

Exact Properties of Measures of Optimal Investment for Institutional Investors

J. Knight

**Department of Economics
University of Western Ontario
London, Ontario, Canada
and Visiting Professor
Birkbeck, University of London**

S. E. Satchell

**Faculty of Economics
University of Cambridge
and Visiting Professor
Birkbeck, University of London**

Abstract

We revisit the problem of calculating the exact distribution of optimal investments in a mean variance world under multivariate normality. The context we consider is where problems in optimisation are addressed through the use of Monte-Carlo simulation. Our findings give clear insight as to when Monte-Carlo simulation will, and will not work. Whilst a number of authors have considered aspects of this exact problem before, we extend the problem by considering the problem of an investor who wishes to maximise quadratic utility defined in terms of alpha and tracking errors. The results derived allow some exact and numerical analysis. Furthermore, they allow us to also derive results for the more traditional non-benchmarked portfolio problem.

JEL Classification: G11.

Keywords: alpha, tracking error, mean-variance, Monte-Carlo

The application of exact distribution theory to mean-variance (MV) analysis has been undertaken by a number of authors, see Jobson and Korkie (1989), Jobson (1991), Britten-Jones (1999), Stein (2002) and Hillier and Satchell (2003). The usual assumptions are that returns are iid multivariate normal and that there may or may not be a riskless asset. However, in all listed cases the analysis is in terms of absolute, i.e. unbenchmarked, portfolios. This is a limitation since most institutional risk analysis is based on MV analysis using returns relative to a benchmark.

The major motivation for this research has been to try and understand the magnitude of estimation error; that is, the extent to which the outcome of the portfolio decision is influenced by parameter uncertainty. Recently, however, Michaud (1998) has proposed a simulated optimisation procedure, the outcome of which can only be really understood by an analysis of the exact properties of optimal portfolios. Michaud's procedure purports to solve some of the problems of portfolio optimisation that arise from estimation error. Other authors have criticised Michaud's approach, see Scherer (2002) and Harvey et al. (2004). One purpose of this paper is to consider expected utility in terms of relative returns and compute the exact properties of the optimal alpha, tracking error, and Sharpe ratio; with a view to assessing Michaud's contribution and the extent to which the various criticisms can be deemed to be valid. Our results, whilst being highly simplified, since we do not impose the myriad of constraints that institutional portfolios typically obey, nevertheless exhibit certain key characteristics that shed light on investment issues. Furthermore, we are able to extend the problem to consider the same case with absolute, not relative, weights. This allows us to derive some new results for this problem. We present the mathematical framework in Section 2. In Section 3, we derive exact results for the relative problem, and consider the absolute problem. In Section 4 we consider some numerical calculations. Section 5 considers the case of restrictions applied to portfolios and also addresses the realistic case of inequality constraints. Our conclusions follow in Section 6.

Section 2

In this section, we discuss the role of portfolio simulation and some of the criticisms of portfolio optimisation. Portfolio optimisation has been criticised for being excessively sensitive to errors in the forecasts of expected returns. This leads to the optimiser choosing implausible portfolios and is a consequence of the difficulties in forecasting expected returns. Furthermore, these MV optimal portfolios lack the diversification deemed desirable by institutional investors, see Green and Hollifield (1992). A number of solutions to this

problem have emerged. In some contexts, Bayesian prices on the expected returns are used to control the sample variability of the means, see, for example, Satchell and Scowcroft (2003). Practitioners often employ large numbers of constraints on the portfolio weights to control the optimiser, we shall refer to this as the practitioner's solution. This solution has been given some support in the context of MV optimisation by Jagannathan and Ma (2002, 2003).

Michaud (1998) has advocated simulating the optimisation. The advantages of this is that we get some sense of the variability of the solution, however we need to understand what the averaging in the simulation will lead to.

To motivate our analysis we consider how Michaud (1998) carries out his resampling methodology. Quoting from Michaud (op. cit, pages 17, 19 and 37).

