Optimal Portfolio Selection: New Strategies for Minimizing Estimation Risk

Raymond Kan, Xiaolu Wang and Guofu Zhou*

JEL classification: G11; G12; C11

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

First version: March, 2015

Current version: April, 2015

*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.

Optimal Portfolio Selection:
New Strategies for Minimizing Estimation Risk

In this paper, we provide two new portfolio rules that mitigate parameter estimation risk in the mean-variance model pioneered by Markowitz (1952) and widely used in practice. While existing studies focus on the riskless asset case, we study the previously unsolved and practically more important case in which there is no riskless asset. We find that the proposed rules are reliably useful for typically samples sizes while the usual plug-in estimated rule can often fail to deliver any positive certainty equivalence returns. In addition, we find that the proposed rules perform well relative to the well known $1/N$ rule in both calibrations and real data sets. Overall, our study shows that the estimation risk can be mitigated substantially as in the riskless asset case.

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 of the main reasons is that many implement issues, such as factor exposures and trading constraints, can be accommodated easily within this framework which allows for analytical insights and fast numerical solutions. Another reason is that the intertemproal hedging demand is found typically small so that independent returns overtime is a workable assumption in the real world. However, to apply the mean-variance frame in practice, the true parameters are unknown and have to be estimated from data. This raises the estimation risk because an optimal portfolio rule based on the estimate parameters is subject to random errors and can be substantially different from the true optimal rule. Brown (1976), Bawa, Brown, and Klein (1979), and Jorion (1986) are examples of earlier work that provide sophisticated portfolio rules to mitigate the estimation risk. Recently, Kan and Zhou (2007), DeMiguel, Garlappi, and Uppal (2009), and Tu and Zhou (2004), among others, provide explicit portfolio rules that are designed to minimize the impact of estimation risk. However, their studies are restricted to the case of the existence of the riskless asset case.

In this paper, we provide two explicit portfolio rules that reduce the impact of estimation risk in the case without the riskless asset, which is a practically more important case because mutual funds and institutional equity funds are required to be fully invested in the stock market. In contrast to the standard plug-in portfolio rule that replaces the unknown moments of asset returns by their maximum likelihood estimates (ML rule), we adjust a decomposition of the ML rule by a suitable function of the data. The function is designed in such a way to optimally account for the estimation risk in estimating the unknown moments, which is entirely ignored in the ML rule. There are two different criteria are used to measure the utility in the presence of estimation risk. The first is the expected utility that takes expectation over the randomness of the parameter estimates, and the second is the *empirical* utility that takes expectation of the portfolio return over the asset risk and estimation risk, and then computes the utility. While the first is the ones in standard decision theory, the second is the one used almost always in practice. Based on the two criteria, we derive

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

two portfolio rules, called QL and UL, respectively. We obtain explicitly the expected returns and risks (standard deviations) of the two portfolios and the associated utilities. Under reasonable sample sizes in practice, we find that QL and UL perform well. They dominates ML which can perform quite poorly.

In their thought provoking paper, DeMiguel, Garlappi, and Uppal (2009) compare the $1/N$ rule not only with the sample-based mean-variance ML rule, but also with a host other sophisticated rules. They find 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." While Tu and Zhou (2011) provide portfolio rules that are generally outperform the $1/N$, nothing has been done in the riskless asset case. Interestingly, we show that, in the no riskless asset case here, the QL and UL rules outperform the $1/N$ in all of our calibrations. Moreover, for the same data sets used by DeMiguel, Garlappi, and Uppal (2009), they also outperform the $1/N$ rule except in two cases where they have similar performances. While insights from DeMiguel, Garlappi, and Uppal (2009) cast some doubt on the value of existing investment theory, our paper re-affirms the usefulness of the mean-variance analysis in the widely relevant no riskless asset case.

The remainder of the paper is organized as follows. Section 2 provides the optimal portfolio rules that accounts for estimation risk. Section 3 compares the performances of various rules. Section 4 concludes.

2. Portfolio Rules

In this section, we review first the mean-variance framework, then study the properties of the ML rule to understand the estimation risk. Armed with this understanding, we proceed to derive the new QL rule and provide analytical results on its distribution.

2.1 The Portfolio Choice Problem

Consider the standard portfolio choice problem in which an investor chooses his optimal portfolio among N risky assets. Denote their returns at time t by R_t, an $N \times 1$ vector. For analytical tractability, we make the common assumption that R_t is independent and identically distributed over time, and has a multivariate normal distribution with mean μ and covariance matrix Σ.

Under the standard mean-variance framework, the investor at time T chooses his portfolio weights w so as to maximize the quadratic utility function

$$U(w) = E[R_{p,T+1}] - \frac{\gamma}{2}\mathrm{Var}[R_{p,T+1}] = w'\mu - \frac{\gamma}{2}w'\Sigma w, \tag{1}$$

where

$$R_{p,T+1} = w'R_{T+1}$$

the future uncertain portfolio return at time $T+1$ and γ is the coefficient of relative risk aversion. It is easy to show that, when both μ and Σ are assumed known, the optimal portfolio weights are

$$w^* = w_g + \frac{1}{\gamma}w_z, \tag{2}$$

where

$$w_g = \frac{\Sigma^{-1}1_N}{1_N'\Sigma^{-1}1_N}, \quad w_z = \Sigma^{-1}(\mu - 1_N\mu_g),$$

with 1_N being an $N \times 1$ vector of ones.

In the familiar mean-variance frontier, w_g is the global minimum-variance portfolio whose weights sum to 1. In contrast, w_z is a zero investment portfolio satisfying $1_N'w_z = 0$. It is clear from (2) that any frontier portfolio is a linear combination of w_g and w_z. As the risk aversion varies, the optimal portfolio from (2) will trace out the upper frontier. Mathematically, maximizing the quadratic utility function is equivalent to the usual portfolio risk minimization for a given level of return. Practitioners (see, e.g., Qian, Hua, and Sorensen (2007)) often use the utility set-up because of its convenient interpretation that $U(w)$ is the risk-adjusted return.

Equation (2) says that holding the optimal portfolio is the same as investing into two funds, w_g and w_z. Since investors here invest 100% into the risky assets, they always hold 100% of w_g. Depending on their degrees of risk aversion, their exposures to w_z vary. Note that in the riskless asset case, investors also hold two funds, but they are the riskless asset and the tangency portfolio. Without the riskless asset, there is no tangency portfolio here.

Denote $R_{p^*,T+1} = \hat{w}^{*'}R_{T+1}$ as the true optimal portfolio based on w^*. To facilitate the discussions below, it will be useful to summarize its properties. The mean and variance of this optimal portfolio are

$$\mu_{p^*} = \mu_g + \frac{\psi^2}{\gamma}, \tag{3}$$

$$\sigma_{p^*}^2 = \sigma_g^2 + \frac{\psi^2}{\gamma^2}, \tag{4}$$

where μ_g and σ_g^2 are the mean and variance of the global minimum-variance portfolio given by

$$\mu_g = 1_N'\Sigma^{-1}\mu/(1_N'\Sigma^{-1}1_N), \quad \sigma_g^2 = 1/(1_N'\Sigma^{-1}1_N),$$

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 frontier. It follows that the utility from holding the optimal portfolio is given by

$$U(w^*) = \mu_g - \frac{\gamma}{2}\sigma_g^2 + \frac{\psi^2}{2\gamma}. \tag{5}$$

This equation shows that w^* outperforms w_g by a certainty equivalent return of $\psi^2/(2\gamma)$. The squared slope of the asymptote as well as the risk aversion coefficient determine the utility difference. For the extreme risk-averse investor with $\gamma = \infty$, he holds only the global minimum-variance portfolio.

2.2 The ML Rule

In practice, however, the optimal portfolio weights, w^*, are not computable because μ and Σ are unknown. Given return data of sample size T, the standard estimates of these parameters are

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t, \tag{6}$$

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})'. \tag{7}$$

Then the (estimated) optimal portfolio rule is

$$\hat{w} = \hat{w}_g + \frac{1}{\gamma}\hat{w}_z, \tag{8}$$

where

$$\hat{w}_g = \frac{\hat{\Sigma}^{-1}1_N}{1_N'\hat{\Sigma}^{-1}1_N}, \quad \hat{w}_z = \hat{\Sigma}^{-1}(\hat{\mu} - 1_N\hat{\mu}_g).$$

Since the unknown moment parameters μ and Σ are estimated by their maximum likelihood estimators, we call $\hat{w}$ the ML rule.

Since w^* is unobservable, $\hat{w}$ is the common feasible alternative. Based on $\hat{w}$, the out-of-sample return for at time $T+1$ is

$$R_{p,T+1} = \hat{w}' R_{T+1}, \tag{9}$$

where $\hat{p}$ indicates that this is an estimated portfolio rule from the data. Note that there are now two risks. The first is the uncertainly associated with future asset return, and the second is the uncertainly associated with using the estimated weights that are subject to random errors. As a result, the distribution of $R_{\hat{p},T+1}$ is rather complex. The following proposition (all proofs are in the Appendix) summarizes its first two moments:

Proposition 1: The mean and variance of $R_{p,T+1}$, the estimated optimal portfolio based on the ML rule, are:

$$\mu_p = \mu_g + \frac{T}{T-N-1}\frac{\psi^2}{\gamma}, \tag{10}$$

$$\sigma_p^2 = \frac{\sigma_g^2(\psi^2+T-2)}{T-N-1} + \frac{T(T-2)[(T+1)\psi^2+N-1]}{\gamma^2(T-N)(T-N-1)(T-N-3)} + \frac{2T^2\psi^4}{\gamma^2(T-N-1)^2(T-N-3)}. \tag{11}$$

In comparison with the true optimal portfolio, it is interesting that μ_p differs from μ_{p^*} only by the scaler $T/(T-N-1)$. As the sample becomes larger and larger, the estimation errors become smaller and smaller, so the scaler converges to 1, and $E[\mu_{\hat{p}}]$ approaches to μ_{p^*}. However, since $T/(T-N-1) > 1$, the expected return on $R_{p,T+1}$ is always greater than that on $R_{p^*,T+1}$. But the risk σ_p^2 is also disproportionally larger too because $R_{p^*,T+1}$ is optimal for a given level of expected return.

Indeed, in comparison with $\sigma_{p^*}^2$, σ_p^2 has now three terms. The first two terms are similar those of σ_p^2 except the scaling coefficients, snd converge to them as the sample size T goes to large. This is expected since when the sample increases, the impact of the estimation errors decrease. The third term is of high order relative to T and is small and negligible. However, even if T is large, as long as N is roughly half of T, the coefficients of the first two terms will be greater than 2 and 8, respectively. In this case, the variance of the $R_{p,T+1}$ will be at least more than double that of the true optimal portfolio. This indicates one should be really careful is using the ML portfolio in practice, especially when the asset and sample size ratio, N/T, is large, say close to 1/2.

To see more clearly how the estimation risk exasperates the volatility of the estimated optimal portfolio $R_{p^*,T+1}$, Figure 1 plots its return distribution for sample sizes $T = 120, 240$ and ∞, with parameters $\mu_g = 0.0127$, $\sigma_g = 0.0487$ and $\psi = 0.176$, calibrated from $N = 10$ value-weighted momentum decile portfolios with monthly data from January 1927 to December 2014 available from French's website. The risk aversion is set at $\gamma = 3$. When $T = \infty$, the density is that of the true optimal portfolio. When $T = 60$, a sample size of 5 years of monthly data, Figure 1 shows that the standard deviation of $R_{p^*,T+1}$ is quite large with long fat tails. When $T = 120$, it improves, but not dramatically. Overall, the figure highlights the impact of estimation risk on the estimated optimal portfolio.

