_0}}{\sqrt{v_2}} + z_1\right)^2. \tag{A63}$$

Using (A59), (A62), and (A63), we get

$$\hat{\psi}_t^2 = z' W^{-1} z - \frac{(e_1' W^{-1} z)^2}{e_1' W^{-1} e_1} = \frac{z_2^2 + u_0}{v_2}, \tag{A64}$$

$$1_N' \hat{\Sigma}_t^{-1} \hat{\mu}_t = \frac{\sqrt{h}}{\sigma_g} e_1' W^{-1} z = \frac{\sqrt{h} y_3}{\sigma_g}, \tag{A65}$$

where

$$y_3 = e_1' W^{-1} z = \frac{a\hat{\psi}_t + z_1}{v_1}. \tag{A66}$$

Using the inverse of partitioned matrix formula, we obtain

$$Q_1'W^{-1}[e_1,\ \xi] = -A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1} = A_{22}^{-\frac{1}{2}}yA_{11\cdot 2}^{-1}. \tag{A67}$$

In addition, Theorem 3.1 and Corollary 3.1 in Dickey (1967) suggests that

$$A_{22}^{-\frac{1}{2}}y = xL^{-1}, \tag{A68}$$

where $x \equiv [x_1,\ x_2]$ is an $(N-2)\times 2$ matrix of independent standard normal random variables, L is a lower diagonal matrix such that $LL' \sim \mathscr{W}_2(h-N+3, I_2)$, and x and L are independent of each other. Using again the Bartlett decomposition, we can write

$$L = \begin{bmatrix} \sqrt{w_1} & 0 \\ -b & \sqrt{w_2} \end{bmatrix}, \tag{A69}$$

with $w_1 \sim \chi^2_{h-N+3}$, $w_2 \sim \chi^2_{h-N+2}$, and $b \sim \mathscr{N}(0,1)$, and they are independent of each other and $A_{11\cdot 2}$ (i.e., v_1, v_2, and a). Taking the inverse of L, we obtain

$$L^{-1} = \begin{bmatrix} \frac{1}{\sqrt{w_1}} & 0 \\ \frac{b}{\sqrt{w_1 w_2}} & \frac{1}{\sqrt{w_2}} \end{bmatrix}, \tag{A70}$$

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

$$\iota = \frac{(I_N - e_1e_1' - \xi\xi')e_2}{\left[e_2'(I_N - e_1e_1' - \xi\xi')e_2\right]^{\frac{1}{2}}} = \frac{(I_N - \xi\xi')e_2}{\sqrt{u_0/(z_2^2+u_0)}} = \frac{\sqrt{z_2^2+u_0}e_2 - z_2\xi}{\sqrt{u_0}}. \tag{A71}$$

Let $\varepsilon_1 = [1,\ 0_{N-3}']'$, we have

$$[h_1,\ h_2] \equiv \iota' W^{-1}[e_1,\ \xi] = \varepsilon_1' Q_1' W^{-1}[e_1,\ \xi] = \varepsilon_1' x L^{-1} A_{11\cdot 2}^{-1} = [x_{11},\ x_{21}]L^{-1}A_{11\cdot 2}^{-1}, \tag{A72}$$

with $x_{11} \sim \mathscr{N}(0,1)$ and $x_{21} \sim \mathscr{N}(0,1)$ being the first element of x_1 and x_2, respectively, and we can express h_1 and h_2 as

$$h_1 = \frac{1}{v_1}\left(\frac{x_{11}}{\sqrt{w_1}} + \frac{bx_{21}}{\sqrt{w_1w_2}} + \frac{ax_{21}}{\sqrt{v_2w_2}}\right), \tag{A73}$$

$$h_2 = \frac{a}{\sqrt{v_2}}h_1 + \frac{x_{21}}{v_2\sqrt{w_2}}. \tag{A74}$$

Using the definition of ι, (A59), (A60), (A61), and (A72), we have

$$e_2'W^{-1}e_1 = \frac{\sqrt{u_0}}{\sqrt{z_2^2+u_0}}h_1 + \frac{e_1'W^{-1}\xi}{\sqrt{z_2^2+u_0}}z_2 = \frac{\sqrt{u_0}}{\sqrt{z_2^2+u_0}}h_1 + \frac{az_2}{v_1\sqrt{v_2}\sqrt{z_2^2+u_0}}, \tag{A75}$$

$$e_2'W^{-1}z = \sqrt{u_0}h_2 + e_2'W^{-1}e_1z_1 + \xi'W^{-1}\xi z_2$$
$$= \frac{a\sqrt{u_0}}{\sqrt{v_2}}h_1 + \frac{x_{21}\sqrt{u_0}}{v_2\sqrt{w_2}} + e_2'W^{-1}e_1z_1 + \left(\frac{1}{v_2} + \frac{a^2}{v_1v_2}\right)z_2. \tag{A76}$$

Substituting these two expressions and (A59) in (A47) and (A48), we obtain

$$\mu_{z,t} = \frac{\sqrt{h}\psi}{v_2}\left(\frac{x_{21}\sqrt{u_0}}{\sqrt{w_2}} + z_2\right), \tag{A77}$$
$$\mu_{g,t} = \mu_g + \frac{\sigma_g\psi}{\sqrt{z_2^2+u_0}}\left(y_1\sqrt{u_0} + \frac{az_2}{\sqrt{v_2}}\right). \tag{A78}$$

where

$$y_1 = \frac{x_{11}}{\sqrt{w_1}} + \frac{bx_{21}}{\sqrt{w_1w_2}} + \frac{ax_{21}}{\sqrt{v_2w_2}}, \tag{A79}$$

In order to obtain those terms that involve W^{-2}, we first write

$$\begin{bmatrix} e_1'W^{-2}e_1 & e_1'W^{-2}\xi \\ e_1'W^{-2}\xi & \xi'W^{-2}\xi \end{bmatrix} = \begin{bmatrix} e_1' \\ \xi' \end{bmatrix} W^{-1}\left([e_1,\ \xi]\begin{bmatrix} e_1' \\ \xi' \end{bmatrix} + Q_1Q_1'\right)W^{-1}[e_1,\ \xi]$$
$$= A_{11\cdot 2}^{-2} + A_{11\cdot 2}^{-1}(L^{-1})'x'xL^{-1}A_{11\cdot 2}^{-1}. \tag{A80}$$

Note that $x'x$ can be written as

$$x'x = \begin{bmatrix} x_{11}^2 & x_{11}x_{21} \\ x_{11}x_{21} & x_{21}^2 \end{bmatrix} + C, \tag{A81}$$

where $C \sim \mathscr{W}_2(N-3, I_2)$, and it is independent of x_{11} and x_{21}. Using the Bartlett decomposition, we can write

$$C = \begin{bmatrix} s_1 + c^2 & c\sqrt{s_2} \\ c\sqrt{s_2} & s_2 \end{bmatrix}, \tag{A82}$$

where $s_1 \sim \chi^2_{N-4}$, $s_2 \sim \chi^2_{N-3}$, and $c \sim \mathscr{N}(0,1)$, and they are independent of each other.[16] Substituting (A58), (A70) and (A81) in (A80) and after simplification, we obtain

$$e_1'W^{-2}e_1 = \frac{1}{v_1^2}\left(y_1^2 + y_2^2 + 1 + \frac{s_1}{w_1} + \frac{a^2}{v_2}\right), \tag{A83}$$
$$\xi'W^{-2}\xi = \left(\frac{ay_1}{v_1\sqrt{v_2}} + \frac{x_{21}}{v_2\sqrt{w_2}}\right)^2 + \frac{a^2}{v_1^2v_2}\left(1 + \frac{s_1}{w_1}\right) + \left(\frac{ay_2}{v_1\sqrt{v_2}} + \frac{\sqrt{s_2}}{v_2\sqrt{w_2}}\right)^2$$

[16]Note that when $N = 3$, C is a zero matrix and we set $s_1 = 0$, $s_2 = 0$, and $c = 0$.

$$+\left(\frac{1}{v_2}+\frac{a^2}{v_1 v_2}\right)^2, \tag{A84}$$

$$e_1' W^{-2}\xi = \frac{a}{v_1^2\sqrt{v_2}}\left(y_1^2+y_2^2+1+\frac{s_1}{w_1}+\frac{a^2}{v_2}\right)+\frac{x_{21}y_1}{v_1 v_2\sqrt{w_2}}+\frac{\sqrt{s_2}y_2}{v_1 v_2\sqrt{w_2}}+\frac{a}{v_1 v_2^{\frac{3}{2}}}, \tag{A85}$$

where

$$y_2 = \frac{c}{\sqrt{w_1}}+\frac{b\sqrt{s_2}}{\sqrt{w_1 w_2}}+\frac{a\sqrt{s_2}}{\sqrt{v_2 w_2}}. \tag{A86}$$

With these expressions, we can write

$$\begin{aligned} e_1' W^{-2} z &= \sqrt{z_2^2+u_0}(e_1' W^{-2}\xi)+z_1(e_1' W^{-2} e_1) \\ &= \frac{y_3}{v_1}\left(y_1^2+y_2^2+1+\frac{s_1}{w_1}+\frac{a^2}{v_2}\right)+\frac{\hat{\psi}_t}{v_1\sqrt{v_2}}\left(\frac{x_{21}y_1+\sqrt{s_2}y_2}{\sqrt{w_2}}+\frac{a}{\sqrt{v_2}}\right), \end{aligned} \tag{A87}$$

$$\begin{aligned} z' W^{-2} z &= (z_2^2+u_0)(\xi' W^{-2}\xi)+2z_1(e_1' W^{-2} z)-z_1^2(e_1' W^{-2} e_1) \\ &= \left(1+\frac{s_1}{w_1}\right)y_3^2+\left(\frac{a y_3+\hat{\psi}_t}{\sqrt{v_2}}\right)^2+\left(y_1 y_3+\frac{x_{21}\hat{\psi}_t}{\sqrt{v_2 w_2}}\right)^2+\left(y_2 y_3+\frac{\sqrt{s_2}\hat{\psi}_t}{\sqrt{v_2 w_2}}\right)^2. \end{aligned} \tag{A88}$$