- “1. Monte Carlo simulate 18 years of monthly returns based on data in Tables 2.3 and 2.4...
2. Compute optimised input parameters from the simulated return data.
3. Compute efficient frontier portfolios...
4. Repeat steps 1-3 500 times...
- 5....Observe the variability in the efficient frontier estimation.”

The assumption behind the Monte Carlo simulation of returns can vary. It can be based on historical returns and involve resampling, or it may involve using means variance and covariance and simulating via multi-variate normality as Michaud details above, his Tables 2.3 and 2.4 contain first and second sample moments. The key feature of such an analysis is that the mean simulated efficient frontier will differ from the “population” efficient frontier based on the information in step 1 by the degree of finite sample bias. Whilst this should be small for $T = 216$ monthly observations, there are lots of portfolio calculations that will be based on much shorter time-periods due to the usual reasons; regime shifts, institutional change and time-varying parameters. Furthermore, we conjecture, and subsequently show, that it is not T that determines bias alone but T and N (the number of stocks) co-jointly. If N is large even for large T , then biases can be very large indeed.

It is worth noting that the emphasis of the above approach is in terms of the MV efficient frontier analysis rather than expected utility. But as we shall show next, maximising quadratic utility gives you a solution that is expressed solely in terms of efficient-set mathematics; the only additional information is the risk aversion coefficient (λ); as we change λ we move along the MV frontier in any case.

Jobson (1991), derives a number of key results in this area for the conventional minimum variance frontier, we shall refer to these results when appropriate. Stein, (2002) has also derived some of our results.

Consider the active weights ω , and the *known* benchmark weights \mathbf{b} , both $(N \times 1)$ vectors and both sum to one i.e. $\omega' \mathbf{i} = \mathbf{b}' \mathbf{i} = 1$. Let μ and Ω be the $(n \times 1)$ mean vector and covariance matrix of the N asset returns where the letter i denotes an $(N \times 1)$ vector of ones.

Our investor chooses to maximise U , where $U = \mu'(\omega - \mathbf{b}) - \frac{\lambda}{2}(\omega - \mathbf{b})' \Omega (\omega - \mathbf{b})$; note that there is also a constraint $(\omega - \mathbf{b})' \mathbf{i} = 0$. This is a classical MV problem equivalent, as we demonstrate, to computing the optimal frontier. It is straightforward to see that as λ ranges from 0 to ∞ we move down the frontier from the maximum expected return portfolio to the global minimum variance portfolio. This framework is widely used in finance, see Sharpe (1981), Grinold and Kahn (1995), Scherer (2002).

Our first-order condition is,

$$\frac{\partial U}{\partial(\omega - \mathbf{b})} = \mu - \lambda \Omega (\hat{\omega} - \mathbf{b}) + \hat{\theta} \mathbf{i} = 0 \quad \text{or} \quad \hat{\omega} = \mathbf{b} + \frac{1}{\lambda} \Omega^{-1} (\mu + \hat{\theta} \mathbf{i}).$$

Using $\mathbf{i}'(\omega - \mathbf{b}) = 0$, we see that $\frac{1}{\lambda}(\beta + \hat{\theta} \gamma) = 0$, where $\beta = \mathbf{i}' \Omega^{-1} \mu$, $\gamma = \mathbf{i}' \Omega^{-1} \mathbf{i}$, and we set $a = \mu' \Omega^{-1} \mu$. Thus $\hat{\omega} = \mathbf{b} + \frac{1}{\lambda} (\Omega^{-1} \mu - \frac{\beta}{\gamma} \Omega^{-1} \mathbf{i})$ and hence active returns α can be computed as

$$\begin{aligned} \alpha &= \mu'(\hat{\omega} - \mathbf{b}) = \frac{1}{\lambda} (a - \frac{\beta^2}{\gamma}) \\ &= \frac{1}{\lambda} (\frac{a\gamma - \beta^2}{\gamma}) \end{aligned} \tag{1}$$