Analytically, we know that the volatilities of $R_{p,T+1}$ increases as N goes up, holding everything else equal. Figure 2 illustrates this more vividly for a modest asset size of $N = 25$, with parameters calibrated from the Fama-French value-weighted size-BM 25 portfolios of the same period as before. In this case, relative to the distribution of the true optimal portfolio, the distribution of the estimated optimal portfolio has a much larger standard deviation and much longer tails than the previous case. Again, the figure indicates the danger of using the estimated ML optimal portfolio by ignoring the estimation risk.

Since $R_{p,T+1}$ has higher expected return than $R_{p^*,T+1}$ though it has higher volatility, the correct way to compare their performance economically is to examine its risk-adjusted return, or the expected utility. Theoretically, the expected utility is

$$E[U(\hat{w})] = E\left([\hat{w}'R_{T+1}] - \frac{\gamma}{2}\text{Var}[\hat{w}'R_{T+1}]\right). \tag{12}$$

Note that here we have not only the randomness of the future asset return R_{T+1}, but also the randomness in $\hat{w}$. The expectation is thus computed in two-steps. First, we compute the usual utility conditional on the portfolio weights. Second, we compute the expected value of this utility over the uncertainty of the estimates. This is the standard utility in decision theory or in a Bayesian framework when there is parameter uncertainty. Following DeMiguel, Garlappi, and Uppal (2009), we can interpret $E[U(\hat{w})]$ intuitively as the average utility realized by an investor who plays the ML estimation strategy infinitely many times. Because of taking the expectation over the estimation uncertainty, the analytical evaluation of $E[U(\hat{w})]$ is complex. We summarize the result as

Proposition 2: The expected utility of the ML portfolio, $R_{p^*,T+1}$, is:

$$E[U(\hat{w})] = U(w_g) - \frac{\gamma}{2}\frac{(N-1)\sigma_g^2}{(T-N-1)} + \frac{k_0\psi^2}{2\gamma} - \frac{(N-1)T(T-2)}{2\gamma(T-N)(T-N-1)(T-N-3)}, \tag{13}$$

where

$$U(w_g) = \mu_g - \frac{\gamma}{2}\sigma_g^2,$$

the expected utility of the global minimum-variance portfolio, and

$$k_0 = \left(\frac{T}{T-N-1}\right)\left[2 - \frac{T(T-2)}{(T-N)(T-N-3)}\right],$$

a constant that is independent of any unknown parameters.

It is clear that, as the sample size gets larger and larger, the second and the four terms converges to zero, and k_1 converges to 1. Hence, as expected, $E[U(\hat{w})]$ approaches the true expected utility given by (5).

It is interesting to note that while individual parameters, components of μ and Σ, play a role in determining the optimal portfolio weights, the optimal utility achievable depend upon only three summary parameters, μ_g, σ_g and ψ^2, the shape of the mean-variance frontier. Now, in the presence of estimation risk of estimating all the parameters, the expected utility from the estimated ML portfolio rule is still a function of μ_g, σ_g and ψ^2 alone without depending on other unknown parameters. Although the three summary parameters are unknown in practice, they can be estimated using their sample analogues.

To understand further the role of the estimation risk, consider first the case when Σ is assumed known, but μ is estimated. In this case, it is easy to show that

$$E[U(\hat{w})|\Sigma] = U(w^*) - \frac{1}{\gamma}\frac{N-1}{2T}. \tag{14}$$

This means that, conditional on knowing Σ, the impact of the estimation risk in estimating μ is irrelevant to the actual magnitude of μ, and is a function only of T, N and γ. When T increases, the estimation errors decrease and the utility difference between $E[U(\hat{w})|\Sigma]$ and the optimal one (without estimation risk) reduces. However, for a given sample size, the number of assets will make the comparison worse. This makes intuitive sense too. As there are more and more assets, the estimation errors in estimating all their means can compound. When the investor's risk aversion γ increases, less money is invested in the zero-cost portfolio that should reduce the estimation to yield a higher value of $E[U(\hat{w})|\Sigma]$. Note that the global minimum-variance portfolio has no estimation risk if Σ is known.

Consider now the case when μ is assumed known, but Σ is estimated. Then we have

$$E[U(\hat{w})] = U(w_g) + \frac{k_0 \psi^2}{2\gamma}. \tag{15}$$

The second and the fourth terms of (13) disappear, and hence they are primarily driven by estimation errors in Σ.

In general, based on Proposition 2, the expected loss of utility due to estimation errors when both μ and Σ are estimated is

$$U(w^*) - E[U(\hat{w})] = \frac{\gamma}{2}\frac{(N-1)\sigma_g^2}{(T-N-1)} + (1-k_0)\frac{\psi^2}{2\gamma} + \frac{(N-1)T(T-2)}{2\gamma(T-N)(T-N-1)(T-N-3)}. \tag{16}$$

The first term is due to estimation error of $\hat{w}_g$. The last two terms are due to estimation error of $\hat{w}_z$. The magnitude of the overall impact will depend on unknown parameters σ_g and ψ. Once these parameters are calibrated from data, it will be straightforward to examine the utility difference.

Figure 3 plots, among others, the true optimal utility and the ML utility for various sample sizes, with the same parameters as in the previous $N = 10$ case. It is seen that the expected utility of the ML rule is negative or zero when $T \leq 120$. But it rises sharply as the sample size goes up from 120 to 350. As the same size increases further, though the utility still rises significantly, the speed of the rise goes down. Figure 4 plots the same utilities as Figure 3 with $N = 25$ and with the same parameters as Figure 2. The pattern is quite similar to Figure 3, but it requires now roughly $T = 325$ for the ML rule to yield positive utilities. This is consistent with the theory that as N increases, the sample size must increase much further to yield the same utility level close to the true optimal one.

Proposition 2 also sheds light on why many practitioners advocate investing into only the global minimal risk portfolio, of which Basak, Jagannathan and Ma (2009) provide an estimator for its estimation risk. Under the iid normal assumption, the first term of (16) provides the total risk of $\hat{w}_g$ explicitly. For fixed parameters, when $N < T$ gets large, the other two terms completely dominate the first term. Therefore, if one is too concerned about the estimation risk and when $N < T$ is modest large, one is much better off investing in $\hat{w}_g$ than in the ML portfolio.

2.3 The QL Rule

Statistically, our objective here is to provide the best estimator of the unknown portfolio weights (not necessarily the best estimator for μ and Σ. Following Brown (1976), Stambaugh (1997), Kan and Zhou (2007), and DeMiguel, Garlappi, and Uppal (2009), among many others, we consider the standard quadratic loss function from statistical decision theory,

$$L(w^*,\tilde{w}) \quad = \quad \frac{\gamma}{2}E\left[(\tilde{w}-w^*)'\Sigma(\tilde{w}-w^*)\right]. \tag{17}$$

It is a quadratic function of the errors in estimating w^*. Mathematically, it is easy to show that

$$L(w^*,\tilde{w}) = U(w^*) - E[U(\tilde{w})], \qquad U(\tilde{w}) \equiv \tilde{w}'\mu - \frac{\gamma}{2}\tilde{w}'\Sigma\tilde{w}. \tag{18}$$

So the quadratic loss criterion is equivalent to maximizing the expected utility of the estimated optimal portfolio rule. Note that the expected utility is defined in the standard way as in the decision theory where the expected utility is computed in two steps. First, conditional on the parameter estimates, the utility is computed is the usual way. Second, further expectation is taken over the uncertainty of the parameter estimates.

How do we find the optimal estimator of w to maximum the expected utility? This is an extremely difficult and open problem in statistics. Instead of search all possible functional forms of the data, the literature has been focuses on finding the optimal estimator in certain special class of estimator. We follow this tradition in this paper too.

Our analysis in the previous subsection indicates that $\hat{p}$ has greater expected return than the true optimal portfolio, but suffers more volatility than compensated for. As a result. we consider the following class of 2-fund rules,

$$\hat{w}(c) = \hat{w}_g + \frac{c}{\gamma}\hat{w}_z. \tag{19}$$

Our objective is to find the best c that maximizes the expected utility. Intuitively, the optimal c must be less than 1 because we want to reduce the expected return on this portfolio to counter balance the non-compensated risk.

Statistically, our procedure is equivalent to using a shrinkage estimator of type $\hat{\Sigma}/c$ for Σ. Jorion (1986) and Jagannathan and Ma (2003), among others, are earlier examples of using shrinkage estimators for Σ in finance. It is interesting to note that $\hat{w}_g$ is invariant to c, so that the shrinkage plays a role only for the $\hat{w}_z$ component. Since our procedure is motivated economically to lower the expected return of $\tilde{w}(c)$ to minimize the total risk, the choice of c amounts to lower the exposure on the zero investment portfolio $\hat{w}_z$ as shown below.

Based on Proposition 1, it is easy to show that

$$\begin{aligned} E[U(\hat{w}(c))] &= E[\mu'\hat{w}_g] + \frac{c}{\gamma}E[\mu'\hat{w}_z] - \frac{\gamma}{2}(E[\hat{w}_g'\Sigma\hat{w}_g] + c^2E[\hat{w}_z'\Sigma\hat{w}_z]) \\ &= \mu_g + \frac{c}{\gamma}\frac{T\psi^2}{(T-N-1)} - \frac{\gamma}{2}\frac{(T-2)\sigma_g^2}{(T-N-1)} + \frac{c^2(T\psi^2+N-1)}{2\gamma(T-N-1)k_1}, \end{aligned} \tag{20}$$

where

$$k_1 = \frac{(T-N)(T-N-3)}{T(T-2)}.$$

Note that this is a quadratic function of c, and so the optimal c is given by

$$c^* = k_1\frac{\psi^2}{\psi^2 + \frac{N-1}{T}} = k_1 g_1(\psi^2), \tag{21}$$

where

$$g_1(x) = x/(x+(N-1)/T).$$

The nations of k_1 and g_1 will be useful to simply expressions below.

Following Kan and Zhou (2007), we can estimate c^* by

$$\hat{c} = k_1 g_1(\hat{\psi}_a^2), \tag{22}$$

where

$$\hat{\psi}_a^2 = \frac{(T-N-1)\hat{\psi}^2 - (N-1)}{T} + \frac{2(\hat{\psi}^2)^{\frac{N-1}{2}}(1+\hat{\psi}^2)^{-\frac{T-2}{2}}}{TB_{\hat{\psi}^2/(1+\hat{\psi}^2)}((N-1)/2,(T-N+1)/2)}, \tag{23}$$

with $\hat{\psi}^2 = \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu} - (1_N'\hat{\Sigma}^{-1}\hat{\mu})^2/(1_N'\hat{\Sigma}^{-1}1_N)$ and $B_x(a,b)$ is the incomplete beta function. Then we can define the QL rule as

$$\hat{w}_q = \hat{w}_g + \frac{k_1}{\gamma}g_1(\hat{\psi}_a^2)\hat{w}_z. \tag{24}$$

This rule is clearly well defined because it a function of data and is independent of any unknown parameters. We call it the QL rule because of its quadratic loss motivation and because of its easy name comparison with the ML rule.