We then obtain

$$\sigma_{g,t}^2 = \sigma_g^2\left(y_1^2+y_2^2+1+\frac{s_1}{w_1}+\frac{a^2}{v_2}\right), \tag{A89}$$

$$\sigma_{z,t}^2 = \frac{h\hat{\psi}_t^2}{v_2}\left(1+\frac{x_{21}^2+s_2}{w_2}\right), \tag{A90}$$

$$\sigma_{gz,t} = \frac{\sqrt{h}\sigma_g\hat{\psi}_t}{\sqrt{v_2}}\left(\frac{a}{\sqrt{v_2}}+\frac{x_{21}}{\sqrt{w_2}}y_1+\frac{\sqrt{s_2}}{\sqrt{w_2}}y_2\right). \tag{A91}$$

Substituting (A65), (A77), (A78), (A89), (A90), (A91) in (A39) and (A40) and after simplification, we obtain

$$\mu_{\mathrm{III},t} = \frac{\sqrt{h}k_3}{\gamma}\left[\frac{g_2(\hat{\psi}_t^2)\psi}{v_2}\left(\frac{x_{21}\sqrt{u_0}}{\sqrt{w_2}}+z_2\right)+\theta_g y_3+\frac{\psi y_3}{\hat{\psi}_t}\left(\frac{\sqrt{u_0}y_1}{\sqrt{v_2}}+\frac{a z_2}{v_2}\right)\right], \tag{A92}$$

$$\begin{aligned} \sigma_{\mathrm{III},t}^2 = \frac{hk_3^2}{\gamma^2}&\left[\left(1+\frac{s_1}{w_1}\right)y_3^2+\left(\frac{a y_3+g_2(\hat{\psi}_t^2)\hat{\psi}_t}{\sqrt{v_2}}\right)^2+\left(y_1 y_3+\frac{x_{21}g_2(\hat{\psi}_t^2)\hat{\psi}_t}{\sqrt{v_2 w_2}}\right)^2\right. \\ &\left.+\left(y_2 y_3+\frac{\sqrt{s_2}g_2(\hat{\psi}_t^2)\hat{\psi}_t}{\sqrt{v_2 w_2}}\right)^2\right]. \end{aligned} \tag{A93}$$

Taking expectations and applying Lemmas 1 and 2 by using the fact that $\hat{\psi}_t^2 \sim \mathscr{G}_{N-1,h-N+1}^{h\psi^2}$, we obtain

$$E[\mu_{\mathrm{III},t}] = \frac{k_3\sqrt{h}\psi}{\gamma}E\left[\frac{g_2(\hat{\psi}_t^2)z_2}{v_2}\right]+\frac{k_3 h}{\gamma(h-N-2)}\left(\theta_g^2+\frac{\psi^2}{h-N-1}\right)$$

$$= \frac{h\psi^2 k_3}{\gamma(h-N-1)} E[g_2(q_3)] + \frac{k_3 h}{\gamma(h-N-2)} \left(\theta_g^2 + \frac{\psi^2}{h-N-1} \right), \tag{A94}$$

$$\begin{aligned} E[\sigma^2_{\mathrm{III},t}] &= \frac{k_3}{\gamma^2} \left[\frac{h\theta_g^2}{h-N-2} + \frac{h-4+h\psi^2}{(h-N-2)(h-N-3)} \right. \\ &\quad \left. + \frac{(h-N-1)(h-N-4)}{h-N} E\left[\frac{2g_2(\hat{\psi}_t^2)\hat{\psi}_t^2}{(h-N-2)v_2} + \frac{g_2^2(\hat{\psi}_t^2)\hat{\psi}_t^2}{v_2} \right] \right] \\ &= \frac{k_3}{\gamma^2} \left[\frac{h\theta_g^2}{h-N-2} + \frac{h-4+h\psi^2}{(h-N-2)(h-N-3)} \right. \\ &\quad \left. + \frac{(h-N-4)}{h-N} E\left[\left(\frac{2g_2(q_4)}{h-N-2} + g_2^2(q_4) \right) q_4 \right] \right], \end{aligned} \tag{A95}$$

where $q_3 \sim \mathcal{G}^{h\psi^2}_{N+1,h-N-1}$ and $q_4 \sim \mathcal{G}^{h\psi^2}_{N-1,h-N-1}$. With the above expressions of $E[\mu_{\mathrm{III},t}]$ and $E[\sigma^2_{\mathrm{III},t}]$, we can get

$$\begin{aligned} E[U(\hat{w}_t^{\mathrm{III}})] = {} & \frac{k_3}{(h-N-2)\gamma} \left[\frac{h\theta_g^2}{2} + \frac{h\psi^2}{h-N-1} - \frac{h-4+h\psi^2}{2(h-N-3)} \right] + \frac{h\psi^2 k_3}{(h-N-1)\gamma} E[g_2(q_3)] \\ & - \frac{k_3(h-N-4)}{2(h-N)\gamma} E\left[\left(\frac{2g_2(q_4)}{h-N-2} + g_2^2(q_4) \right) q_4 \right]. \end{aligned} \tag{A96}$$

This completes the proof.

Proof of Proposition 3

Note that

$$\mu_{p,t} = \mu_{g,t} + \frac{1}{\gamma}\mu_{z,t}, \tag{A97}$$

$$\sigma^2_{p,t} = \sigma^2_{g,t} + \frac{2}{\gamma}\sigma_{gz,t} + \frac{1}{\gamma^2}\sigma^2_{z,t}, \tag{A98}$$

where $\mu_{g,t}$, $\mu_{z,t}$, $\sigma^2_{g,t}$, $\sigma^2_{z,t}$ are the conditional mean and variance of portfolios $\hat{w}_{g,t}$ and $\hat{w}_{z,t}$, and $\sigma_{gz,t}$ is the conditional covariance of the two portfolios. The expressions of these conditional mean, variance, and covariance are given in (A77), (A78), (A89), (A90), and (A91). Taking their expectations, we obtain

$$E[\mu_{g,t}] = \mu_g, \tag{A99}$$

$$E[\mu_{z,t}] = \frac{h\psi^2}{h-N-1}, \tag{A100}$$

$$E[\sigma^2_{g,t}] = \frac{(h-2)\sigma_g^2}{h-N-1}, \tag{A101}$$

$$E[\sigma^2_{z,t}] = \frac{h(h-2)(h\psi^2+N-1)}{(h-N)(h-N-1)(h-N-3)}, \tag{A102}$$

$$E[\sigma_{gz,t}] = 0. \tag{A103}$$

It follows that

$$E[\mu_{p,t}] = \mu_g + \frac{h\psi^2}{\gamma(h-N-1)}, \tag{A104}$$

$$E[\sigma^2_{p,t}] = \frac{(h-2)\sigma^2_g}{h-N-1} + \frac{h(h-2)(h\psi^2+N-1)}{\gamma^2(h-N)(h-N-1)(h-N-3)}, \tag{A105}$$

which gives the expression of the expected out-of-sample utility of portfolio p. This completes the proof.

Proof of Proposition 4

Note that

$$\mu_{q,t} = \mu_{g,t} + \frac{\tilde{k}_3 g_3(\hat{\psi}^2_t)}{\gamma}\mu_{z,t}, \tag{A106}$$

$$\sigma^2_{q,t} = \sigma^2_{g,t} + \frac{\tilde{k}^2_3 g^2_3(\hat{\psi}^2_t)}{\gamma^2}\sigma^2_{z,t} + \frac{2\tilde{k}_3 g_3(\hat{\psi}^2_t)}{\gamma}\sigma_{gz,t}, \tag{A107}$$

where $\mu_{g,t}$, $\mu_{z,t}$, $\sigma^2_{g,t}$, $\sigma^2_{z,t}$ are the conditional mean and variance of portfolios $\hat{w}_{g,t}$ and $\hat{w}_{z,t}$, and $\sigma_{gz,t}$ is the conditional covariance of the two portfolios. The expressions of these conditional mean, variance, and covariance are given in (A77), (A78), (A89), (A90), and (A91). Plugging these expressions in (A106) and (A107), we obtain the expressions of $\mu_{q,t}$ and $\sigma^2_{q,t}$ in the Proposition. Take expectations and applying Lemmas 1 and 2, we get

$$E[\mu_{q,t}] = \mu_g + \frac{\sqrt{h}\psi\tilde{k}_3}{\gamma}E\left[\frac{g_3(\hat{\psi}^2_t)z_2}{v_2}\right] = \mu_g + \frac{h\psi^2\tilde{k}_3}{\gamma(h-N-1)}E[g_3(q_3)], \tag{A108}$$

$$E[\sigma^2_{q,t}] = \frac{(h-2)\sigma^2_g}{h-N-1} + \frac{\tilde{k}^2_3 h(h-2)}{\gamma(h-N)}E\left[\frac{g^2_3(\hat{\psi}^2_t)\hat{\psi}^2_t}{v_2}\right] = \frac{(h-2)\sigma^2_g}{h-N-1} + \frac{\tilde{k}_3(h-N-3)E[g^2_3(q_4)q_4]}{\gamma(h-N-1)}, \tag{A109}$$

where $q_3 \sim \mathscr{G}^{h\psi^2}_{N+1,h-N-1}$, $q_4 \sim \mathscr{G}^{h\psi^2}_{N-1,h-N-1}$. Using these expressions, we obtain the expected out-of-sample utility of portfolio q. This completes the proof.

REFERENCES

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

Black, Fischer, and Robert Litterman, 1992, Global portfolio optimization, *Financial Analysts Journal* 48, 28–43.

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

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

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

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.

Grinold, Richard C., and Ronald N. Kahn, 1999, *Active Portfolio Management: Quantitative Theory and Applications* (McGraw-Hill, New York).

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

Kan, Raymond, and Daniel Smith, 2008, The distribution of the sample minimum-variance frontier, *Management Science* 54, 1364–1380.

Kan, Raymond, and Xiaolu Wang, 2015, On the economic value of alphas, working paper, University of Toronto.

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

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

Litterman, Bob, 2003, *Modern Investment Management: An Equilibrium Approach* (Wiley, New York).

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

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

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

Okhrin, Yarema, and Wolfgang Schmid, 2006, Distributional properties of portfolio weights, *Journal of Econometrics* 134, 235–256.

Qian, Edward, Ronald Hua, and Eric Sorensen, 2007, *Quantitative Equity Portfolio Management: Modern Techniques and Applications* (Chapman & Hall, New York).

Sargan, John D., 1976, The existence of the moments of estimated reduced form coefficients, Discussion Paper A6, London School of Economics Econometrics Program.