Other terms of interest can be calculated. For example, we have

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{\lambda^2} (\Omega^{-1} \mu - \frac{\beta}{\gamma} \Omega^{-1} \mathbf{i})' \Omega (\Omega^{-1} \mu - \frac{\beta}{\gamma} \Omega^{-1} \mathbf{i}) \\
&= \frac{1}{\lambda^2} (a - \frac{\beta^2}{\gamma} + \frac{\beta^2}{\gamma}) \\
&= \frac{1}{\lambda^2} (a - \frac{\beta^2}{\gamma}),
\end{aligned} \tag{2}$$

we will focus on $\hat{\sigma}$, the tracking error or standard deviation of relative returns. Finally,

$$\begin{aligned}
E(U) &= \frac{1}{\lambda} (\frac{a\gamma - \beta^2}{\gamma}) - \frac{\lambda}{2} (\frac{1}{\lambda^2} (\frac{a\gamma - \beta^2}{\gamma})) \\
&= \frac{1}{2\lambda} (\frac{a\gamma - \beta^2}{\gamma}).
\end{aligned} \tag{3}$$

It is straightforward to compute the information ratio defined as $\frac{\alpha}{\hat{\sigma}}$. Notice that in this problem all terms depend essentially on a single term $(\frac{a\gamma - \beta^2}{\gamma})$ or functions of it.

Remark

A related formulation of the above problem is the following $\min \frac{1}{2} (\omega - \mathbf{b})' \Omega (\omega - \mathbf{b})$ subject to $(\omega - \mathbf{b})' \mathbf{i} = 0$ and $(\omega - \mathbf{b})' \mu = \pi$. Here the Lagrangean is given by

$$L = \frac{1}{2} (\omega - \mathbf{b})' \Omega (\omega - \mathbf{b}) - \theta_1 (\omega - \mathbf{b})' \mathbf{i} - \theta_2 ((\omega - \mathbf{b})' \mu - \pi)$$

resulting in the 1st order conditions

$$\frac{\partial L}{\partial \omega} = (\omega - \mathbf{b}) - \theta_1 \mathbf{i} - \theta_2 \mu = 0$$

$$\frac{\partial L}{\partial \theta_1} = (\omega - \mathbf{b})' \mathbf{i} = 0$$

$$\frac{\partial L}{\partial \theta_2} = (\omega - \mathbf{b})' \mu - \pi = 0$$

Solving we have $\omega - \mathbf{b} = \theta_1 \Omega^{-1} \mathbf{i} + \theta_2 \Omega^{-1} \mu$ with $\theta_1 = \frac{\beta \pi}{a\gamma - \beta^2}$ and $\theta_2 = \frac{\gamma \pi}{a\gamma - \beta^2}$.

Thus

$$\hat{\omega} = \mathbf{b} + \frac{\pi\gamma}{a\gamma - \beta^2} (\Omega^{-1}\mu - \frac{\beta}{\gamma}\Omega^{-1}\mathbf{i})$$

$$\hat{\omega} = \mathbf{b} + \pi\hat{\hat{\omega}}$$

and consequently,

$$\hat{\sigma}^2 = \pi^2 \hat{\hat{\omega}}' \Omega \hat{\hat{\omega}}$$

$$\hat{\sigma}^2 = \frac{\pi^2 \gamma}{a\gamma - \beta^2}$$

Comparing with (2) we see immediately that $\pi = \frac{1}{\lambda} (a - \beta^2/\gamma)$.

This second problem is simply the computation of the minimum variance frontier. It differs from the earlier version in that it explicitly specifies π , the expected rate of return, rather than λ , the risk aversion coefficient.

Section 3. Finite Sample Properties of Estimators of Alpha and Tracking Error.