Let $R_{q,t+1}$ be the portfolio return at time $t+1$, the portfolio associated with using weights $\hat{w}_q$. Then we have

Proposition 3: The mean and variance of $R_{q,t+1}$, the estimated optimal portfolio based on the QL rule, are:

$$\mu_q = \mu_g + \frac{k_1 T\psi^2}{\gamma(T-N-1)} E[g_1(f_2)], \tag{25}$$

$$\sigma_q^2 = \frac{(T-2+\psi^2)\sigma_g^2}{T-N-1} + \frac{Tk_1^2(T-2+\psi^2)E[g_1^2(f_1)f_1]}{\gamma^2(T-N)(T-N-1)} - \frac{T^2k_1^2\psi^4E^2[g_1(f_2)]}{\gamma^2(T-N-1)^2} + \frac{Tk_1^2\psi^2\left(E[g_1^2(f_3)]+T\psi^2E[g_1^2(f_4)]\right)}{\gamma^2(T-N)(T-N-3)}, \tag{26}$$

where $f_1,\ldots,f_4$ are mutually independent random variables with distributions,

$$f_1 \sim \frac{\chi^2_{N-1}(T\psi^2)}{\chi^2_{T-N-1}}, \quad f_2 \sim \frac{\chi^2_{N+1}(T\psi^2)}{\chi^2_{T-N-1}}, \quad f_3 \sim \frac{\chi^2_{N+1}(T\psi^2)}{\chi^2_{T-N-3}}, \quad f_4 \sim \frac{\chi^2_{N+3}(T\psi^2)}{\chi^2_{T-N-3}}.$$

These explicit expressions for the mean and variance of the QL portfolio are much more complex than those of the ML portfolio. This is necessarily so because the weights involve the use of the complex $\hat{c}$ estimator. However, numerically, they are easily evaluated as only simple one-dimensional integrations are needed. While simulations may be used alternatively to evaluation the mean and variance, it will be necessary to specify all the parameters and it requires essentially to compute an integral with dimensionality $N+N(N-1)/2$ with random errors and computationally demanding.

The explicit expressions provide interesting analytical insights. Since $k_1 < 1$ and $g_1(x) < 1$, we know that $R_{q,t+1}$ has lower expected return than $R_{p,t+1}$. This is exactly our goal to reduce the risk. Numerically, it is easy to verify that it is indeed the case that $\sigma_q^2 < \sigma_p^2$.

Figure 5 plots the distribution of $R_{q,t+1}$, with the the same parameters used to plot Figure 1. In comparison with Figure 1, the distribution of $R_{q,t+1}$ is much tighter than $R_{p,t+1}$ around its mean for a fixed T, reflecting our objective of taking a smaller risk than $R_{p,t+1}$. Although not reported here for brevity, similar patterns hold in the $N = 25$ case.

Now let us consider the expected utility of the ML rule. We have

Proposition 4: The expected utility of the QL portfolio, $R_{q,t+1}$, is:

$$E[U(\hat{w}_q)] = \mu_g - \frac{\gamma(T-2)\sigma_g^2}{2(T-N-1)} + \frac{k_1 T \psi^2 E[g_1(f_2)]}{\gamma(T-N-1)} - \frac{k_1^2 T(T-2)E[g_1^2(f_1)f_1]}{2\gamma(T-N)(T-N-1)}. \tag{27}$$

In comparison with Proposition 2, the first two terms are the same as they should be, which is due to the contribution of holding the same global minimum-variance portfolio. Numerically, it is easy to verify that the third term minus the fourth one here are larger, due to the minimization of the estimation errors with the use of the QL rule.

To assess the improvement of the QL rule numerically, Figure 3 and 4 plot its expected along with the others. When $N = 10$, Figure 3 shows that the QL expected utility is positive and achieves about half of the true utility already. However, when $N = 25$, Figure 4 shows that the QL expected utility, though positive, achieves only a small fraction of the true utility. In the cases, while QL expected utility improves the ML substantially, the speed of convergence to the optimal utility is still slow. Even when the sample size is as large as 600, the difference remains significant. This may be due to the inherited problem of parameter uncertainty or the possibility for improving the QL rule. While extensions of the QL seem very difficult, we, in the next subsection, study an alternative.

2.4 The UL Rule

Consider now again a two-fund rule of the type,

$$\hat{w}(\tau) = \hat{w}_g + \frac{\tau}{\gamma}\hat{w}_z. \tag{28}$$

Our objective is to find the best τ that maximizes the following empirical utility,

$$U^e(\tau) = E[\hat{w}(\tau)' R_{T+1}] - \frac{\gamma}{2}\mathrm{Var}[\hat{w}(\tau)' R_{T+1}]. \tag{29}$$

In contrast to the usual expected utility under parameter uncertainty, (12), here the utility is computed over the mean and variance of the estimated optimal portfolio.

We call $U^e(\tau)$ empirical utility because it is exactly the utility computed in almost all empirical studies, such as DeMiguel, Garlappi, and Uppal (2009) in comparing portfolio rules and Hong, Tu and Zhou (2007) in assessing asymmetries. Mathematically, it is the utility after taking over all the uncertainties of the portfolio, the asset risk and estimation risk, and then we evaluation the utility. Theoretical studies, such as Ferson and Siegel (2001) who study the uncertainty in signals and not the estimation risk here, also use this type of utility functions.

Since the empirical utility is evaluated without conditional on first the portfolio weights, it may be interpreted as the unconditional utility. Hence, we call the portfolio rule associated with maximizing $U^e(\tau)$, which is minimizes an unconditional loss function, the UL rule in what follows.

It can be shown that the optimal choice of τ is

$$\tau^* = \frac{(T-N)(T-N-1)(T-N-3)\psi^2}{(N-1)(T-2)(T-N-1)+(T+1)(T-2)(T-N-1)\psi^2+2T(T-N)\psi^4}. \tag{30}$$

Replacing ψ^2 by $\hat{\psi}_a^2$ as before, we get an estimator, $\hat{\tau}$, of τ^*. Then the portfolio weights of the UL rule is well defined by

$$\hat{w}_u = \hat{w}_g + \frac{\hat{\tau}}{\gamma}\hat{w}_z. \tag{31}$$

Denote the associated portfolio return by $R_{u,t+1}$. We have

Proposition 5: The mean and variance of $R_{u,t+1}$, the estimated optimal portfolio based on the UL rule, are:

$$\mu_u = \mu_g + \frac{k_2 T\psi^2}{\gamma(T-N-1)}E[g_2(f_2)], \tag{32}$$

$$\sigma_u^2 = \frac{(T-2+\psi^2)\sigma_g^2}{T-N-1} + \frac{Tk_2^2(T-2+\psi^2)E[g_2^2(f_1)f_1]}{\gamma^2(T-N)(T-N-1)} - \frac{T^2k_2^2\psi^4E^2[g_2(f_2)]}{\gamma^2(T-N-1)^2} + \frac{Tk_2^2\psi^2\left(E[g_2^2(f_3)]+T\psi^2E[g_2^2(f_4)]\right)}{\gamma^2(T-N)(T-N-3)}, \tag{33}$$

where $k_2 = (T-N)(T-N-1)(T-N-3)$ and the g_2 function is defined by

$$g_2(x) = \frac{x}{(T-2)(T-N-1)[N-1+(T+1)x]+2T(T-N)x^2}.$$

In contrast with Proposition 3, the UL rule has the same expressions except that now the previous k_1 and g_1 are replaced by k_2 and g_2, respectively. Despite of the complexity of the formulas, their values are fairly close to each other numerically.

Similar to Figure 5, Figure 6 plots the distribution of $R_{u,t+1}$ for $T = 60, 120$ and infinity. The two figures are virtually identical. The results may not be too surprising. Intuitively, the difference between the standard expected utility under parameter uncertainty and the empirical utility cannot be too large as they are both the criteria for investors to make optimal decisions under parameter uncertainty. Their small difference should imply small disagreements in the criteria. Another reason for their small difference is that both criteria must converge to the same standard utility of knowing true parameters, as the sample size goes to infinity.

Since the UL rule is motivated from the empirical utility, it is of interest to know what level of the empirical utility it can achieve. We have

Proposition 6: The empirical utility of the UL portfolio is:

$$
\begin{aligned}
U^e(\hat{w}_u) = \mu_g &- \frac{\gamma(T-2+\psi^2)\sigma_g^2}{2(T-N-1)} + \frac{Tk_2\psi^2 E[g_2(f_2)]}{\gamma(T-N-1)} + \frac{h^2k_2^2\psi^4E^2[g_2(f_2)]}{2\gamma(T-N-1)^2} \\
&- \frac{Tk_2^2(T-2+\psi^2)E[g_2^2(f_1)f_1]}{2\gamma(T-N)(T-N-1)} - \frac{Tk_2^2\psi^2(E[g_2^2(f_3)]+T\psi^2E[g_2^2(f_4)])}{2\gamma(T-N)(T-N-3)}.
\end{aligned}
\tag{34}
$$

Table 1 provides a comparison of all the rules in term of empirical utility (the $1/N$ rule in the table will be discussed in the next subsection). In Panel A, there are $N = 10$ assets where the assumed true parameters are the same those used in Figure 1, calibrated from monthly returns of value-weighted momentum decile portfolios from January 1927 to December 2014 are used. The true optimal portfolio has an empirical utility value of 1.43, which, independent of T, is not subject to estimations errors. For a typical sample size of 5 or 10 years data, the estimated ML rule performs poorly to have negative utility values. In contrast, the QL and UL perform well even when $T = 60$, yielding about half of the true utility level. As T increases to 240, the ML rule does much better now, with a positive utility of 0.67, about half of what is possible. But QL and UL achieve a level of 1.07. Theoretically, UL dominates QL by design as it is the optimal rule in the class to yield the greatest empirical utility. However, their values are identical under those calibrated parameters, and the only slight difference occurs when $T = 60$. When the sample used for estimation goes up to 960, there are virtually no differences among the three estimated rules. In this case, the data seem enough to provide good estimates of the parameters, and the ML rule is adequate.

Panel B of the table provides the results with $N = 25$. When $T = 60$, N is almost half of T. As we discussed earlier, the estimation errors are large in this case. Indeed, the ML rule has a huge negative value of -38.38%. Since its expected return is higher than the expected return of the true optimal portfolio, the negative utility value reflects huge risk of the ML portfolio resulted from estimation risk. Moreover, the ML rule still has a negative utility value until $T = 480$. In contrast, the QL and UL rules performs well with positive utility values of 0.26% and 0.27%, respectively, when T is as small as $T = 60$. However, these values are much lower than those for the $N = 10$ case (accounting for the difference in the true optimal portfolios) due to the increase of estimation errors when N is large relative to T. Overall, the ML rule can perform quite poorly when the sample size is less than 240, and the QL and UL rules dominate the ML rule in all cases.

2.5 The $1/N$ Rule

Since the $1/N$ has received a lot of attention recently due in part to DeMiguel, Garlappi, and Uppal (2009), and since it is one of the major rules to be compared below, we here formally introduce it and summarize its properties.

The $1/N$ rule invests $1/N$ of one dollar equally into all the risky assets, so the portfolio weights are

$$w_e = 1_N/N. \tag{35}$$

The associated expected utility is then

$$U(w_e) = \mu' w_e - \frac{\gamma}{2} w_e' \Sigma w_e, \tag{36}$$

where θ is the excess expected asset return.

In comparison with the true optimal expected utility, equation (reftrueU) or

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

the $1/N$ can perform quite poorly. This especially occurs when γ is small. As $\psi \neq 0$ is in general, then the true optimal utility can be very large for the very aggressive investors due to the third term above, but utility of the $1/N$ barely increases as γ becomes smaller.

In general, the true optimal portfolio weights can be quite different from the $1/N$ rule. In this case, it is clear that the $1/N$ will under-perform the true portfolio weights. However, in comparison with the estimated portfolio rules, the ML, QL and UL, it is unclear whether $1/N$ will do better or worse as the latter rules are subject to estimation errors. Nevertheless, it is important to point out that the estimated portfolio rules are *consistent* in that they all converge to the true optimal portfolio weights as the sample size goes to infinity. In contrast, the $1/N$ is data independent whose performance is fixed, and will eventually under-perform unless the true optimal portfolio weights happens to be exactly equal to the $1/N$.