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

Table 1: Portfolio performance comparison for the case with risk-free asset using empirical data over 1963/7–2004/11 (with normalization)

This table reports the empirical utility (in percentage points) and the Sharpe ratio of the *ex post* tangency portfolio (MV in-sample), the equally weighted portfolio of risky assets ($1/N$), the normalized ML rule (ML), and the normalized three-fund rule of Kan and Zhou (2007) (KZ3) for $h = 120$ and $\gamma = 1$, using the same six datasets as in DeMiguel, Garlappi, and Uppal (2009). One sided p-values of the performance difference between the optimal portfolios and the $1/N$ rule are also reported.

	Industry $N = 11$	International $N = 9$	MKT/SMB/HML $N = 3$	FF+1-factor $N = 21$	FF+3-factor $N = 23$	FF+4-factor $N = 24$
A. CEQ						
MV (in-sample)	0.0098	0.0082	0.0044	0.0296	−0.2075	−0.0752
$1/N$	0.0050	0.0051	0.0042	0.0073	0.0071	0.0072
ML	−0.1191	−0.1344	0.0041	−0.6426	−0.0034	−0.0157
p-value	1.0000	1.0000	0.5448	1.0000	0.7692	0.8749
KZ3	−0.0508	−0.0013	0.0041	−0.0567	0.0143	−0.0025
p-value	0.9996	0.9770	0.5390	0.9998	0.0565	0.7429
B. Sharpe ratio						
MV (in-sample)	0.2146	0.1969	0.2770	0.5231	0.5260	0.5591
$1/N$	0.1365	0.1375	0.2351	0.1628	0.1683	0.1761
ML	−0.0152	−0.0372	0.2034	−0.0227	0.1240	0.1443
p-value	0.9757	0.9886	0.7254	0.9970	0.7393	0.6707
KZ3	−0.0093	0.0142	0.2420	0.0089	0.2218	0.1315
p-value	0.9722	0.9893	0.4349	0.9892	0.2202	0.7304

Table 2: CEQ comparison for the case with risk-free asset using empirical data over 1927/1–2014/12 (with normalization)

This table reports the empirical utility (in percentage points) of the *ex post* tangency portfolio (MV in-sample), the equally weighted portfolio of risky assets ($1/N$), the normalized ML rule, and the normalized three-fund rule of Kan and Zhou (2007) (KZ3) for different combinations of h (120 or 240) and γ (1 or 3), using the same six datasets as in DeMiguel, Garlappi, and Uppal (2009) but over an extended sample period. One sided p-values of the performance difference between the optimal portfolios and the $1/N$ rule are also reported.

	Industry $N=11$	International $N=9$	MKT/SMB/HML $N=3$	FF+1-factor $N=21$	FF+3-factor $N=23$	FF+4-factor $N=24$
A. $h=120$ and $\gamma=1$						
MV (in-sample)	0.0092	0.0066	0.0043	0.0191	−0.0237	−0.0316
$1/N$	0.0059	0.0047	0.0038	0.0070	0.0067	0.0068
ML	−0.2304	−3.6631	0.0027	−3.0487	0.0016	−0.0028
p-value	1.0000	1.0000	0.7061	1.0000	0.8058	0.8811
KZ3	−0.0584	−0.2733	0.0029	−0.4240	0.0073	0.0009
p-value	1.0000	1.0000	0.6982	1.0000	0.3978	0.8300
B. $h=120$ and $\gamma=3$						
MV (in-sample)	0.0073	0.0046	0.0038	0.0152	−0.0278	−0.0371
$1/N$	0.0040	0.0024	0.0033	0.0036	0.0038	0.0041
ML	−0.7309	−11.0738	−0.0013	−9.0334	−0.0293	−0.0604
p-value	1.0000	1.0000	0.9893	1.0000	1.0000	1.0000
KZ3	−0.1999	−0.8111	0.0001	-1.2344	0.0046	−0.0314
p-value	1.0000	1.0000	0.9733	1.0000	0.3678	1.0000
C. $h=240$ and $\gamma=1$						
MV (in-sample)	0.0091	0.0129	0.0041	0.0197	−0.0109	−0.0160
$1/N$	0.0060	0.0034	0.0036	0.0069	0.0066	0.0067
ML	0.0062	0.0013	0.0038	−0.0561	−0.8185	−0.2734
p-value	0.4715	0.8254	0.3877	1.0000	1.0000	1.0000
KZ3	0.0063	0.0032	0.0033	0.0128	−0.5827	−0.1777
p-value	0.4336	0.5804	0.7679	0.0481	1.0000	1.0000
D. $h=240$ and $\gamma=3$						
MV (in-sample)	0.0074	0.0089	0.0038	0.0166	−0.0116	−0.0172
$1/N$	0.0044	0.0011	0.0032	0.0043	0.0044	0.0047
ML	0.0004	−0.0017	0.0033	−0.2377	−2.5849	−0.9454
p-value	0.9148	0.8836	0.4843	1.0000	1.0000	1.0000
KZ3	0.0038	0.0014	0.0029	0.0031	−1.8555	−0.6327
p-value	0.6172	0.3877	0.7733	0.6291	1.0000	1.0000

Table 3: Sharpe ratio comparison for the case with risk-free asset using empirical data over 1927/1–2014/12 (with normalization)

This table reports the Sharpe ratio of the *ex post* tangency portfolio (MV in-sample), the equally weighted portfolio of risky assets ($1/N$), the normalized ML rule, and the normalized three-fund rule (KZ3) of Kan and Zhou (2007) for different combinations of h (120 or 240), using the same six datasets as in DeMiguel, Garlappi, and Uppal (2009) but over an extended sample period. One sided p-values of the performance difference between the optimal portfolios and the $1/N$ rule are also reported.

	Industry $N=11$	International $N=9$	MKT/SMB/HML $N=3$	FF+1-factor $N=21$	FF+3-factor $N=23$	FF+4-factor $N=24$
A. $h=120$						
MV (in-sample)	0.2332	0.1696	0.1982	0.3379	0.3401	0.3908
$1/N$	0.1573	0.1222	0.1791	0.1486	0.1506	0.1569
ML	0.0281	0.0155	0.0741	−0.0230	0.0970	0.1085
p-value	0.9956	0.9464	0.9966	0.9999	0.8860	0.8554
KZ3	0.0328	−0.0060	0.0818	−0.0208	0.1656	0.0950
p-value	0.9950	0.9767	0.9958	0.9999	0.3683	0.9107
B. $h=240$						
MV (in-sample)	0.2411	0.2354	0.2284	0.3827	0.3886	0.4433
$1/N$	0.1686	0.0950	0.2009	0.1612	0.1638	0.1714
ML	0.1197	0.0502	0.1806	0.0815	0.0487	0.0763
p-value	0.8556	0.8412	0.7700	0.9472	0.9888	0.9712
KZ3	0.1513	0.0967	0.1818	0.1796	0.0476	0.0738
p-value	0.6652	0.4708	0.7894	0.3404	0.9894	0.9742

Table 4: CEQ comparison for the case with risk-free asset using empirical data over 1927/1–2014/12 (without normalization)

This table reports the empirical utility (in percentage points) of the *ex post* optimal portfolio (MV in-sample), the $1/N$ rule constructed based on (39), the ML rule, the two-fund rule (KZ2), and the three-fund rule (KZ3) of Kan and Zhou (2007) for different combinations of h (120 or 240) and γ (1 or 3), using the same six datasets as in DeMiguel, Garlappi, and Uppal (2009) but over an extended sample period. One sided p-values of the performance difference between the optimal portfolios and the $1/N$ rule are also reported.

	Industry $N=11$	International $N=9$	MKT/SMB/HML $N=3$	FF+1-factor $N=21$	FF+3-factor $N=23$	FF+4-factor $N=24$
A. $h=120$ and $\gamma=1$						
MV (in-sample)	0.0272	0.0144	0.0196	0.0571	0.0578	0.0764
$1/N$	0.0116	0.0031	0.0130	0.0080	0.0081	0.0090
ML	−0.0516	−0.0440	0.0065	−0.1721	−0.2158	−0.2526
p-value	1.0000	0.9996	0.8183	1.0000	1.0000	1.0000
KZ2	0.0043	−0.0060	0.0124	0.0281	0.0264	0.0350
p-value	0.8471	0.9635	0.5448	0.0647	0.0849	0.0415
KZ3	0.0073	−0.0060	0.0150	0.0304	0.0209	0.0313
p-value	0.7113	0.9520	0.3421	0.0492	0.1838	0.0768
B. $h=120$ and $\gamma=3$						
MV (in-sample)	0.0091	0.0048	0.0065	0.0190	0.0193	0.0255
$1/N$	0.0039	0.0010	0.0043	0.0027	0.0027	0.0030
ML	−0.0172	−0.0147	0.0022	−0.0574	−0.0719	−0.0842
p-value	1.0000	0.9996	0.8183	1.0000	1.0000	1.0000
KZ2	0.0014	−0.0020	0.0041	0.0094	0.0088	0.0117
p-value	0.8471	0.9635	0.5448	0.0647	0.0849	0.0415
KZ3	0.0024	−0.0020	0.0050	0.0101	0.0070	0.0104
p-value	0.7113	0.9520	0.3421	0.0492	0.1838	0.0768
C. $h=240$ and $\gamma=1$						
MV (in-sample)	0.0291	0.0277	0.0261	0.0732	0.0755	0.0982
$1/N$	0.0043	0.0007	0.0127	0.0050	0.0055	0.0065
ML	−0.0158	−0.0138	0.0093	−0.0333	−0.0430	−0.0354
p-value	0.9464	0.9638	0.7070	0.9339	0.9661	0.9089
KZ2	0.0044	−0.0030	0.0119	0.0231	0.0230	0.0402
p-value	0.4953	0.7894	0.5654	0.1153	0.1221	0.0292
KZ3	0.0052	−0.0018	0.0126	0.0248	0.0209	0.0385
p-value	0.4553	0.7356	0.5105	0.0878	0.1581	0.0382
D. $h=240$ and $\gamma=3$						
MV (in-sample)	0.0097	0.0092	0.0087	0.0244	0.0252	0.0327
$1/N$	0.0014	0.0002	0.0042	0.0017	0.0018	0.0022
ML	−0.0053	−0.0046	0.0031	−0.0111	−0.0143	−0.0118
p-value	0.9464	0.9638	0.7070	0.9339	0.9661	0.9089
KZ2	0.0015	−0.0010	0.0040	0.0077	0.0077	0.0134
p-value	0.4953	0.7894	0.5654	0.1153	0.1221	0.0292
KZ3	0.0017	−0.0006	0.0042	0.0083	0.0070	0.0128
p-value	0.4553	0.7356	0.5105	0.0878	0.1581	0.0382

Table 5: Sharpe ratio comparison for the case with risk-free asset using empirical data over 1927/1–2014/12 (without normalization)

This table reports the Sharpe ratio of the *ex post* optimal portfolio (MV in-sample), the $1/N$ rule constructed based on (39), the ML rule, the two-fund rule (KZ2) and the three-fund rule (KZ3) of Kan and Zhou (2007) for different combinations of h (120 or 240), using the same six datasets as in DeMiguel, Garlappi, and Uppal (2009) but over an extended sample period. One sided p-values of the performance difference between the optimal portfolios and the $1/N$ rule are also reported.