Consider

$$\begin{aligned} \mathbf{Q} &= (\mu, \mathbf{i})' \Omega^{-1} (\mu, \mathbf{i}) \\ &= \begin{pmatrix} a & \beta \\ \beta & \gamma \end{pmatrix} \end{aligned} \tag{4}$$

It is well known that, under normality $\hat{\mu} \sim N(\mu, \frac{1}{T}\Omega)$, and $\hat{\mu}$ and S are independent, where S is the sample covariance matrix. Firstly, by Theorem 3.2.11 of Muirhead (1982), conditional on $\hat{\mu}$,

$\hat{\mathbf{Q}}^{-1} = ((\hat{\mu}, \mathbf{i})' \hat{\Omega}^{-1} (\hat{\mu}, \mathbf{i}))^{-1}$ has a central Wishart: $W_2(T - N + 1, \frac{1}{T}\bar{\mathbf{Q}}^{-1})$ where $\bar{\mathbf{Q}} = (\hat{\mu}, \mathbf{i})' \Omega^{-1} (\hat{\mu}, \mathbf{i})$. The statistic of interest is given by $\hat{h} = \frac{\hat{\gamma}}{\hat{a}\hat{\gamma} - \hat{\beta}^2}$ and is the first principal element of $\hat{\mathbf{Q}}^{-1}$. Formally, we have $\hat{h} = (1, 0)\hat{\mathbf{Q}}^{-1}(1, 0)'$ and again from Muirhead

(1982) Theorem 3.2.5 we have $\hat{\mathbf{h}} | \hat{\boldsymbol{\mu}} \sim W_1(T - N + 1, \frac{1}{T}(1,0)\bar{\mathbf{Q}}^{-1}(1,0)')$ and thus, letting $\boldsymbol{\varphi} = \frac{1}{T}(1,0)\bar{\mathbf{Q}}^{-1}(1,0)'$, we have

$$\frac{\hat{\mathbf{h}}}{\boldsymbol{\varphi}} | \boldsymbol{\varphi} \sim \chi^2_{(\nu)}, \text{ where } \nu = T - N + 1$$

and consequently, this result holds unconditionally.

Next we examine $\boldsymbol{\varphi}$, noting immediately that $T\boldsymbol{\varphi}$ is the first principal element in $\bar{\mathbf{Q}}^{-1}$. That is

$$\boldsymbol{\varphi} = \frac{1}{T} \frac{\mathbf{i}' \boldsymbol{\Omega}^{-1} \mathbf{i}}{(\hat{\boldsymbol{\mu}}' \boldsymbol{\Omega}^{-1} \hat{\boldsymbol{\mu}})(\mathbf{i}' \boldsymbol{\Omega}^{-1} \mathbf{i}) - (\hat{\boldsymbol{\mu}}' \boldsymbol{\Omega}^{-1} \mathbf{i})(\mathbf{i}' \boldsymbol{\Omega}^{-1} \hat{\boldsymbol{\mu}})} \quad (5)$$

Now $\hat{\boldsymbol{\mu}} \sim N(\boldsymbol{\mu}, \frac{1}{T}\boldsymbol{\Omega})$ and thus letting $\boldsymbol{\omega} = \sqrt{T}\boldsymbol{\Omega}^{-1/2}\hat{\boldsymbol{\mu}}$ we have $\boldsymbol{\omega} \sim N(\sqrt{T}\boldsymbol{\Omega}^{-1/2}\boldsymbol{\mu}, \mathbf{I}_N)$.

Further, letting $\mathbf{c} = \boldsymbol{\Omega}^{-1/2}\mathbf{i}$ we have that

$$\begin{aligned} \boldsymbol{\varphi} &= \frac{1}{T} \left(\frac{\mathbf{T}\mathbf{c}'\mathbf{c}}{\boldsymbol{\omega}'\boldsymbol{\omega}\mathbf{c}'\mathbf{c} - \boldsymbol{\omega}'\mathbf{c}\mathbf{c}'\boldsymbol{\omega}} \right) \\ &= \frac{1}{\boldsymbol{\omega}'(\mathbf{I} - \mathbf{c}(\mathbf{c}'\mathbf{c})^{-1}\mathbf{c}')\boldsymbol{\omega}} = \frac{1}{\boldsymbol{\omega}'\bar{\mathbf{P}}_c\boldsymbol{\omega}} \end{aligned} \quad (6)$$