However, under ceratin restrictive conditions, the $1/N$ can perform well. For example, consider the case when all the assets have identical variance σ^2. Then the true optimal expected utility simplifies to

$$U(w^*) = \mu' w_e - \frac{\gamma}{2N}\sigma^2 + \frac{\psi^2}{2\gamma},$$

so the first two terms are now identical, and the only difference occurs in the third term. Note that the third term is zero if all the asset expected return are the same.

So, sometimes when the true optimal portfolio weights happen to be close to the $1/N$ rule, the $1/N$ will perform well and likely to outperform the estimated rules. Therefore, it is an empirical matter whether or not the estimated rules can outperform the $1/N$ in realistic scenarios.

Table 1 provides the expected utility (which is identical to the empirical utility in the $1/N$ case because there are no estimation risk) under those parameters calibrated earlier. In either the $N = 10$ or $N = 25$ case, the $1/N$ rule outperforms the ML rule when the sample size T is less than 120. However, the ML rule eventually performs better than the $1/N$ rule when $T = 480$ or greater. But the QL and UL rules outperform the $1/N$ rule in all cases except one when $T = 60$ and $N = 25$. Again, a caveat is that if the true optimal portfolio implies by the calibrated parameters are close to the $1/N$, it will be hard for any estimated rule to outperform the $1/N$. But as along as they are different, the estimated rules will eventually dominate the $1/N$ as they will converge to the true optimal portfolio when the sample size gets larger and larger.

3. Empirical Applications

In this section, we apply the various rules of the previous section to six real data sets and compare their performances.[2]

The data sets are all those used by DeMiguel, Garlappi, and Uppal (2009) except an unavailable

[2]For brevity, we omit the comparison with rules of Jorion (1986), MacKinlay and Ľuboš Pástor (2000), and Tu and Zhou (2011). In the no riskless asset case here, if the UL rule is combined with the $1/N$, the result changes little.

one.[3] The first set is the standard 10 industry portfolios: Consumer-Discretionary, Consumer-Staples, Manufacturing, Energy, High-Tech, Telecommunication, Wholesale and Retail, Health, Utilities, and Others. The second set is returns in 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. The third set is the Fama-French three factors, Mkt, SMB and HML. The fourth set is the Fama-French 20 portfolios sorted by size and book-to-market plus the market factor. The fifth set is an extension of the fourth by adding the HML, SMB, and the momentum factors. The six set is Fama-French (1993) 25 portfolios and factors, which replaces the unavailable set. Exceptional the international data available from January 1970 to December 2014, all other data are from Kenneth Frenchs Web site from January 1929 to December 2014.

Following DeMiguel, Garlappi, and Uppal (2009) and Tu and Zhou (2014), we use a rolling estimation approach with two estimation windows of length $M = 120$ and 240 months, respectively. Specifically, in each month t, starting from $t = M$, we use the data in the most recent M months up to month t to compute the various portfolio rules and the associated portfolios. For example, if $w_{z,t}$ be the estimated optimal portfolio rule in month t for a given rule 'z', and if r_{t+1} is the realized in month $t+1$. The realized excess return on the portfolio is $r_{z,t+1} = w_{z,t}^{'} r_{t+1}$. Then we can compute the average return of the $T-M$ realized returns, $\hat{\mu}_z$, and the standard deviation, $\hat{\sigma}_z$.

The *certainty-equivalent* return or empirical utility is thus given by

$$CER_z = \hat{\mu}_z - \frac{\gamma}{2}\hat{\sigma}_z^2, \tag{37}$$

which can be interpreted as the return that an investor earns by using portfolio rule z. Clearly the greater the value of the CER, the better the rule. As before, we set the risk aversion coefficient γ to 3.

Table 2 reports the empirical results. Consider first the ML rule. When $M = 120$, the certainty-equivalent returns have huge negative numbers except for the third data set for which it has a value of 0.10. When the rolling window increases to $M = 240$, the ML rule performs much better, but still are all negative except for the third data set. Consistent with the findings in the riskless asset case of Kan and Zhou (2007), MacKinlay and Ľuboš Pástor (2000), DeMiguel, Garlappi, and Uppal (2009) and Tu and Zhou (2014), and many others, we find that the standard sample ML estimates

[3]The unavailable set is S&P sectors (see their Table 4) previously constructed by Roberto Wessels.

of the optimal portfolio rule performs poorly in practice, in the important case of investing in risky assets only.

The QL and UL rules perform much better than the ML rule. Their values are always positive and sizable for the last three data sets no matter $M = 120$ or 240. Consistent with our earlier theoretical analysis, they dominate the ML rule in the empirical applications. The results suggest strongly that the QM and UL rules are useful portfolio strategies for improving the ML rule. As expected from theory, the UL rule should perform better the QL by design. However, their differences are application dependent, and are small here for the 6 data sets.

Finally, consider the performance of the $1/N$ rule. In their influential study, DeMiguel, Garlappi, and Uppal (2009) find the $1/N$ is hard to beat in typical sample sizes by existing rules. While Tu and Zhou (2011) provide alternative rules to overcome their concern, nothing is known in the no riskless asset case. We show theoretically earlier that QL and UL rules can be good rules to outperform the $1/N$ in many cases, though not all. Empirically, our results below also show that the UL and QL perform well relative to the $1/N$.

Indeed, when the estimation window is as short as $M = 120$, the QL and UL performs better than the $1/N$ in four of the data sets. When $M = 240$, the performances of the QL and UL rules generally improve and they beat the $1/N$ in all cases except one. The exceptional case is on the third data set where the ex post optimal portfolio (estimated based on all the data) happens to be close to the $1/N$. Hence, it is very difficult, if not impossible, for any estimated rule to beat the $1/N$.

In short, in the no riskless asset case, the $1/N$ is still a good benchmark for portfolio rules and it beats the standard ML rule in the empirical applications here and in our theoretical analysis earlier when the estimation window is about 120. Nevertheless, our proposed new portfolio strategies, the QL and UL rules, can generally outperform the $1/N$ even with the 120 estimation window. This suggests there is value added of investment theory to the naive $1/N$ rule. However, our rules cannot always beat the $1/N$ in certain situations, especially when the true optimal true happens to be close to the $1/N$. How to detect such situations ex ante and how to improve the QL and UL rules further in these situations is an important question for future research.

4. Conclusion

In this paper, we provide new portfolios strategies that mitigate parameter estimation risk in the mean-variance model pioneered by Markowitz (1952) and widely used in practice. The strategies are especially suitable for the more important practical case of managing equity funds without using a riskless asset. We find that the new strategies are reliably for typically samples sizes while the usual plug-in estimated rule can often fail to perform adequately. This is true not only theoretically, but also empirically when the rules are applied to real data sets.

We also compare our new strategies with the $1/N$ in the no riskless asset case. While DeMiguel, Garlappi, and Uppal (2009) find in the riskless case 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", we show that, in the common no riskless asset case, our new strategies can in general outperform the $1/N$ theoretically for typical sizes of estimation window. Empirically, we find that they can indeed perform well relative to the $1/N$ in 6 real data sets. However, solutions to estimation risk are complex and more research is called for. For example, in addition to the estimation risk, it will be of theoretical interest to extend our study to account for time varying expected returns and risks, beyond the scope of the Markowitz framework.

Appendix A: Proofs

A.1. Proof of Proposition 1

We compute the mean and variance in two steps. First, we simplify them conditional on the estimated weights. Then, based on the simplifications, we obtain the unconditional moments analytically.

Denote for brevity $\hat{p}$ as $R_{p,T+1}$. The conditional mean and variance of $\hat{p}$ at time t can be written as

$$\mu_{\hat{p},t} = \frac{1_N'\hat{\Sigma}^{-1}\mu}{1_N'\hat{\Sigma}^{-1}1_N}\left(1 - \frac{1_N'\hat{\Sigma}^{-1}\hat{\mu}}{\gamma}\right) + \frac{\mu'\hat{\Sigma}^{-1}\hat{\mu}}{\gamma}, \tag{A1}$$

$$\sigma^2_{\hat{p},t} = \frac{1_N'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N}{(1_N'\hat{\Sigma}^{-1}1_N)^2}\left(1 - \frac{1_N\hat{\Sigma}^{-1}\hat{\mu}}{\gamma}\right)^2 + \frac{2}{\gamma}\frac{1_N'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\mu}}{1_N'\hat{\Sigma}^{-1}1_N}\left(1 - \frac{1_N'\hat{\Sigma}^{-1}\hat{\mu}}{\gamma}\right) + \frac{\hat{\mu}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\mu}}{\gamma^2}. \tag{A2}$$

Let P be an $N \times N$ orthonormal matrix with the first two columns being ν and η, $P = [\nu,\ \eta,\ P_1]$, where

$$\nu = \frac{\Sigma^{-\frac{1}{2}}1_N}{\left(1_N'\Sigma^{-1}1_N\right)^{\frac{1}{2}}} = \sigma_g\Sigma^{-\frac{1}{2}}1_N, \tag{A3}$$

$$\eta = \frac{(I_N - \nu\nu')\Sigma^{-\frac{1}{2}}\mu}{\left[\mu'\Sigma^{-\frac{1}{2}}(I_N - \nu\nu')\Sigma^{-\frac{1}{2}}\mu\right]^{\frac{1}{2}}} = \frac{\Sigma^{-\frac{1}{2}}(\mu - 1_N\mu_g)}{\psi}. \tag{A4}$$

Define

$$z = \sqrt{h}P'\Sigma^{-\frac{1}{2}}\hat{\mu} \sim N(\mu_z, I_N), \tag{A5}$$

$$W = hP'\Sigma^{-\frac{1}{2}}\hat{\Sigma}\Sigma^{-\frac{1}{2}}P \sim W_N(h-1, I_N), \tag{A6}$$

which are independent of each other. It is easy to show that

$$\mu_z = \sqrt{h}P'\Sigma^{-\frac{1}{2}}\mu = \begin{bmatrix} \sqrt{h}\theta_g \\ \sqrt{h}\psi \\ 0_{N-2} \end{bmatrix} = \sqrt{h}\theta_g e_1 + \sqrt{h}\psi e_2, \tag{A7}$$

where $\theta_g = \mu_g/\sigma_g$; and e_1 and e_2 are the unit vectors. Denote the first two elementns of z as z_1 and z_2, we have then

$$z_1 \sim N(\sqrt{h}\theta_g, 1), \tag{A8}$$

$$z_2 \sim N(\sqrt{h}\psi, 1), \tag{A9}$$

$$z'z = u = z_1^2 + u_1 = z_1^2 + z_2^2 + u_0, \tag{A10}$$

where $u_0 \sim \chi^2_{N-2}$, $u_1 \sim \chi^2_{N-1}(h\psi^2)$, and $u \sim \chi^2_N(h\theta_g^2 + h\psi^2)$.