	Industry $N = 11$	International $N = 9$	MKT/SMB/HML $N = 3$	FF+1-factor $N = 21$	FF+3-factor $N = 23$	FF+4-factor $N = 24$
A. $h = 120$						
MV (in-sample)	0.2332	0.1696	0.1982	0.3379	0.3401	0.3908
$1/N$	0.1534	0.0812	0.1623	0.1277	0.1292	0.1359
ML	0.1371	0.0161	0.1738	0.2735	0.2684	0.3154
p-value	0.6793	0.9074	0.3159	0.0002	0.0004	0.0000
KZ2	0.1353	−0.0118	0.1637	0.2677	0.2629	0.2995
p-value	0.7049	0.9657	0.4759	0.0004	0.0008	0.0001
KZ3	0.1669	0.0491	0.1851	0.2843	0.2561	0.2961
p-value	0.3269	0.8369	0.1430	0.0000	0.0016	0.0001
B. $h = 240$						
MV (in-sample)	0.2411	0.2354	0.2284	0.3827	0.3886	0.4433
$1/N$	0.1098	0.0611	0.1619	0.1096	0.1132	0.1214
ML	0.1448	0.0228	0.1648	0.2835	0.2822	0.3490
p-value	0.1806	0.8047	0.4550	0.0001	0.0001	0.0000
KZ2	0.1319	−0.0096	0.1596	0.2611	0.2606	0.3227
p-value	0.2798	0.9408	0.5340	0.0005	0.0007	0.0000
KZ3	0.1514	0.0631	0.1688	0.2699	0.2576	0.3209
p-value	0.1050	0.4673	0.3828	0.0000	0.0009	0.0000

Table 6: CEQ comparison for the case without risk-free asset using empirical data over 1927/1–2014/12

This table reports the empirical utility (in percentage points) for the *ex post* optimal portfolio, the equally weighted portfolio of risky assets ($1/N$), the ML rule, and the QL rule for the case without a risk-free asset for different combinations of h (120 or 240) and γ (1 or 3), using the same six datasets as in DeMiguel, Garlappi, and Uppal (2009) but over an extended sample period. One sided p-values of the performance difference between the optimal portfolios and the $1/N$ rule are also reported.

	Industry $N=11$	International $N=9$	MKT/SMB/HML $N=3$	FF+1-factor $N=21$	FF+3-factor $N=23$	FF+4-factor $N=24$
A. $h=120$ and $\gamma=1$						
MV (in-sample)	0.0159	0.0079	0.0083	0.0442	0.0576	0.0761
$1/N$	0.0059	0.0047	0.0038	0.0070	0.0067	0.0068
ML	−0.0588	−0.0394	−0.0020	−0.1284	−0.1876	-0.2251
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
QL	−0.0028	−0.0023	0.0063	0.0216	0.0268	0.0361
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
B. $h=120$ and $\gamma=3$						
MV (in-sample)	0.0080	0.0046	0.0045	0.0178	0.0191	0.0253
$1/N$	0.0040	0.0024	0.0033	0.0036	0.0038	0.0041
ML	−0.0173	−0.0114	0.0010	−0.0392	−0.0628	−0.0753
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
QL	0.0014	0.0010	0.0037	0.0104	0.0088	0.0118
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
C. $h=240$ and $\gamma=1$						
MV (in-sample)	0.0164	0.0220	0.0095	0.0523	0.0743	0.0973
$1/N$	0.0060	0.0034	0.0036	0.0069	0.0066	0.0067
ML	−0.0154	−0.0113	−0.0029	−0.0282	−0.0356	−0.0245
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
QL	0.0057	0.0013	−0.0000	0.0205	0.0232	0.0422
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
D. $h=240$ and $\gamma=3$						
MV (in-sample)	0.0082	0.0091	0.0050	0.0214	0.0247	0.0323
$1/N$	0.0044	0.0011	0.0032	0.0043	0.0044	0.0047
ML	−0.0022	−0.0026	0.0006	−0.0047	−0.0121	-0.0084
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
QL	0.0046	0.0013	0.0016	0.0112	0.0076	0.0139
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

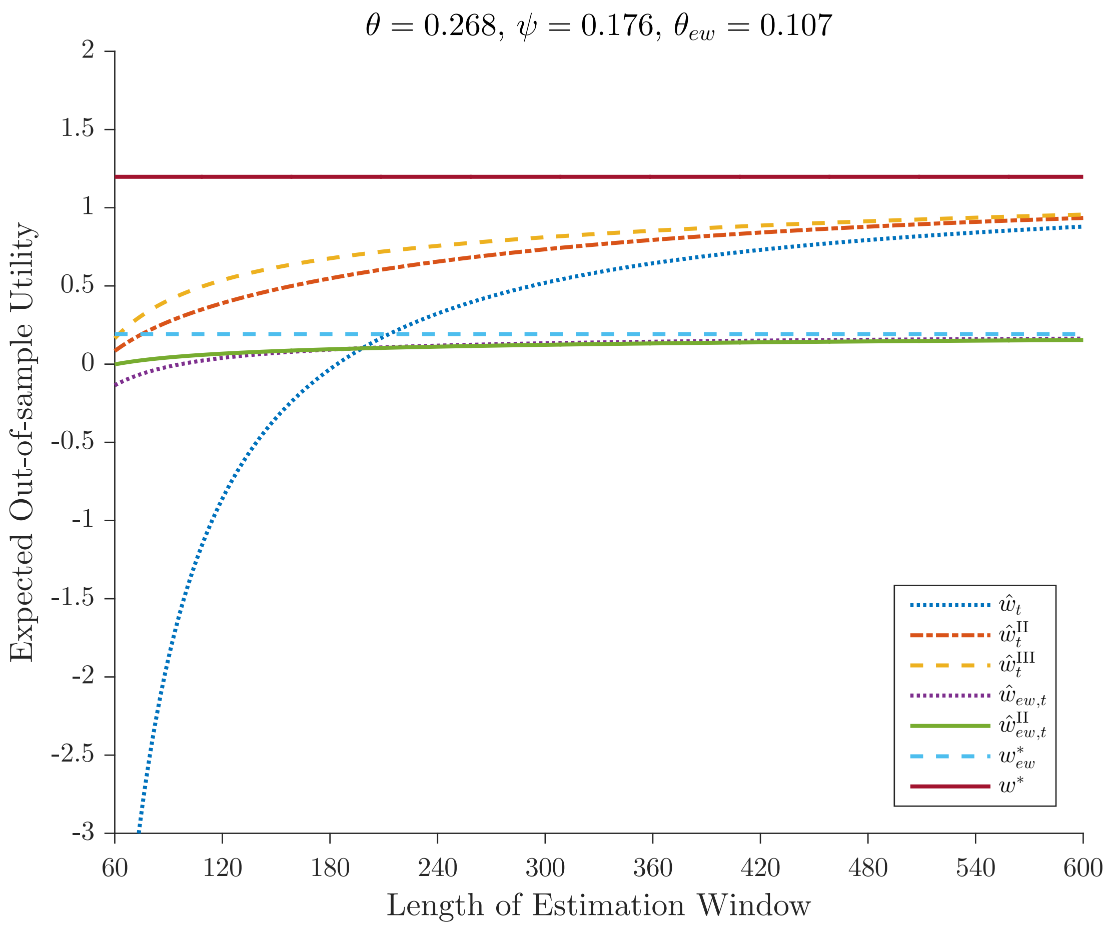


Figure 1: **Expected out-of-sample utility of various portfolio rules for the case with a risk-free asset and 10 risky assets**

This figure plots the expected out-of-sample utilities (in percentage points) of portfolios $\hat{w}_t$, $\hat{w}_t^{\mathrm{II}}$, $\hat{w}_t^{\mathrm{III}}$, $\hat{w}_{ew,t}$, $\hat{w}_{ew,t}^{\mathrm{II}}$, w_{ew}^*, and w^* as a function of the length of estimation window (h), with parameters estimated using excess monthly returns of the 10 momentum portfolios over the period of 1927/1–2014/12. The risk aversion coefficient is set to three ($\gamma = 3$).