and it follows immediately that $\boldsymbol{\omega}'\bar{\mathbf{P}}_c\boldsymbol{\omega} \sim \chi^2_{(N-1, \lambda^*)}$ where $\lambda^* = \mathbf{T}\boldsymbol{\mu}'\boldsymbol{\Omega}^{-1/2}\bar{\mathbf{P}}_c\boldsymbol{\Omega}^{-1/2}\boldsymbol{\mu} = \mathbf{T}/\mathbf{h}$

with $\mathbf{h} = \left(\frac{\gamma}{\mathbf{a}\gamma - \beta^2} \right)$. Therefore $\boldsymbol{\varphi}^{-1} \sim \chi^2_{(N-1, \mathbf{T}/\mathbf{h})}$ and thus the distribution of $\hat{\mathbf{h}}$ will be given

by the following ratio:

$$\hat{\mathbf{h}} \sim \frac{\chi^2_{(\nu)}}{\chi^2_{(N-1, \mathbf{T}/\mathbf{h})}} \quad (7)$$

where the two χ^2 variables in (7) are independent. This result also appears in Stein (2002).

Thus, pdf($\hat{\mathbf{h}}$) can be easily found using results related to non-central F distributions. In this regard we have from Johnson and Kotz (1972, p. 191) that the pdf($1/\mathbf{h} = \mathbf{g}$) is, noting $B(\cdot)$ and ${}_1F_1(\cdot)$ to be Beta and confluent hypergeometric functions respectively,

$$\text{pdf}(\mathbf{g}) = \frac{e^{-\mathbf{T}/2\mathbf{h}} \mathbf{g}^{\frac{N-1}{2}-1}}{B\left(\frac{N-1}{2}, \frac{\nu}{2}\right)(1+\mathbf{g})^{\frac{N-1+\nu}{2}}} {}_1F_1\left(\frac{N-1+\nu}{2}, \frac{N-1}{2}, \frac{\mathbf{T}}{2\mathbf{h}} \frac{\mathbf{g}}{(1+\mathbf{g})}\right)$$

Using this result and simple transformations one can readily derive the pdf density functions for the quantities of interest, viz, $\hat{\alpha} = 1/\lambda\hat{h}$. Tracking Error = $\overline{\text{TE}} = 1/\lambda\sqrt{\hat{h}}$ and the Information Ratio = $\overline{\text{IR}} = 1/\sqrt{\hat{h}}$. Thus we have

$$\text{pdf}(\hat{\alpha} = \omega) = \frac{\lambda e^{-T/2h} (\lambda\omega)^{\frac{N-1}{2}-1}}{\text{B}(\frac{N-1}{2}, \frac{\nu}{2})(1+\lambda\omega)^{\frac{N-1+\nu}{2}}} {}_1F_1\left(\frac{N-1+\nu}{2}, \frac{N-1}{2}, \frac{T}{2h} \frac{\lambda\omega}{(1+\lambda\omega)}\right), \quad \omega > 0$$

$$\text{pdf}(\overline{\text{IR}} = x) = \frac{2xe^{-T/2h} (x^2)^{\frac{N-1}{2}-1}}{\text{B}(\frac{N-1}{2}, \frac{\nu}{2})(1+x^2)^{\frac{N-1+\nu}{2}}} {}_1F_1\left(\frac{N-1+\nu}{2}, \frac{N-1}{2}, \frac{T}{2h} \left(\frac{x^2}{(1+x^2)}\right)\right), \quad x > 0$$

$$\text{pdf}(\overline{\text{TE}} = y) = \frac{2\lambda^2 ye^{-T/2h} (\lambda^2 y^2)^{\frac{N-1}{2}-1}}{\text{B}(\frac{N-1}{2}, \frac{\nu}{2})(1+\lambda^2 y^2)^{\frac{N-1+\nu}{2}}} {}_1F_1\left(\frac{N-1+\nu}{2}, \frac{N-1}{2}, \frac{T}{2h} \left(\frac{\lambda^2 y^2}{(1+\lambda^2 y^2)}\right)\right), \quad y > 0$$