In terms of z and W, we obtain the conditional moments as

$$\mu_{\hat{p},t} = \mu_g + \sigma_g \psi \frac{e_1' W^{-1} e_2}{e_1' W^{-1} e_1} + \frac{\psi\sqrt{h}}{\gamma}\left(e_2' W^{-1} z - \frac{e_1' W^{-1} e_2}{e_1' W^{-1} e_1} e_1' W^{-1} z\right), \tag{A11}$$

$$\begin{aligned}\sigma^2_{\hat{p},t} = {} & \frac{\sigma_g^2 (e_1' W^{-2} e_1)}{(e_1' W^{-1} e_1)^2}\left(1 - \frac{\sqrt{h} e_1' W^{-1} z}{\gamma\sigma_g}\right)^2 + \frac{2\sigma_g \sqrt{h} e_1' W^{-2} z}{\gamma e_1' W^{-1} e_1}\left(1 - \frac{\sqrt{h} e_1' W^{-1} z}{\gamma\sigma_g}\right) \\ & + \frac{h z' W^{-2} z}{\gamma^2}.\end{aligned} \tag{A12}$$

Let

$$\xi = \frac{(I_N - e_1 e_1')z}{\left[z'(I_N - e_1 e_1')z\right]^{\frac{1}{2}}} = \frac{(I_N - e_1 e_1')z}{\sqrt{u_1}}$$

be the first two columns of $Q = [e_1,\ \xi,\ Q_1]$ and define

$$A = (Q' W^{-1} Q)^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \sim W_N(h-1, I_N), \tag{A13}$$

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

$$A_{11\cdot 2} \equiv A_{11} - A_{12} A_{22}^{-1} A_{21} \sim W_2(h - N + 1, I_2), \tag{A14}$$

$$\mathrm{vec}(y) \equiv \mathrm{vec}(-A_{22}^{-\frac{1}{2}} A_{21}) \sim N(0_{2N-4}, I_{2N-4}), \tag{A15}$$

$$A_{22} \sim W_{N-2}(h-1, I_{N-2}), \tag{A16}$$

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

$$A_{11\cdot 2} = \begin{pmatrix} v_1 + a^2 & -a\sqrt{v_2} \\ -a\sqrt{v_2} & v_2 \end{pmatrix}, \tag{A17}$$

where $v_1 \sim \chi^2_{h-N}$, $v_2 \sim \chi^2_{h-N+1}$, and $a \sim N(0,1)$, all of which are independent of each other. Since

$$A_{11\cdot 2}^{-1} = \begin{pmatrix} \frac{1}{v_1} & \frac{a}{v_1\sqrt{v_2}} \\ \frac{a}{v_1\sqrt{v_2}} & \frac{1}{v_2} + \frac{a^2}{v_1 v_2} \end{pmatrix}, \tag{A18}$$

it follows that

$$e_1' W^{-1} e_1 = \frac{1}{v_1}, \tag{A19}$$

$$e_1' W^{-1} \xi = \frac{a}{v_1\sqrt{v_2}}, \tag{A20}$$

$$\xi' W^{-1} \xi = \frac{1}{v_2} + \frac{a^2}{v_1 v_2}. \tag{A21}$$

Using the definition of ξ, we obtain

$$e_1' W^{-1} z = \sqrt{u_1} e_1' W^{-1} \xi + \frac{z_1}{v_1} = \frac{a\sqrt{u_1}}{v_1\sqrt{v_2}} + \frac{z_1}{v_1}. \tag{A22}$$

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

$$\iota = \frac{(I_N - e_1 e_1' - \xi\xi')e_2}{\left[e_2'(I_N - e_1 e_1' - \xi\xi')e_2\right]^{\frac{1}{2}}} = \frac{(I_N - \xi\xi')e_2}{\sqrt{u_0/u_1}}. \tag{A23}$$

Using the fact thst

$$Q_1' W^{-1}[e_1,\ \xi] = A_{22}^{-\frac{1}{2}} y A_{11\cdot 2}^{-1}, \tag{A24}$$

and Theorem 3.1. and Corollary 3.1. in Dickey (1967), we can write

$$A_{22}^{-\frac{1}{2}} y = x B^{-\frac{1}{2}}, \tag{A25}$$

where

$$\mathrm{vec}(x) = \mathrm{vec}([x_1,\ x_2]) \sim N(0_{2N-4}, I_{2N-4}), \tag{A26}$$

$$B \sim W_2(h-N+3, I_2), \tag{A27}$$

and x and B are independent of each other. Using again the Bartlett decomposition, we can write

$$B = LL' = \begin{pmatrix} w_1 & -b\sqrt{w_1} \\ -b\sqrt{w_1} & b^2 + w_2 \end{pmatrix}, \tag{A28}$$

where

$$L = \begin{pmatrix} \sqrt{w_1} & 0 \\ -b & \sqrt{w_2} \end{pmatrix}, \tag{A29}$$

with $w_1 \sim \chi^2_{h-N+3}$, $w_2 \sim \chi^2_{h-N+2}$, and $b \sim N(0,1)$ (which are independent of each other). Let $\varepsilon_1 = [1, 0'_{N-3}]'$, we have

$$[q_1,\, q_2] \equiv \iota' W^{-1}[e_1,\, \xi] = \varepsilon_1' x L^{-1} A_{11\cdot 2}^{-1} = [x_{11},\, x_{21}] B^{-\frac{1}{2}} A_{11\cdot 2}^{-1}, \tag{A30}$$

where

$$q_1 = \frac{1}{v_1}\left(\frac{x_{11}}{\sqrt{w_1}} + \frac{b x_{21}}{\sqrt{w_1 w_2}} + \frac{a x_{21}}{\sqrt{v_2 w_2}}\right), \tag{A31}$$

$$q_2 = \frac{a}{\sqrt{v_2}} q_1 + \frac{x_{21}}{v_2\sqrt{w_2}}, \tag{A32}$$

and $x_{11} \sim N(0,1)$ and $x_{21} \sim N(0,1)$ are the first elements of x_1 and x_2.

Now using the definition of ι, (A19), (A20), (A21), and (A30), we have

$$\begin{aligned} e_2' W^{-1} e_1 &= \frac{\sqrt{u_0}}{\sqrt{u_1}} q_1 + \frac{e_1' W^{-1}\xi}{\sqrt{u_1}} z_2 \\ &= \frac{1}{v_1}\left(\frac{x_{11}}{\sqrt{w_1}} + \frac{b x_{21}}{\sqrt{w_1 w_2}} + \frac{a x_{21}}{\sqrt{v_2 w_2}}\right)\sqrt{\frac{u_0}{u_1}} + \frac{a z_2}{v_1\sqrt{v_2}\sqrt{u_1}}, \end{aligned} \tag{A33}$$

$$\begin{aligned} e_2' W^{-1} z &= \sqrt{u_0} q_2 + e_2' W^{-1} e_1 z_1 + \xi' W^{-1} \xi z_2 \\ &= \frac{a\sqrt{u_0}}{v_1\sqrt{v_2}}\left(\frac{x_{11}}{\sqrt{w_1}} + \frac{b x_{21}}{\sqrt{w_1 w_2}} + \frac{a x_{21}}{\sqrt{v_2 w_2}}\right) + \frac{x_{21}\sqrt{u_0}}{v_2\sqrt{w_2}} + e_2' W^{-1} e_1 z_1 \\ &\quad + \left(\frac{1}{v_2} + \frac{a^2}{v_1 v_2}\right) z_2. \end{aligned} \tag{A34}$$

Because of these two equations, we also obtain

$$e_2' W^{-1} z - \frac{e_2' W^{-1} e_1}{e_1' W^{-1} e_1} e_1' W^{-1} z = \frac{1}{v_2}\left(\frac{\sqrt{u_0}}{\sqrt{w_2}} x_{21} + z_2\right). \tag{A35}$$

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

$$\begin{aligned} \begin{bmatrix} e_1' \\ \xi' \end{bmatrix} W^{-2}[e_1,\, \xi] &= \begin{bmatrix} e_1' \\ \xi' \end{bmatrix} W^{-1}\left([e_1,\, \xi]\begin{bmatrix} e_1' \\ \xi' \end{bmatrix} + Q_1 Q_1'\right) W^{-1}[e_1,\, \xi] \\ &= A_{11\cdot 2}^{-2} + A_{11\cdot 2}^{-1} B^{-\frac{1}{2}'} x' x B^{-\frac{1}{2}} A_{11\cdot 2}^{-1}. \end{aligned} \tag{A36}$$

Note that $x'x$ can be written as

$$x'x = \begin{pmatrix} x_{11}^2 & x_{11}x_{21} \\ x_{11}x_{21} & x_{21}^2 \end{pmatrix} + C, \tag{A37}$$

where $C \sim W_2(N-3, I_2)$, independent of x_{11} and x_{21}. Using the Bartlett decomposition, we can write

$$C = \begin{pmatrix} s_1 + c^2 & c\sqrt{s_2} \\ c\sqrt{s_2} & s_2 \end{pmatrix}, \tag{A38}$$

where $s_1 \sim \chi^2_{N-4}$, $s_2 \sim \chi^2_{N-3}$, and $c \sim N(0,1)$, and they are independent of each other. We have

$$\begin{aligned}
& I_2 + B^{-\frac{1}{2}'} x'xB^{-\frac{1}{2}} \\
=& I_2 + \begin{pmatrix} \frac{1}{\sqrt{w_1}} & \frac{b}{\sqrt{w_1 w_2}} \\ 0 & \frac{1}{\sqrt{w_2}} \end{pmatrix} \begin{pmatrix} x_{11}^2 + s_1 + c^2 & x_{11}x_{21} + c\sqrt{s_2} \\ x_{11}x_{21} + c\sqrt{s_2} & x_{21}^2 + s_2 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{w_1}} & 0 \\ \frac{b}{\sqrt{w_1 w_2}} & \frac{1}{\sqrt{w_2}} \end{pmatrix} \\
=& \begin{pmatrix} 1 + \frac{s_1}{w_1} + \left(\frac{x_{11}}{\sqrt{w_1}} + \frac{bx_{21}}{\sqrt{w_1 w_2}}\right)^2 + \left(\frac{c}{\sqrt{w_1}} + \frac{b\sqrt{s_2}}{\sqrt{w_1 w_2}}\right)^2 & \frac{x_{21}}{\sqrt{w_2}}\left(\frac{x_{11}}{\sqrt{w_1}} + \frac{bx_{21}}{\sqrt{w_1 w_2}}\right) + \frac{\sqrt{s_2}}{\sqrt{w_2}}\left(\frac{c}{\sqrt{w_1}} + \frac{b\sqrt{s_2}}{\sqrt{w_1 w_2}}\right) \\ \frac{x_{21}}{\sqrt{w_2}}\left(\frac{x_{11}}{\sqrt{w_1}} + \frac{bx_{21}}{\sqrt{w_1 w_2}}\right) + \frac{\sqrt{s_2}}{\sqrt{w_2}}\left(\frac{c}{\sqrt{w_1}} + \frac{b\sqrt{s_2}}{\sqrt{w_1 w_2}}\right) & 1 + \frac{x_{21}^2 + s_2}{w_2} \end{pmatrix}.
\end{aligned} \tag{A39}$$