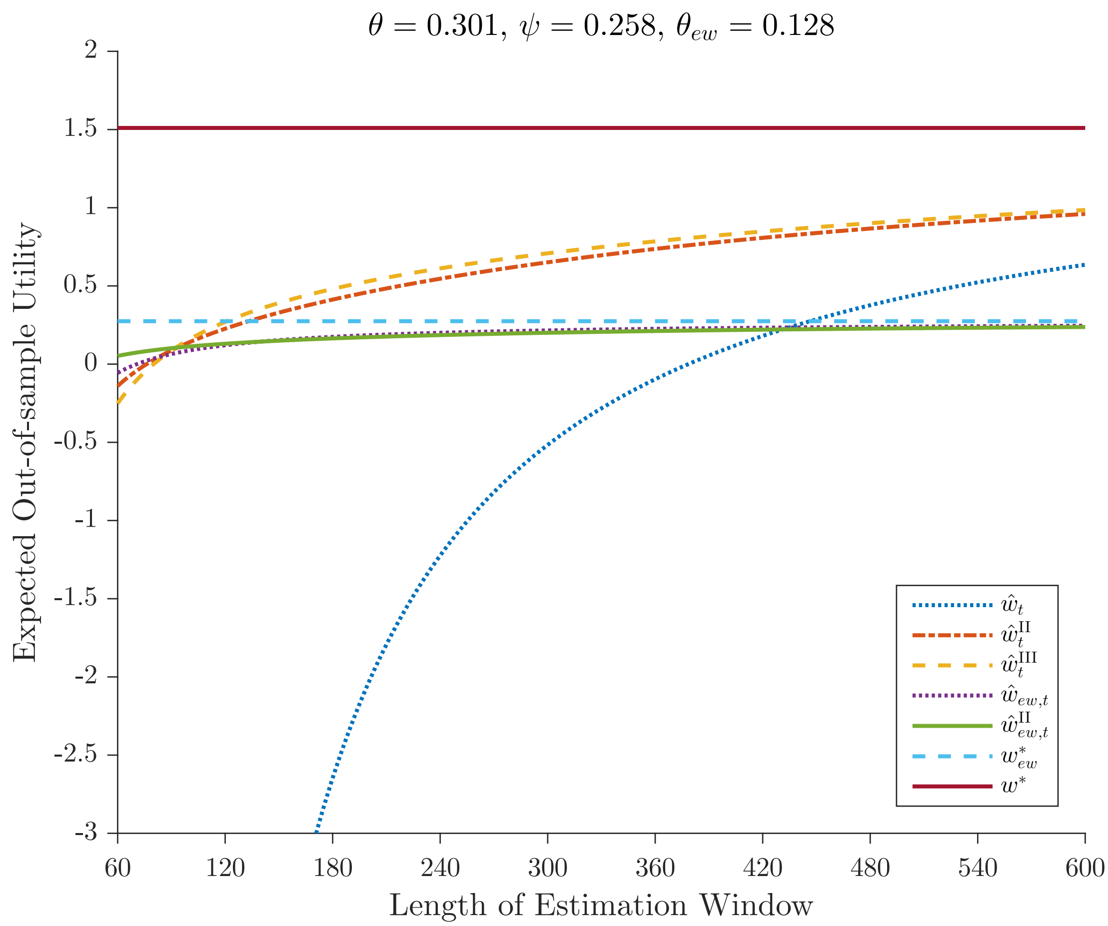
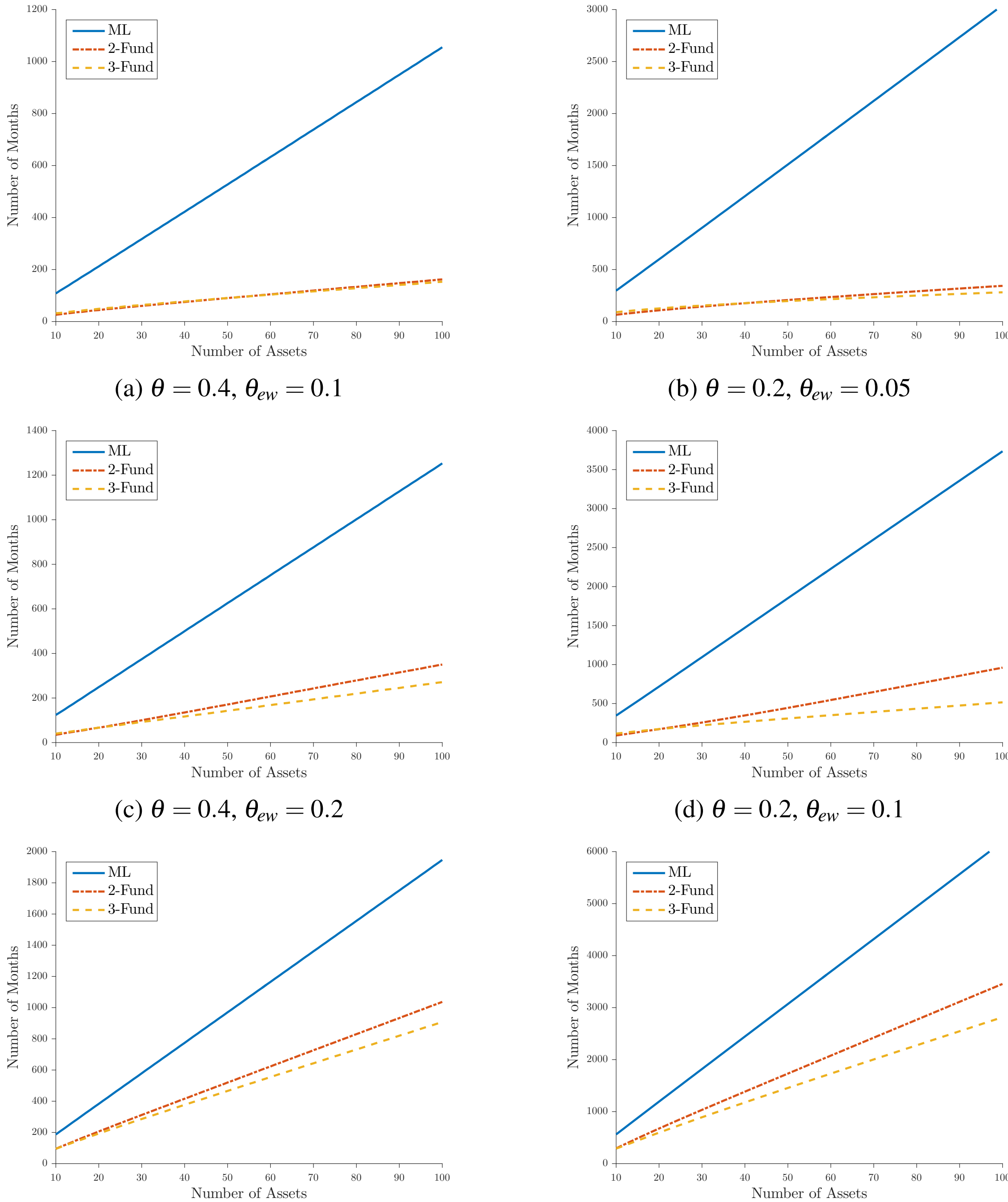


Figure 2: **Expected out-of-sample utility of various portfolio rules for the case with a risk-free asset and 25 risky assets**

This figure plots the expected out-of-sample utilities (in percentage points) of portfolios $\hat{w}_t$, $\hat{w}_t^{\mathrm{II}}$, $\hat{w}_t^{\mathrm{III}}$, $\hat{w}_{ew,t}$, $\hat{w}_{ew,t}^{\mathrm{II}}$, w_{ew}^*, and w^* as a function of the length of estimation window (h), with parameters estimated using excess monthly returns of the Fama-French 5×5 size and book-to-market ranked portfolios over the period of 1927/1–2014/12. The risk aversion coefficient is set to three ($\gamma = 3$).



(a) $\theta = 0.4$, $\theta_{ew} = 0.1$

(b) $\theta = 0.2$, $\theta_{ew} = 0.05$

(c) $\theta = 0.4$, $\theta_{ew} = 0.2$

(d) $\theta = 0.2$, $\theta_{ew} = 0.1$

(e) $\theta = 0.4$, $\theta_{ew} = 0.3$

(f) $\theta = 0.2$, $\theta_{ew} = 0.15$

Figure 3: **Number of estimation months required to outperform the $1/N$ rule**
This figure plots the required number of estimation months for various optimal portfolio rules (ML, two-fund, and three-fund) to outperform the $1/N$ rule as a function of the number of risky assets (N) when a risk-free asset is included in the optimal portfolios. The six panels report results for different combinations of θ (0.4 or 0.2) and θ_{ew}/θ (0.25, 0.5, or 0.75), with the Sharpe ratio of the global minimum-variance portfolio (θ_g) set equal to $\theta/2$ in each case.

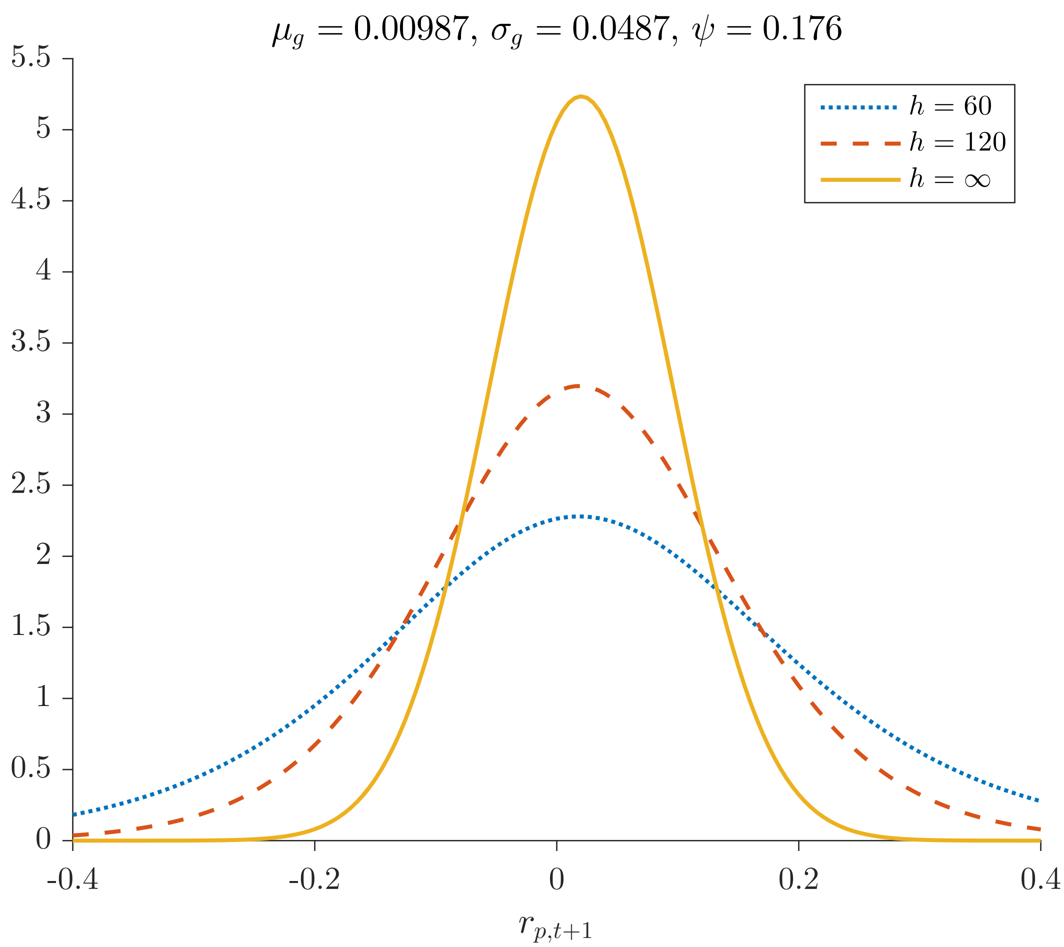


Figure 4: **Unconditional distribution of out-of-sample return of the ML rule with 10 risky assets**
This figure plots the unconditional distribution of $r_{p,t+1}$ for $h = 60$ months and 120 months with parameters estimated using excess monthly returns of the 10 momentum portfolios over the period of 1927/1–2014/12. The risk aversion coefficient is set to three ($\gamma = 3$). For comparison, the return distribution of the true optimal portfolio is also reported.