From these pdfs or via that of \hat{h} or g , we can easily find moments. That is, since

$$\hat{h}^{-1} = g = \chi_{(N-1, T/h)}^2 / \chi_{(\nu)}^2$$

$$\text{E}(g^k) = \text{E}[(\chi_{(N-1, T/h)}^2)^k] \text{E}[(\chi_{(\nu)}^2)^{-k}]$$

and since

$$\text{E}[(\chi_{(N-1, T/h)}^2)^k] = \frac{e^{-T/2h} 2^k}{\Gamma(\frac{N-1}{2})} \Gamma\left(\frac{N-1}{2} + k\right) {}_1F_1\left(\frac{N-1}{2} + k, \frac{N-1}{2}; T/2h\right)$$

and

$$\text{E}[(\chi_{(\nu)}^2)^{-k}] = \frac{\Gamma(\frac{\nu}{2} - k)}{2^k \Gamma(\frac{\nu}{2})}, \quad \nu > 2k$$

$$\text{E}[g^k] = \frac{e^{-T/2h} \Gamma(\frac{N-1}{2} + k) \Gamma(\frac{\nu}{2} - k)}{\Gamma(\frac{N-1}{2}) \Gamma(\frac{\nu}{2})} {}_1F_1\left(\frac{N-1}{2} + k, \frac{N-1}{2}; T/2h\right)$$

Therefore:

$$\begin{aligned} E[(\hat{\alpha})^k] &= \lambda^{-k} E(g^k) \\ &= \frac{e^{-T/2h} \Gamma\left(\frac{N-1}{2} + k\right) \Gamma\left(\frac{\nu}{2} - k\right) 2^k}{\lambda^k \Gamma\left(\frac{N-1}{2}\right) \Gamma\left(\frac{\nu}{2}\right)} {}_1F_1\left(\frac{N-1}{2} + k, \frac{N-1}{2}; T/2h\right) \end{aligned} \quad (8)$$

$$\begin{aligned} E((\overline{TE})^k) &= \lambda^{-k} E(g^{k/2}) \\ &= \frac{e^{-T/2h} \Gamma\left(\frac{N-1}{2} + \frac{k}{2}\right) \Gamma\left(\frac{\nu}{2} - \frac{k}{2}\right) 2^k}{\lambda^k \Gamma\left(\frac{N-1}{2}\right) \Gamma\left(\frac{\nu}{2}\right)} {}_1F_1\left(\frac{N-1}{2} + \frac{k}{2}, \frac{N-1}{2}; T/2h\right) \end{aligned} \quad (9)$$

$$E((\overline{IR})^k) = E(g^{k/2}) \quad (10)$$

In particular, if we consider the means of the three quantities we have:

$$\begin{aligned} E(\hat{\alpha}) &= \frac{\Gamma\left(\frac{N-1}{2} + 1\right) \Gamma\left(\frac{\nu-1}{2} - 1\right)}{\lambda \Gamma\left(\frac{N-1}{2}\right) \Gamma\left(\frac{\nu}{2}\right)} e^{-T/2h} {}_1F_1\left(\frac{N-1}{2} + 1, \frac{N-1}{2}; T/2h\right) \\ &= \frac{(N-1)}{\lambda(\nu-2)} {}_1F_1\left(-1, \frac{N-1}{2}; -T/2h\right) \\ &= \frac{N-1}{\lambda(\nu-2)} \left[1 + \frac{T}{h(N-1)} \right] \\ &= \frac{N-1}{\lambda(\nu-2)} + \frac{T}{\lambda h(\nu-2)}; \quad \nu = T - N + 1. \end{aligned}$$

Since the true $\alpha = 1/2h$ we can readily develop an unbiased estimator of α via a simple transformation:

$$E\left[\frac{(\nu-2)}{T} \hat{\alpha} - \frac{N-1}{T\lambda}\right] = \alpha$$

For other quantities we have :

$$E(\overline{TE}) = \frac{\Gamma\left(\frac{N-1}{2} + \frac{1}{2}\right) \Gamma\left(\frac{\nu}{2} - \frac{1}{2}\right)}{\lambda \Gamma\left(\frac{N-1}{2}\right) \Gamma\left(\frac{\nu}{2}\right)} {}_1F_1\left(-\frac{1}{2}, \frac{N-1}{2}; -T/2h\right)$$

and

$$E(\overline{\mathbf{IR}}) = \lambda E(\overline{\mathbf{TE}}).$$

Also, note that since $E(\hat{\sigma}^2) = E(\overline{\mathbf{TE}}^2) = \frac{1}{\lambda} E(\hat{\alpha})$ and an unbiased estimator of σ^2 is easily derived to be

$$\hat{\sigma}^2 = \frac{1}{\lambda} \left[\frac{\nu - 2}{T} \hat{\alpha} - \frac{(N-1)}{T\lambda} \right].$$