The above results allow us to obtain

$$\begin{aligned}
e_1'W^{-2}e_1 =& \left[\frac{1}{v_1}, \ \frac{a}{v_1\sqrt{v_2}}\right] \left(I_2 + B^{-\frac{1}{2}'} x'xB^{-\frac{1}{2}}\right) \begin{bmatrix} \frac{1}{v_1} \\ \frac{a}{v_1\sqrt{v_2}} \end{bmatrix} \\
=& \frac{1}{v_1^2}\left(\frac{x_{11}}{\sqrt{w_1}} + \frac{bx_{21}}{\sqrt{w_1 w_2}} + \frac{ax_{21}}{\sqrt{v_2 w_2}}\right)^2 + \frac{1}{v_1^2}\left(\frac{c}{\sqrt{w_1}} + \frac{b\sqrt{s_2}}{\sqrt{w_1 w_2}} + \frac{a\sqrt{s_2}}{\sqrt{v_2 w_2}}\right)^2 \\
&+ \frac{1}{v_1^2}\left(1 + \frac{s_1}{w_1} + \frac{a^2}{v_2}\right),
\end{aligned} \tag{A40}$$

$$\begin{aligned}
\xi'W^{-2}\xi =& \left[\frac{a}{v_1\sqrt{v_2}}, \ \frac{1}{v_2} + \frac{a^2}{v_1 v_2}\right] \left(I_2 + B^{-\frac{1}{2}'} x'xB^{-\frac{1}{2}}\right) \begin{bmatrix} \frac{a}{v_1\sqrt{v_2}} \\ \frac{1}{v_2} + \frac{a^2}{v_1 v_2} \end{bmatrix} \\
=& \left[\frac{a}{v_1\sqrt{v_2}}\left(\frac{x_{11}}{\sqrt{w_1}} + \frac{bx_{21}}{\sqrt{w_1 w_2}} + \frac{ax_{21}}{\sqrt{v_2 w_2}}\right) + \frac{x_{21}}{v_2\sqrt{w_2}}\right]^2 + \frac{a^2}{v_1^2 v_2}\left(1 + \frac{s_1}{w_1}\right) \\
&+ \left[\frac{a}{v_1\sqrt{v_2}}\left(\frac{c}{\sqrt{w_1}} + \frac{b\sqrt{s_2}}{\sqrt{w_1 w_2}} + \frac{a\sqrt{s_2}}{\sqrt{v_2 w_2}}\right) + \frac{\sqrt{s_2}}{v_2\sqrt{w_2}}\right]^2 + \left(\frac{1}{v_2} + \frac{a^2}{v_1 v_2}\right)^2
\end{aligned} \tag{A41}$$

and

$$
\begin{aligned}
e_1'W^{-2}\xi &= \left[\frac{1}{v_1},\ \frac{a}{v_1\sqrt{v_2}}\right]\left(I_2+B^{-\frac{1}{2}'}x'xB^{-\frac{1}{2}}\right)\begin{bmatrix} \frac{a}{v_1\sqrt{v_2}} \\ \frac{1}{v_2}+\frac{a^2}{v_1v_2} \end{bmatrix} \\
&= \frac{a}{v_1^2\sqrt{v_2}}\left(\frac{x_{11}}{\sqrt{w_1}}+\frac{bx_{21}}{\sqrt{w_1w_2}}+\frac{ax_{21}}{\sqrt{v_2w_2}}\right)^2+\frac{x_{21}}{v_1v_2\sqrt{w_2}}\left(\frac{x_{11}}{\sqrt{w_1}}+\frac{bx_{21}}{\sqrt{w_1w_2}}+\frac{ax_{21}}{\sqrt{v_2w_2}}\right) \\
&\quad+\frac{a}{v_1^2\sqrt{v_2}}\left(\frac{c}{\sqrt{w_1}}+\frac{b\sqrt{s_2}}{\sqrt{w_1w_2}}+\frac{a\sqrt{s_2}}{\sqrt{v_2w_2}}\right)^2+\frac{\sqrt{s_2}}{v_1v_2\sqrt{w_2}}\left(\frac{c}{\sqrt{w_1}}+\frac{b\sqrt{s_2}}{\sqrt{w_1w_2}}+\frac{a\sqrt{s_2}}{\sqrt{v_2w_2}}\right) \\
&\quad+\frac{a}{v_1^2\sqrt{v_2}}\left(1+\frac{s_1}{w_1}+\frac{a^2}{v_2}\right)+\frac{a}{v_1\sqrt{v_2^3}}. \qquad \text{(A42)}
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
e_1'W^{-2}z &= \sqrt{u_1}(e_1'W^{-2}\xi)+z_1(e_1'W^{-2}e_1) \\
&= \left(\frac{a\sqrt{u_1}}{v_1^2\sqrt{v_2}}+\frac{z_1}{v_1^2}\right)\left[\left(\frac{x_{11}}{\sqrt{w_1}}+\frac{bx_{21}}{\sqrt{w_1w_2}}+\frac{ax_{21}}{\sqrt{v_2w_2}}\right)^2+\left(\frac{c}{\sqrt{w_1}}+\frac{b\sqrt{s_2}}{\sqrt{w_1w_2}}+\frac{a\sqrt{s_2}}{\sqrt{v_2w_2}}\right)^2\right. \\
&\quad\left.+\left(1+\frac{s_1}{w_1}+\frac{a^2}{v_2}\right)\right]+\frac{x_{21}\sqrt{u_1}}{v_1v_2\sqrt{w_2}}\left(\frac{x_{11}}{\sqrt{w_1}}+\frac{bx_{21}}{\sqrt{w_1w_2}}+\frac{ax_{21}}{\sqrt{v_2w_2}}\right) \\
&\quad+\frac{\sqrt{s_2}\sqrt{u_1}}{v_1v_2\sqrt{w_2}}\left(\frac{c}{\sqrt{w_1}}+\frac{b\sqrt{s_2}}{\sqrt{w_1w_2}}+\frac{a\sqrt{s_2}}{\sqrt{v_2w_2}}\right)+\frac{a\sqrt{u_1}}{v_1\sqrt{v_2^3}} \qquad \text{(A43)}
\end{aligned}
$$

and

$$
\begin{aligned}
z'W^{-2}z &= u_1(\xi'W^{-2}\xi)+2z_1(e_1'W^{-2}z)-z_1^2(e_1'W^{-2}e_1) \\
&= \left(1+\frac{s_1}{w_1}\right)\left(\frac{a\sqrt{u_1}}{v_1\sqrt{v_2}}+\frac{z_1}{v_1}\right)^2+\left(\frac{az_1}{v_1\sqrt{v_2}}+\frac{\sqrt{u_1}}{v_2}+\frac{a^2\sqrt{u_1}}{v_1v_2}\right)^2 \\
&\quad+\left[\frac{x_{21}\sqrt{u_1}}{v_2\sqrt{w_2}}+\left(\frac{x_{11}}{\sqrt{w_1}}+\frac{bx_{21}}{\sqrt{w_1w_2}}+\frac{ax_{21}}{\sqrt{v_2w_2}}\right)\left(\frac{a\sqrt{u_1}}{v_1\sqrt{v_2}}+\frac{z_1}{v_1}\right)\right]^2 \\
&\quad+\left[\left(\frac{c}{\sqrt{w_1}}+\frac{b\sqrt{s_2}}{\sqrt{w_1w_2}}+\frac{a\sqrt{s_2}}{\sqrt{v_2w_2}}\right)\left(\frac{a\sqrt{u_1}}{v_1\sqrt{v_2}}+\frac{z_1}{v_1}\right)+\frac{\sqrt{s_2}\sqrt{u_1}}{v_2\sqrt{w_2}}\right]^2. \qquad \text{(A44)}
\end{aligned}
$$

Let

$$
y_1 = \frac{x_{11}}{\sqrt{w_1}}+\frac{bx_{21}}{\sqrt{w_1w_2}}+\frac{ax_{21}}{\sqrt{v_2w_2}}, \qquad \text{(A45)}
$$

$$
y_2 = \frac{c}{\sqrt{w_1}}+\frac{b\sqrt{s_2}}{\sqrt{w_1w_2}}+\frac{a\sqrt{s_2}}{\sqrt{v_2w_2}}, \qquad \text{(A46)}
$$

we then obtain

$$\mu_{g,t} = \mu_g + \frac{\sigma_g \psi}{\sqrt{u_1}} \left(\sqrt{u_0} y_1 + \frac{a z_2}{\sqrt{v_2}} \right), \tag{A47}$$

$$\mu_{z,t} = \frac{\sqrt{h}\psi}{v_2} \left(\frac{x_{21}\sqrt{u_0}}{\sqrt{w_2}} + z_2 \right), \tag{A48}$$

$$\sigma^2_{g,t} = \sigma^2_g \left(y_1^2 + y_2^2 + 1 + \frac{s_1}{w_1} + \frac{a^2}{v_2} \right), \tag{A49}$$

$$\sigma^2_{z,t} = \frac{h u_1}{v_2^2} \left(1 + \frac{x_{21}^2 + s_2}{w_2} \right), \tag{A50}$$

$$\sigma_{gz,t} = \frac{\sigma_g \sqrt{h}\sqrt{u_1}}{v_2} \left(\frac{a}{\sqrt{v_2}} + \frac{x_{21}}{\sqrt{w_2}} y_1 + \frac{\sqrt{s_2}}{\sqrt{w_2}} y_2 \right), \tag{A51}$$

where $\sigma_{gz,t} = w'_{g,t} \Sigma w_{z,t}$. Therefore, we can compute analytically the expectation of these terms as

$$E[\mu_{g,t}] = \mu_g, \tag{A52}$$

$$E[\mu_{z,t}] = \frac{h\psi^2}{h-N-1}, \tag{A53}$$

$$E[\sigma^2_{g,t}] = \frac{(h-2)\sigma^2_g}{h-N-1}, \tag{A54}$$

$$E[\sigma^2_{z,t}] = \frac{h(h-2)(h\psi^2+N-1)}{(h-N)(h-N-1)(h-N-3)}, \tag{A55}$$

$$E[\sigma_{gz,t}] = 0. \tag{A56}$$

Note that

$$\mu_{\hat{p},t} = \mu_{g,t} + \frac{1}{\gamma}\mu_{z,t}, \tag{A57}$$

$$\sigma^2_{\hat{p},t} = \sigma^2_{g,t} + \frac{1}{\gamma^2}\sigma^2_{z,t} + \frac{2}{\gamma}\sigma_{gz,t}. \tag{A58}$$

We have

$$
\begin{aligned}
\mu_{\hat{p},t} &= \mu_g + \frac{\sigma_g \psi}{\sqrt{u_1}}\left(\sqrt{u_0}y_1 + \frac{az_2}{\sqrt{v_2}}\right) + \frac{\sqrt{h}\psi}{\gamma v_2}\left(\frac{x_{21}\sqrt{u_0}}{\sqrt{w_2}} + z_2\right) \\
&= \mu_g + \frac{\psi z_2}{\sqrt{u_1}}\left(\frac{\sigma_g a}{\sqrt{v_2}} + \frac{\sqrt{h}\sqrt{u_1}}{\gamma v_2}\right) + \frac{\psi\sqrt{u_0}}{\sqrt{u_1}}\left(\sigma_g y_1 + \frac{\sqrt{h}x_{21}\sqrt{u_1}}{\gamma v_2\sqrt{w_2}}\right), \\
\sigma^2_{\hat{p},t} &= \left(\frac{\sigma_g a}{\sqrt{v_2}} + \frac{\sqrt{h}\sqrt{u_1}}{\gamma v_2}\right)^2 + \left(\sigma_g y_1 + \frac{\sqrt{h}x_{21}\sqrt{u_1}}{\gamma v_2\sqrt{w_2}}\right)^2 \\
&\quad + \sigma_g^2\left(1 + \frac{s_1}{w_1}\right) + \left(\sigma_g y_2 + \frac{\sqrt{h}\sqrt{s_2 u_1}}{\gamma v_2\sqrt{w_2}}\right)^2. \qquad \text{(A59)}
\end{aligned}
$$

Note that the joint distribution of $(\mu_{\hat{p},t}, \sigma^2_{\hat{p},t})$ depends on 13 independent univariate random variables: a, b, c, x_{11}, x_{21}, z_2, u_0, s_1, s_2, w_1 w_2, u_1, and v_2. However, with a simple transformation, we can replace x_{11}, b and c with two standard normal random variables.

Finally, based on above results, we can easily show that

$$E[\mu_{\hat{p},t}] = \mu_g + \frac{h\psi^2}{\gamma(h-N-1)}, \qquad \text{(A60)}$$

$$E[\sigma^2_{\hat{p},t}] = \frac{h(h-2)(h\psi^2+N-1)}{\gamma^2(h-N)(h-N-1)(h-N-3)} + \frac{(h-2)\sigma_g^2}{h-N-1}. \qquad \text{(A61)}$$

This implies Proposition 1. Q.E.D.