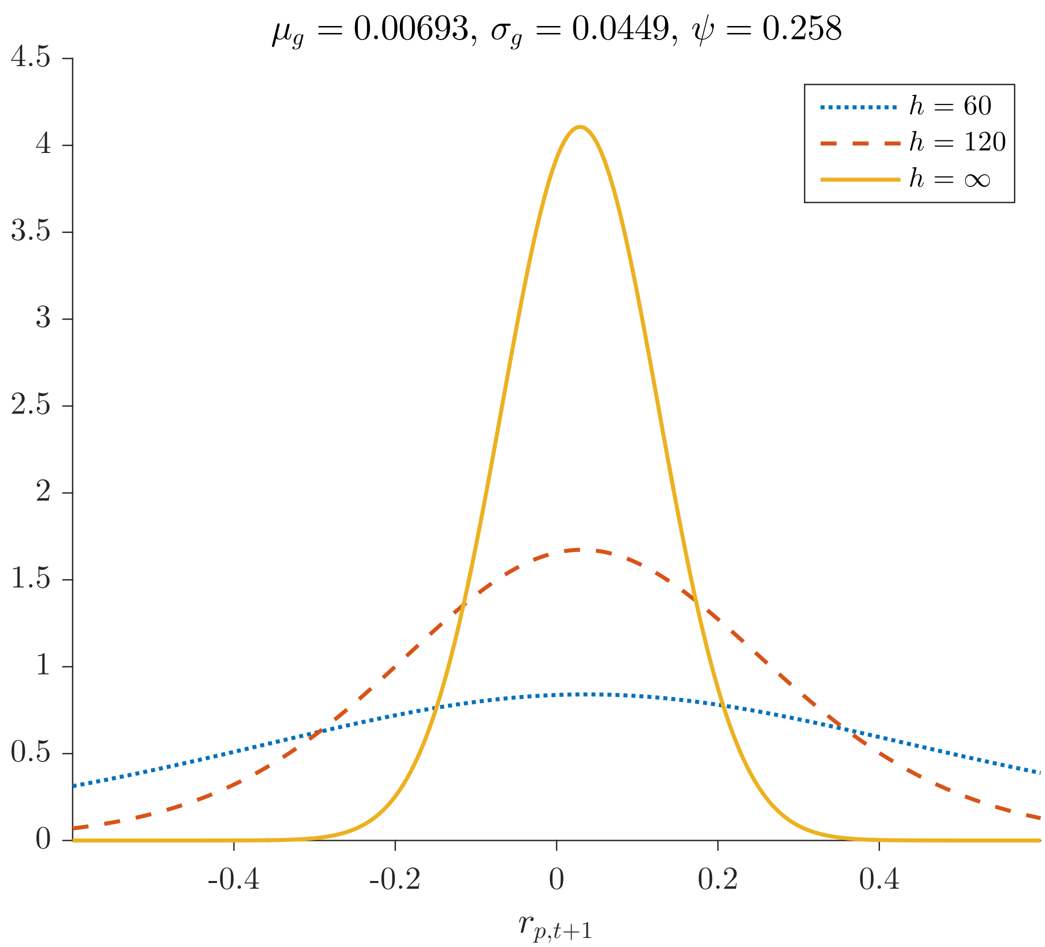


Figure 5: **Unconditional distribution of out-of-sample return of the ML rule with 25 risky assets**
This figure plots the unconditional distribution of $r_{p,t+1}$ for $h = 60$ months and 120 months with parameters estimated using excess monthly returns of the Fama-French 5×5 size and book-to-market ranked portfolios over the period of 1927/1–2014/12. The risk aversion coefficient is set to three ($\gamma = 3$). For comparison, the return distribution of the true optimal portfolio is also reported.

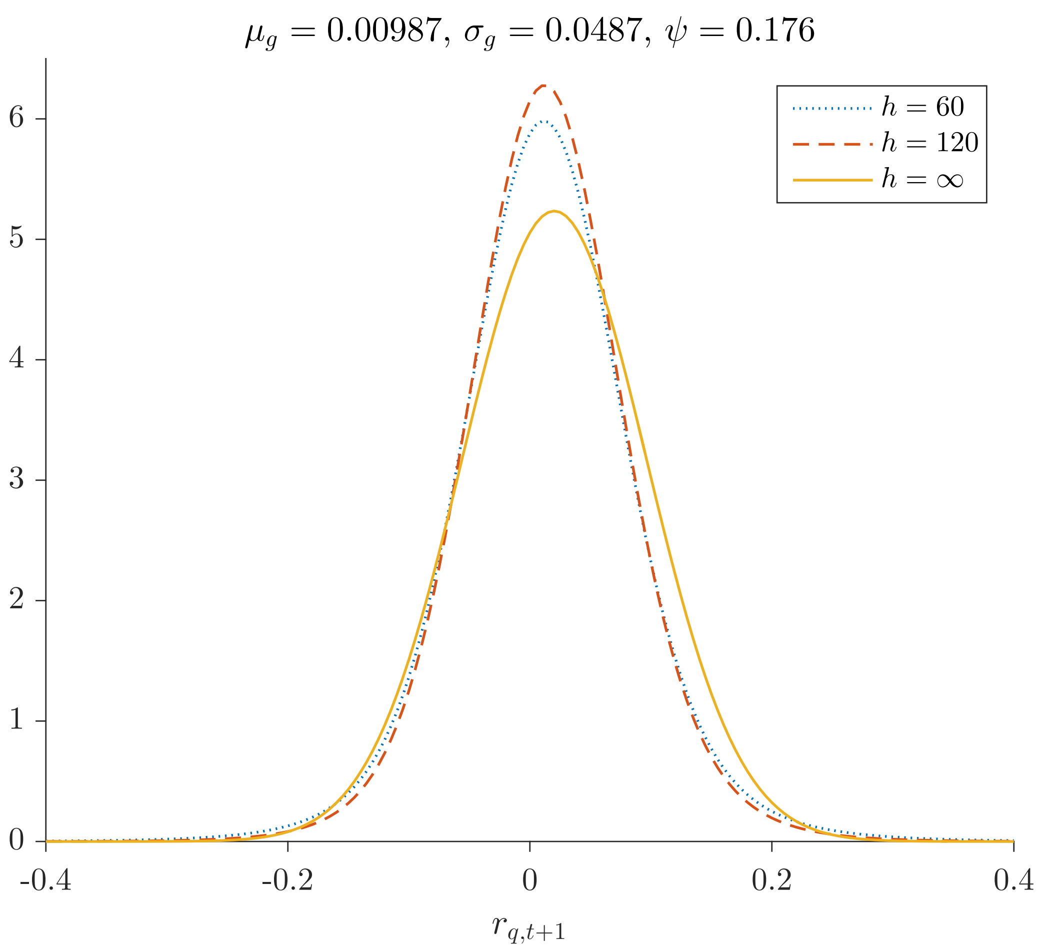


Figure 6: **Unconditional distribution of out-of-sample return of the QL rule with 10 risky assets**
This figure plots the unconditional distribution of $r_{q,t+1}$ for $h = 60$ months and 120 months with parameters estimated using excess monthly returns of the 10 momentum portfolios over the period of 1927/1–2014/12. The risk aversion coefficient is set to three ($\gamma = 3$). For comparison, the return distribution of the true optimal portfolio is also reported.

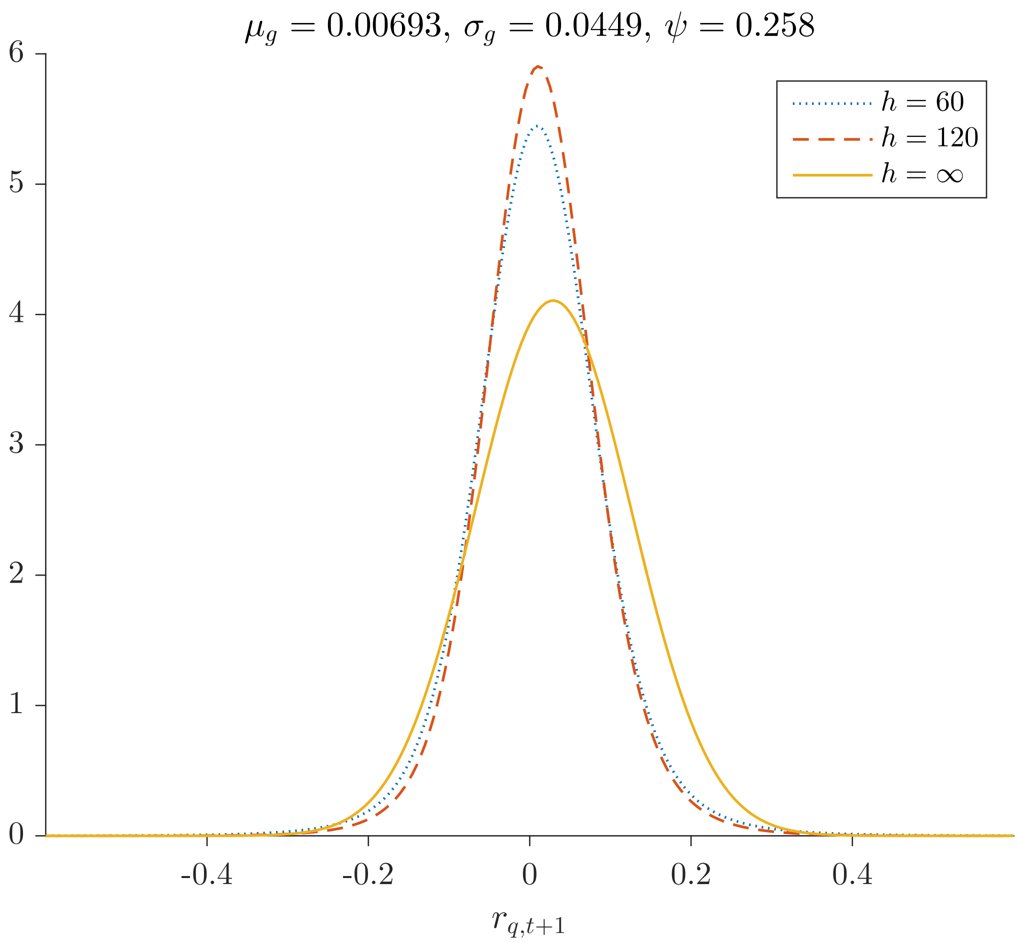


Figure 7: **Unconditional distribution of out-of-sample return of the QL rule with 25 risky assets**

This figure plots the unconditional distribution of $r_{q,t+1}$ for $h = 60$ months and 120 120 months with parameters estimated using excess monthly returns of the Fama-French 5×5 size and book-to-market ranked portfolios over the period of 1927/1–2014/12. The risk aversion coefficient is set to three ($\gamma = 3$). For comparison, the return distribution of the true optimal portfolio is also reported.