A.1. Proof of Proposition 2

A.1. Proof of Proposition 3

A.1. Proof of Proposition 4

A.1. Proof of Proposition 5

A.1. Proof of Proposition 6

REFERENCES

Basak, Gopal K., Ravi Jagannathan, Tongshu Ma, 2009, Jackknife Estimator for Tracking Error Variance of Optimal Portfolios, Management Science 55, 990–1002.

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

Berger, James O., 1985, *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag, New York).

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

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

Chincarini, Ludwig B., and Daehwan Kim, 2006, *Quantitative Equity Portfolio Management* (McGraw-Hill, New York).

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

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.

Ferson, Wayne E., and Andrew F. Siegel, 2001, The efficient use of conditioning information in portfolios, *Journal of Finance* 56, 967–982.

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

Lehmann, E. L., and George Casella, 1998, *Theory of Point Estimation* (Springer-Verlag, New York).

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

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

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

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

Michaud, Richard., 1998, *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation* (Harvard Business School Press, MA).

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

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

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

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

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

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
Empirical utility

This table reports empirical utility (in percentage points) of the true optimal portfolio, the ML rule, the QL rule, the UL rule, and the $1/N$ rule, respectively, for various sample sizes of the estimation window (T). The assumed true parameter values are calibrated from real data. In Panel A, monthly returns of the value-weighted momentum decile portfolios from January 1927 to December 2014 are used. In Panel B, monthly returns of Fama-French Size-yBM 25 portfolios from the same period are used. The risk aversion coefficient γ is set at 3.

T	60	120	240	480	960	2000
Panel A: Ten Momentum Portfolios						
True Optimal	1.43	1.43	1.43	1.43	1.43	1.43
ML	-3.38	-0.33	0.67	1.07	1.26	1.35
QL	0.70	0.93	1.07	1.19	1.28	1.35
UL	0.71	0.93	1.07	1.19	1.28	1.35
$1/N$	0.38	0.38	0.38	0.38	0.38	0.38
Panel B: FF 25 portfolios						
True Optimal	1.79	1.79	1.79	1.79	1.79	1.79
ML	-38.38	-5.78	-0.77	0.72	1.30	1.56
QL	0.26	0.69	0.97	1.22	1.44	1.60
UL	0.27	0.70	0.97	1.22	1.44	1.60
$1/N$	0.45	0.45	0.45	0.45	0.45	0.45

Table 2
Certainty-equivalent returns based on real data sets

This table reports the certainty-equivalent returns (in percentage points) of a mean-variance investor who uses various investment rules: the ML, the QL, the UL and the $1/N$. The estimated rules are based on a rolling sample with an estimation window $M = 120$ (Panel A) or 240 (Panel B). The real data sets are the five data sets used by DeMiguel, Garlappi, and Uppal (2009), and an additional one, the Fama-French 25 size and book-to-market portfolios and their three factors. The international data is from January 1970 to December 2014. All other data sets are from January 1929 to December 2014. The risk aversion coefficient γ is set at 3.

Rules	Industry portfolios N=11	Inter'l portfolios N=9	Mkt/ SMB/HML N=3	FF- 1-factor N=21	FF- 4-factor N=24	FF25 3-factor N=28
Panel A: M=120						
ML	-1.73	-1.14	0.10	-3.92	-7.53	-9.72
QL	0.14	0.10	0.37	1.04	1.18	0.84
UL	0.14	0.10	0.37	1.05	1.21	0.85
$1/N$	0.40	0.24	0.33	0.36	0.41	0.39
Panel B: M=240						
ML	-0.22	-0.26	0.06	-0.47	-0.84	-2.18
QL	0.46	0.13	0.16	1.12	1.39	0.73
UL	0.46	0.13	0.16	1.12	1.41	0.74
$1/N$	0.44	0.11	0.32	0.43	0.47	0.45

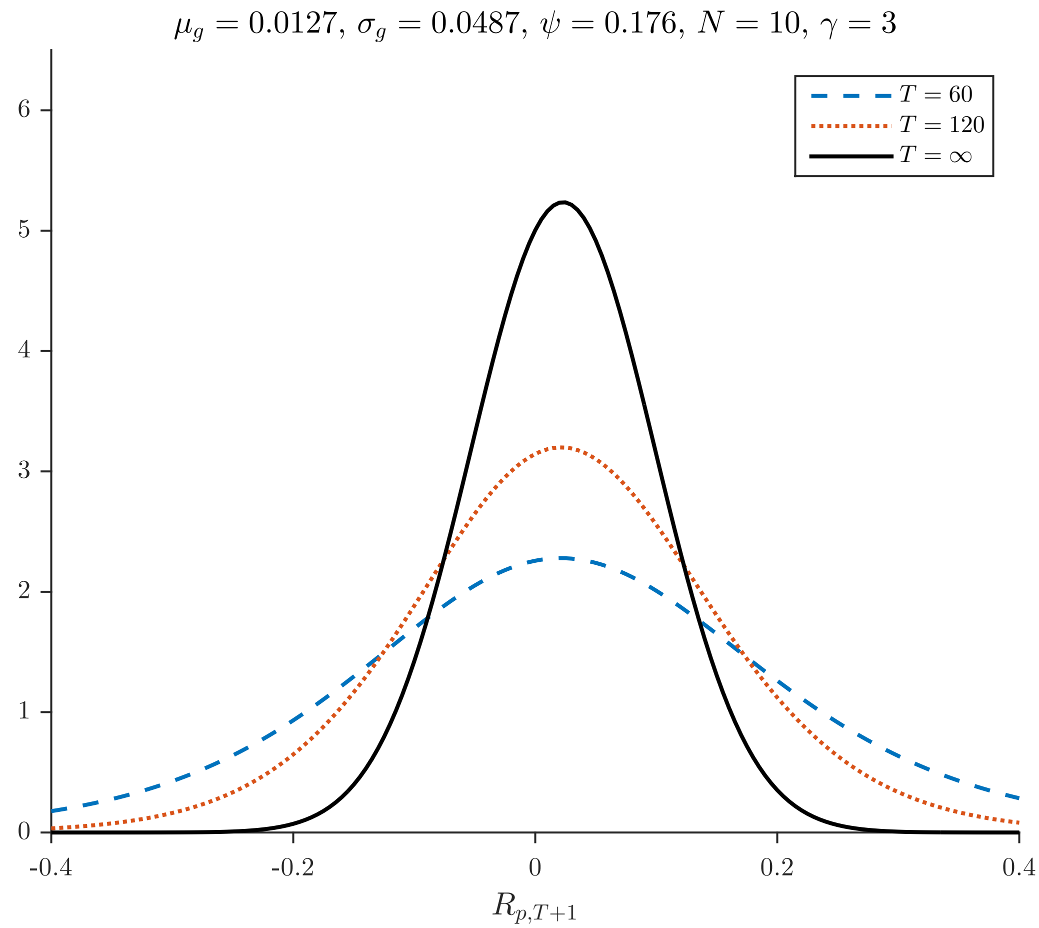

Figure 1
Distribution of out-of-sample portfolio return of the ML rule ($N = 10$)

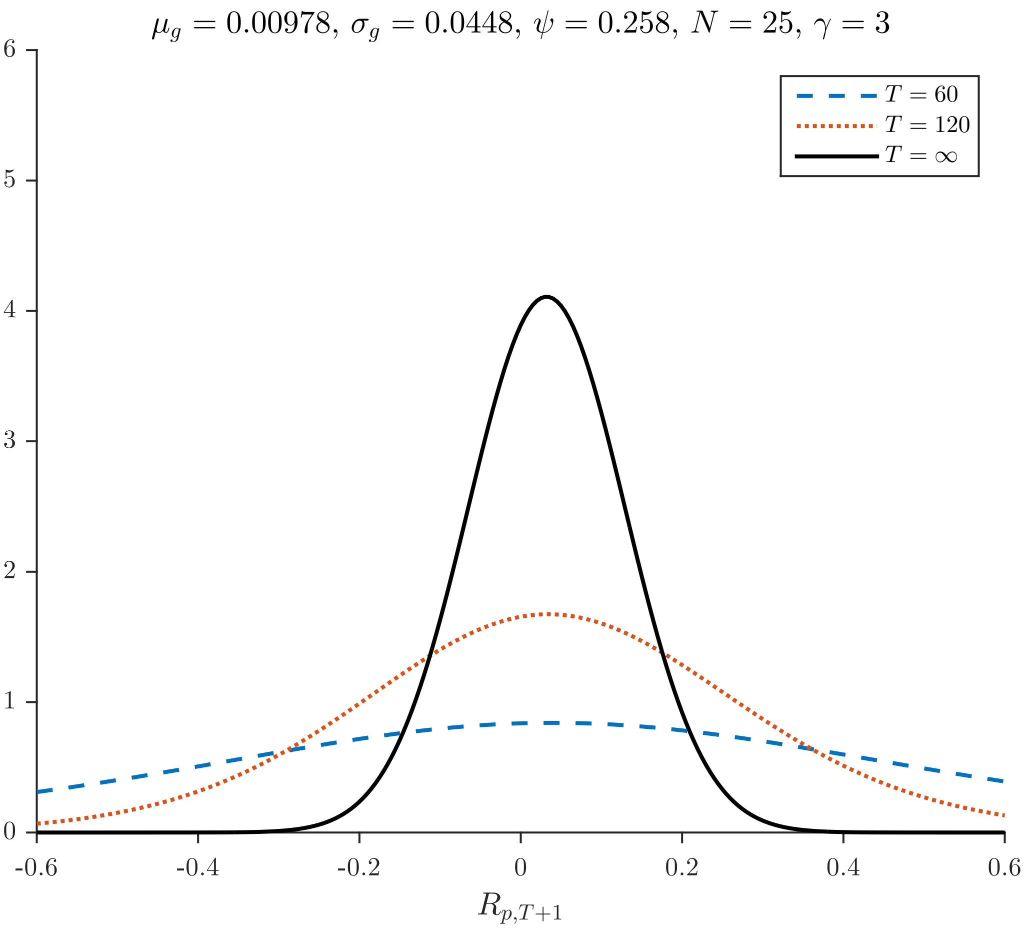

Figure 2
Distribution of out-of-sample portfolio return of the ML rule ($N = 25$)