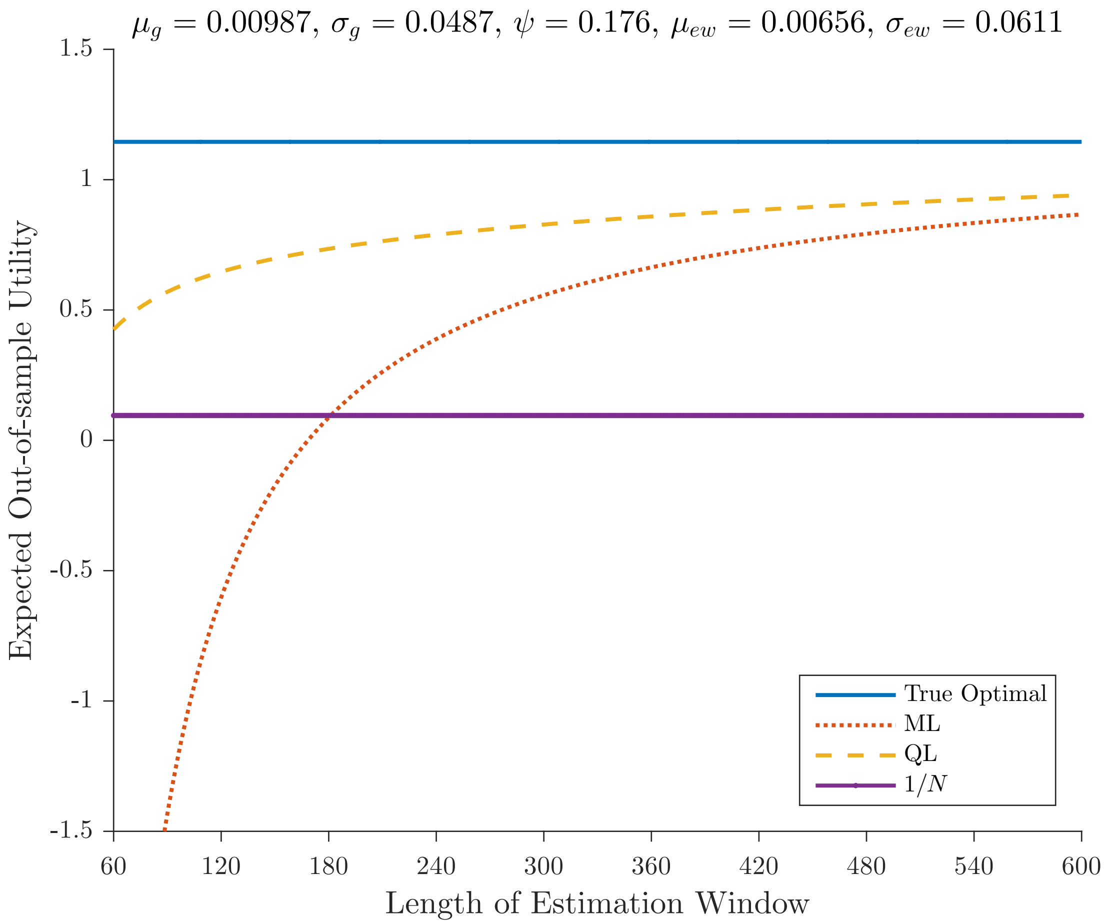


Figure 8: **Expected out-of-sample utility of various portfolio rules for the case with 10 risky assets**

This figure plots the expected out-of-sample utility (in percentage points) of the true optimal portfolio, the ML rule, the QL rule, and the $1/N$ rule as a function of the length of the estimation window (h), with parameters estimated using excess monthly returns of the 10 momentum portfolios over the period of 1927/1-2014/12. The risk aversion coefficient is set to three ($\gamma = 3$).

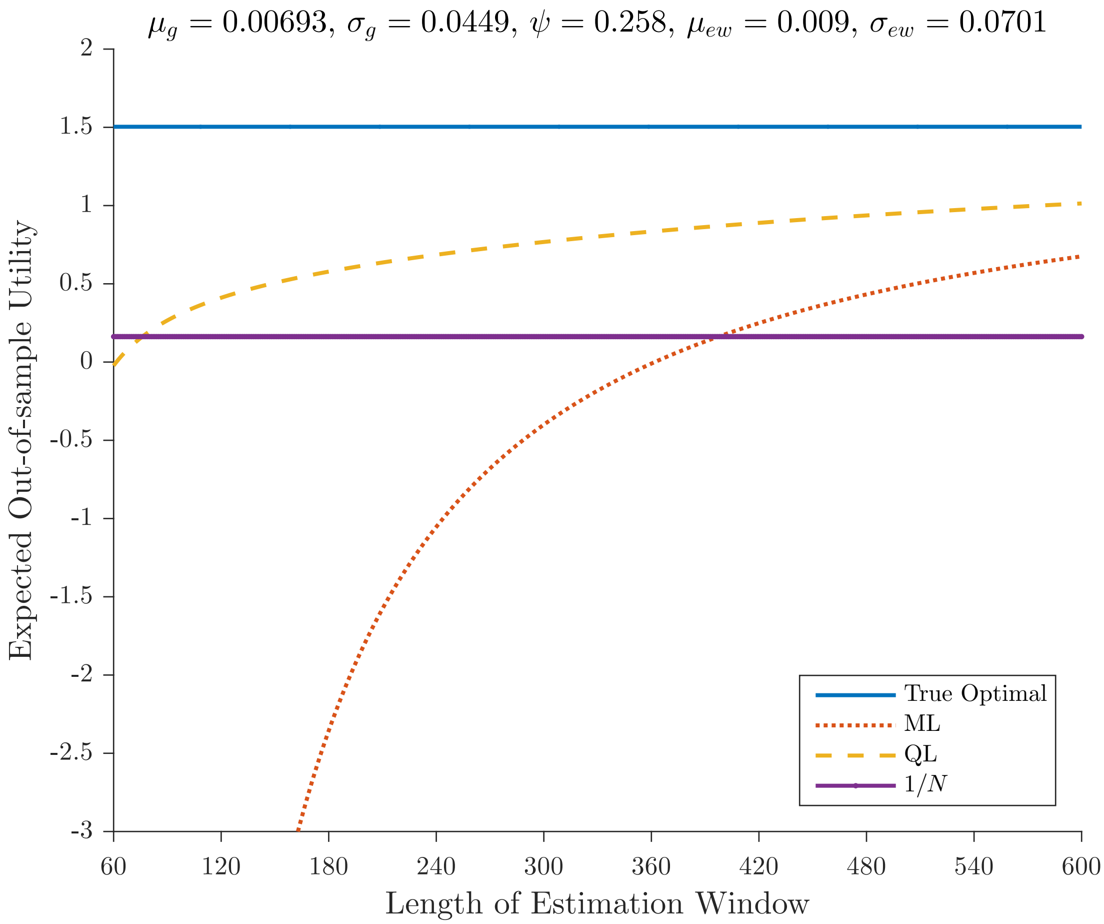


Figure 9: **Expected out-of-sample utility of various portfolio rules for the case with $N = 25$ risky assets**

This figure plots the expected out-of-sample utility (in percentage points) of the true optimal portfolio, the ML rule, the QL rule, and the $1/N$ rule as a function of the length of the estimation window (h), with parameters estimated using excess monthly returns of Fama-French 5×5 size and book-to-market ranked portfolios over the period of 1927/1-2014/12. The risk aversion coefficient is set to three ($\gamma = 3$).

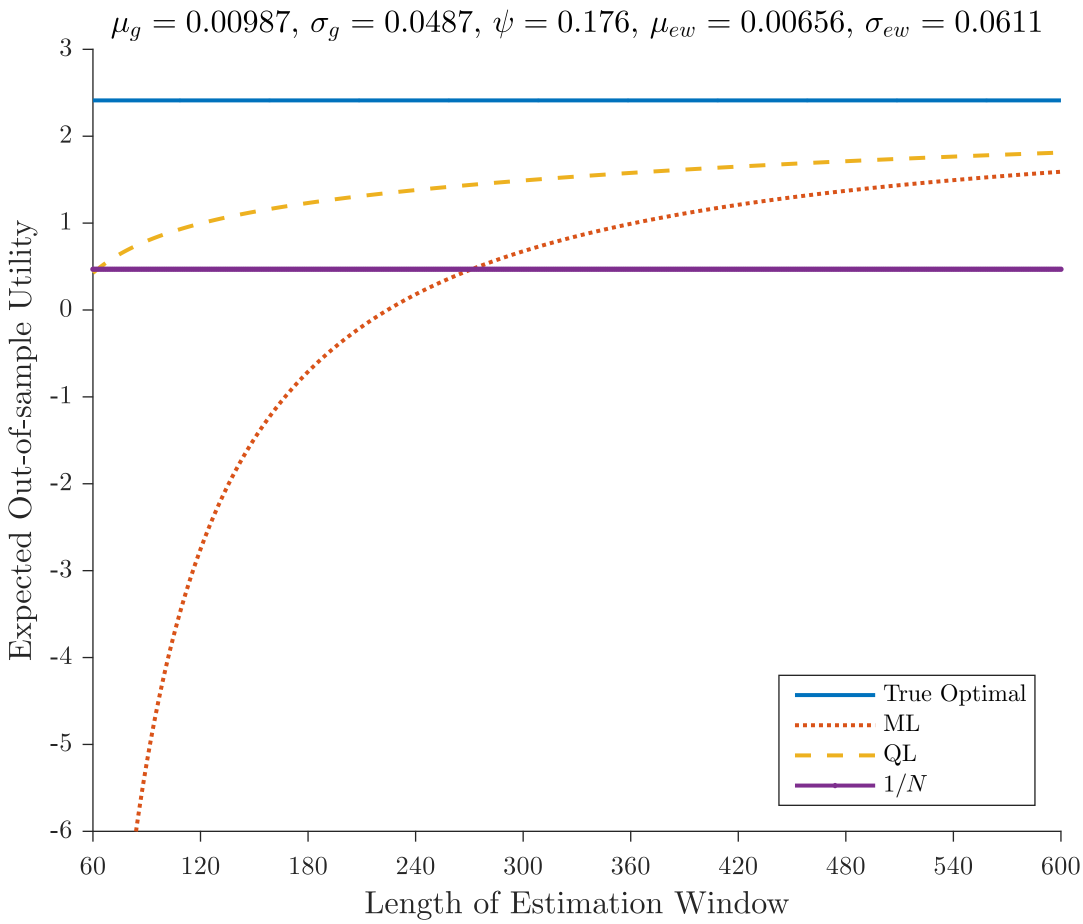


Figure 10: **Expected out-of-sample utility of various portfolio rules for the case with 10 risky assets**
This figure plots the expected out-of-sample utility (in percentage points) of the true optimal portfolio, the ML rule, the QL rule, and the $1/N$ rule as a function of the length of the estimation window (h), with parameters estimated using excess monthly returns of the 10 momentum portfolios over the period of 1927/1-2014/12. The risk aversion coefficient is set to one ($\gamma = 1$).

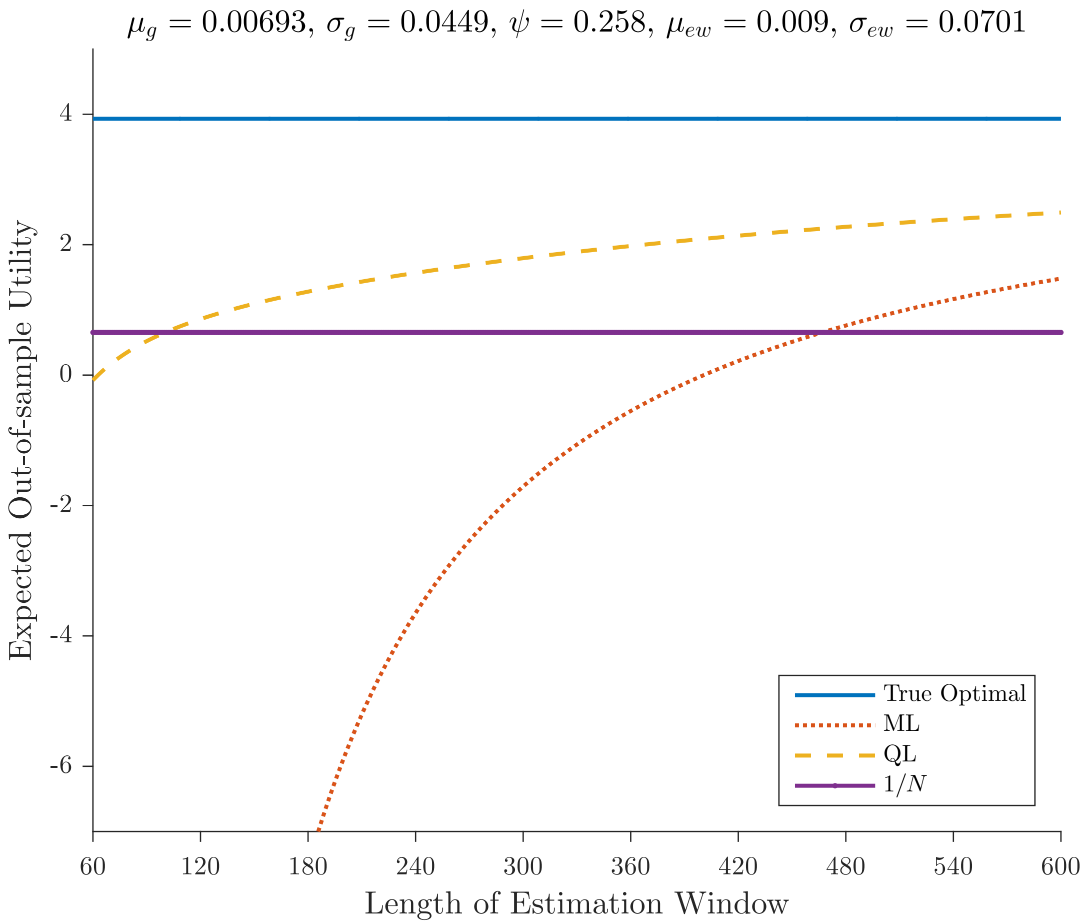
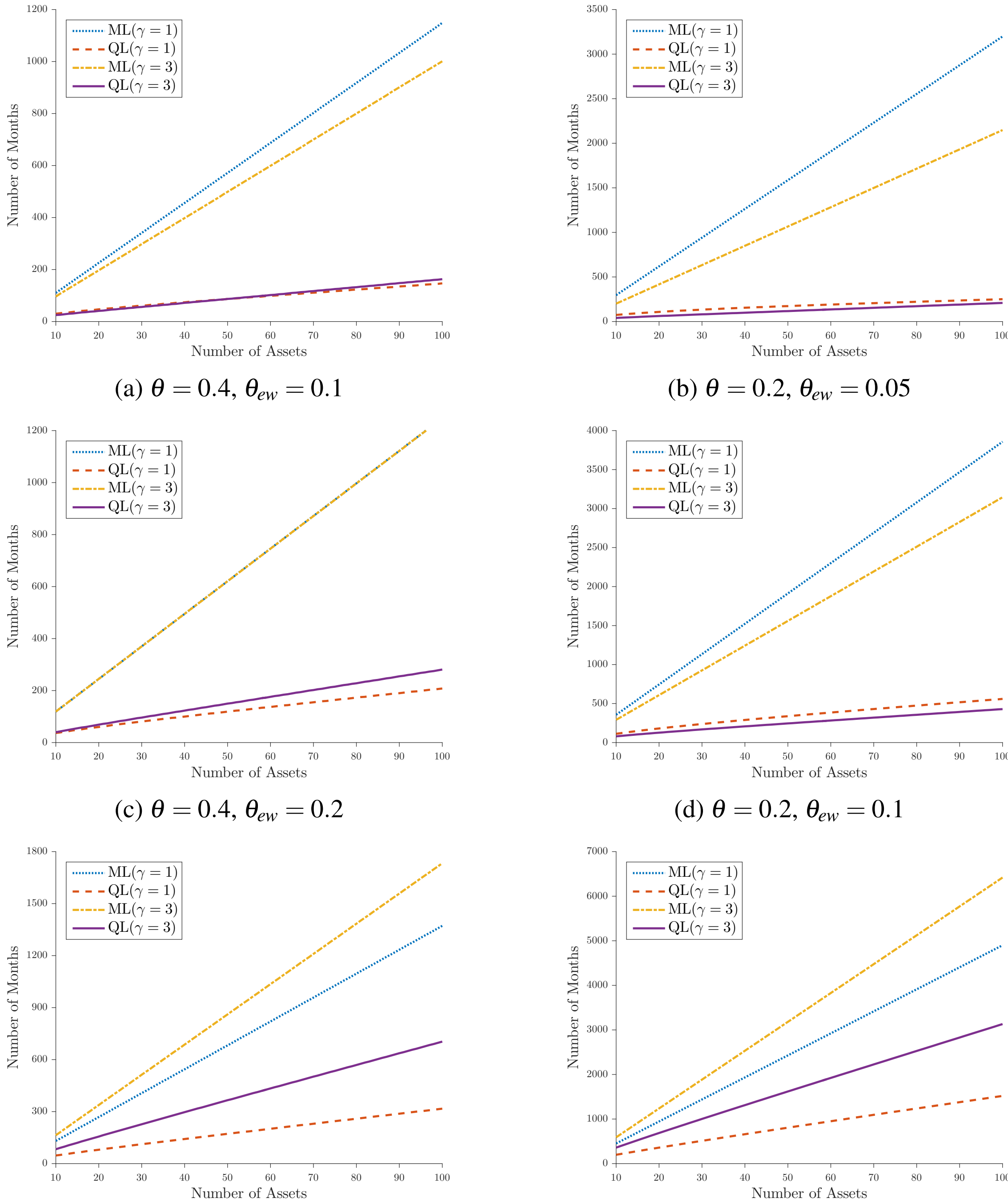


Figure 11: **Expected out-of-sample utility of various portfolio rules for the case with 25 risky assets**

This figure plots the expected out-of-sample utility (in percentage points) of the true optimal portfolio, the ML rule, the QL rule, and the $1/N$ rule as a function of the length of the estimation window (h), with parameters estimated using excess monthly returns of Fama-French 5×5 size and book-to-market ranked portfolios over the period of 1927/1-2014/12. The risk aversion coefficient is set to one ($\gamma = 1$).



(a) $\theta = 0.4$, $\theta_{ew} = 0.1$

(b) $\theta = 0.2$, $\theta_{ew} = 0.05$

(c) $\theta = 0.4$, $\theta_{ew} = 0.2$

(d) $\theta = 0.2$, $\theta_{ew} = 0.1$

(e) $\theta = 0.4$, $\theta_{ew} = 0.3$

(f) $\theta = 0.2$, $\theta_{ew} = 0.15$

Figure 12: **Number of estimation months required to outperform the $1/N$ rule**

This figure plots the required number of estimation months for the ML and the QL rule to outperform the $1/N$ rule as a function of the number of risky assets (N) when a risk-free asset is not available. The six panels report results for different combinations of θ (0.4 or 0.2) and θ_{ew}/θ (0.25, 0.5, or 0.75), with the Sharpe ratio of the global minimum-variance portfolio (θ_g) set equal to $\theta/2$ in each case. In addition, we assume $\sigma_g = 0.05$ and $\sigma_{ew} = 0.065$.