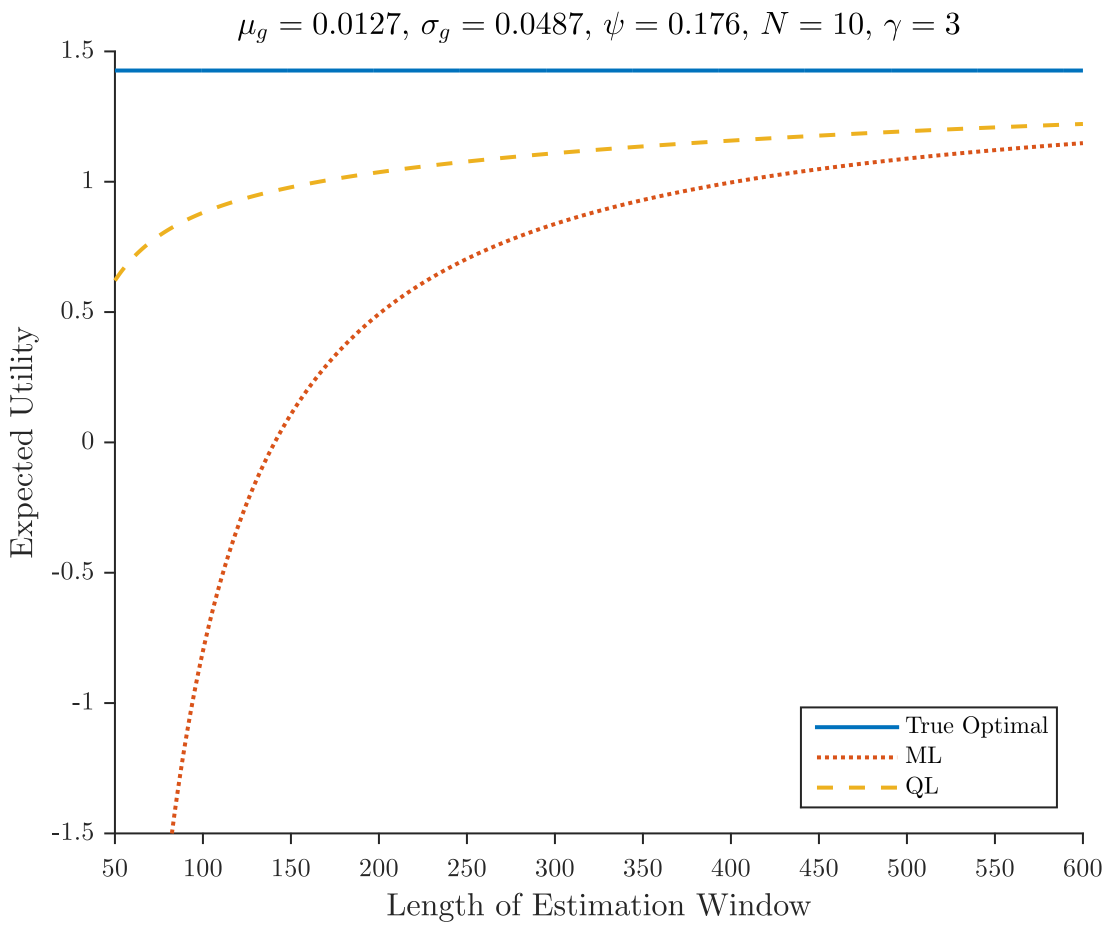

Figure 3
Expected utility ($N = 10$)

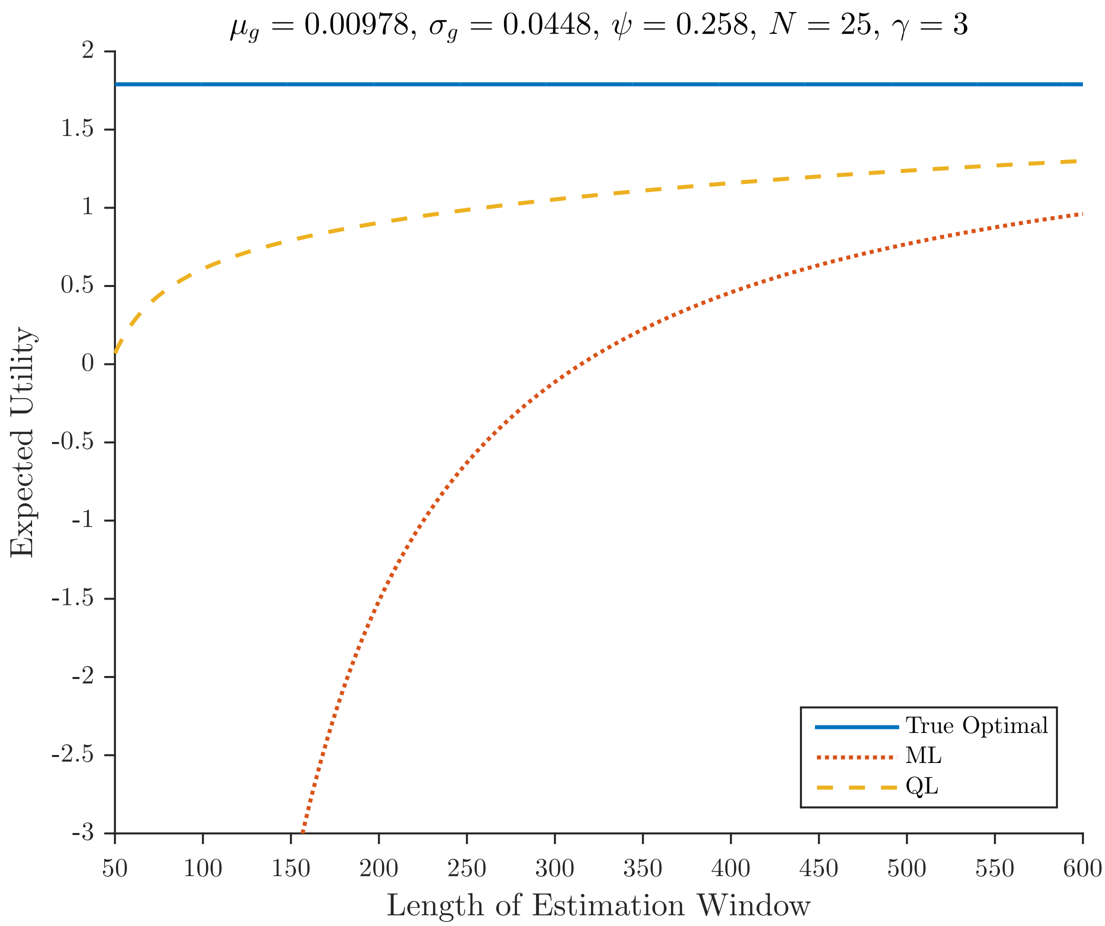

Figure 4
Expected utility ($N = 25$)

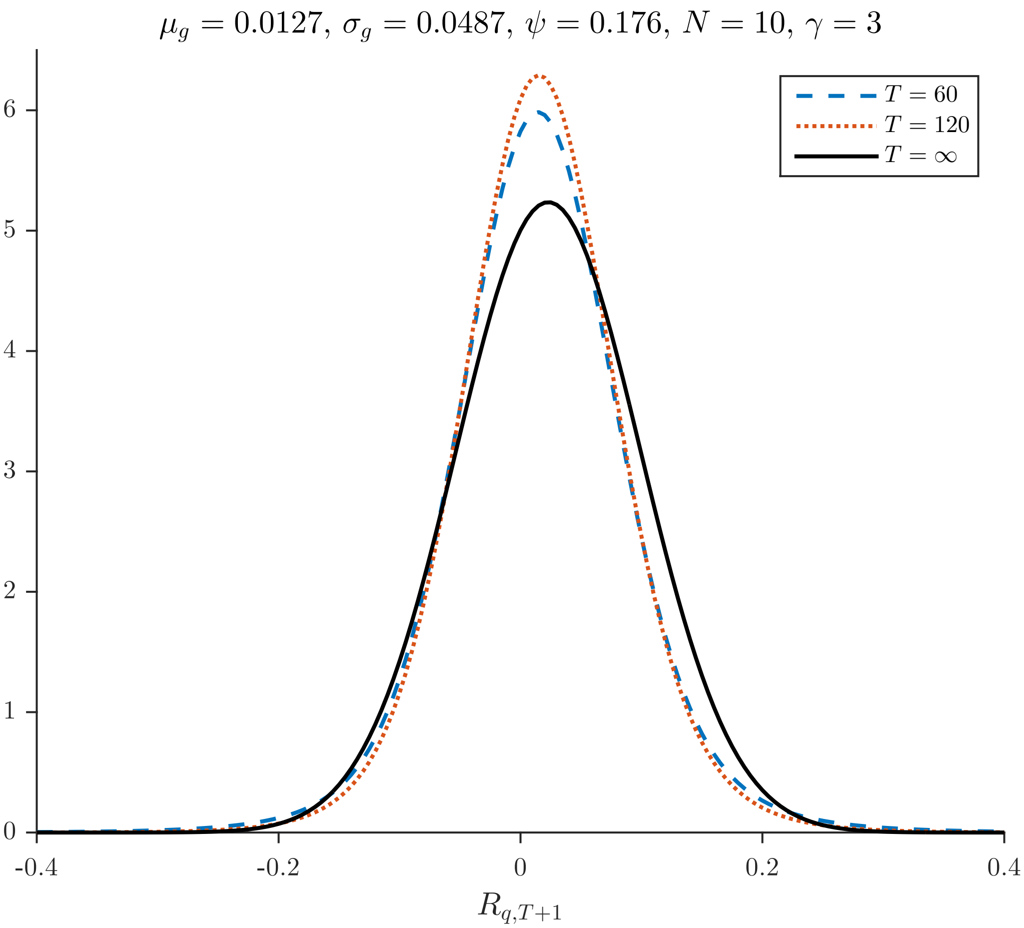

Figure 5
Distribution of out-of-sample portfolio return of the QL rule **($N = 10$)**

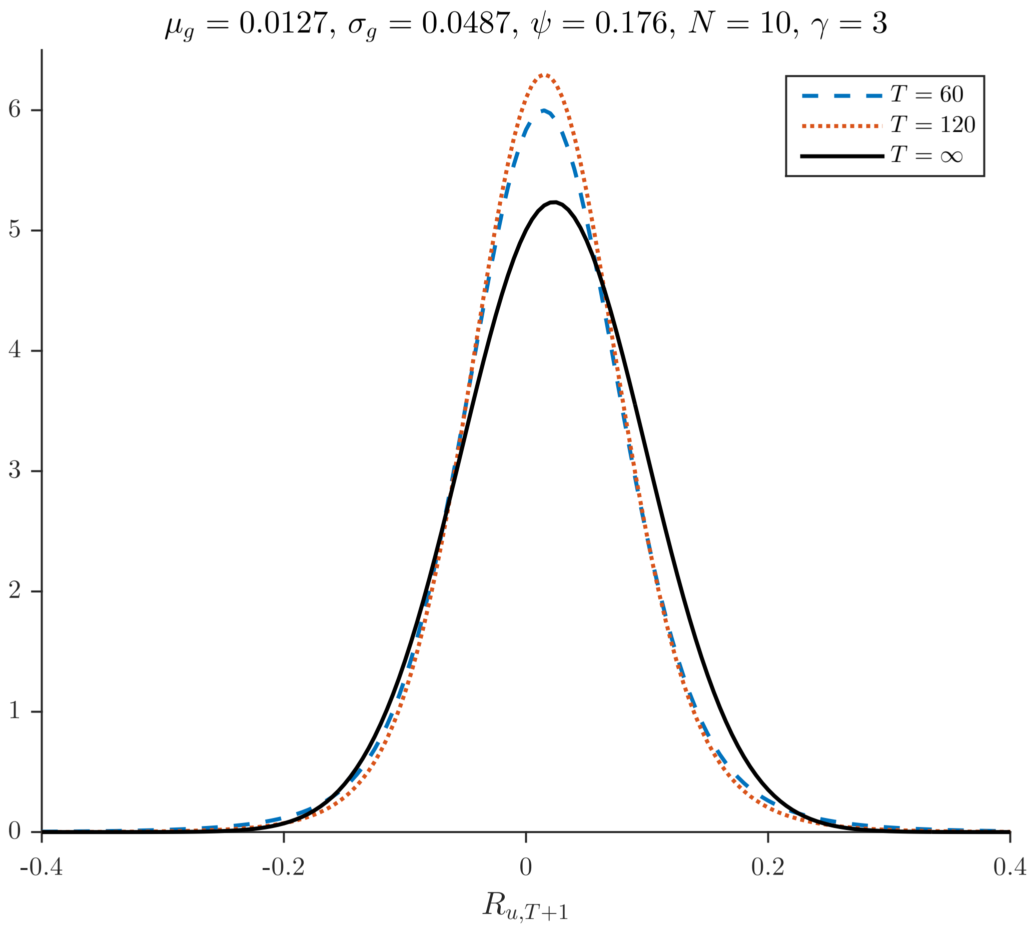

Figure 6
Distribution of out-of-sample portfolio return of the UL rule ($N = 10$)