While little progress can be made with exact expressions for the expectation of $\overline{\mathbf{TE}}$ and $\overline{\mathbf{IR}}$, we can get more insight by considering approximations. Jobson (1991) gives similar results in that he derives the means and variances of \mathbf{a} , β and γ and h^{-1} and determines their marginal distributions. Stein (2002) also derives some formulae similar to ours.

We now examine a situation in which both N , the number of stocks or assets and T , the sample size, increase in such a way that the ratio $\frac{N-1}{T}$ remains constant. We note \mathbf{a} , β and γ and hence h also depend upon N and thus asymptotics here require that the terms limit to a constant, or at least we need to make assumptions about $\frac{1}{h}$ as a function of N . For the moment we shall not consider this possible influence. Thus we now let $\frac{N-1}{T} = n$ so that $N-1 = T.n$. By letting $T \rightarrow \infty$ we can readily see the effect on the moments of large N and T . For $\hat{\alpha}$ we have from our exact result

$$E(\hat{\alpha}) = \frac{Tn}{\lambda(T(1-n) - 2)} \left[1 + \frac{1}{hn} \right]$$

and therefore as $T \rightarrow \infty$ we find

$$E(\hat{\alpha}) \rightarrow \frac{1}{\lambda} \left[\frac{n}{1-n} + \frac{1}{h(1-n)} \right].$$

The corresponding results for $\overline{\mathbf{IR}}$ and $\overline{\mathbf{TE}}$ are given by

$$E(\overline{IR}) \rightarrow \left(\frac{n}{1-n} + \frac{1}{h(1-n)} \right)^{\frac{1}{2}}$$

and

$$E(\overline{TE}) \rightarrow \frac{1}{\lambda} \left(\frac{n}{1-n} + \frac{1}{h(1-n)} \right)^{\frac{1}{2}}.$$

We now examine portfolio optimization without a benchmark. Here we maximize $\omega' \mu - \frac{\lambda}{2} \omega' \Omega \omega$ subject to $\omega' \mathbf{i} = 1$. The associated Lagrangian is given by:

$$W = \omega' \mu - \frac{\lambda}{2} \omega' \Omega \omega - \theta (\omega' \mathbf{i} - 1) \quad (11)$$

with

$$\frac{\partial W}{\partial \omega} = \mu - \lambda \omega - \theta \mathbf{i} = 0$$

implying

$$\omega = \frac{1}{\lambda} (\Omega^{-1} \mu - \theta \Omega^{-1} \mathbf{i})$$

since $\mathbf{i}' \omega = 1$ we have immediately that

$$\theta = (\mathbf{i}' \Omega^{-1} \mu - \lambda) / \mathbf{i}' \Omega^{-1} \mathbf{i}$$

i.e. $\theta = (\beta - \lambda) / \gamma$.

Consequently,

$$\hat{\omega} = \frac{1}{\lambda} (\hat{\Omega}^{-1} \hat{\mu} - \left(\frac{\hat{\beta} - \lambda}{\hat{\gamma}} \right) \hat{\Omega}^{-1} \mathbf{i})$$

and thus

$$\tilde{\alpha} = \hat{\mu}' \hat{\omega} = \frac{1}{\lambda} \hat{\mu}' \hat{\Omega}^{-1} \hat{\mu} - \left(\frac{\hat{\beta} - \lambda}{\hat{\gamma}} \right) \hat{\mu}' \hat{\Omega}^{-1} \mathbf{i}$$

$$\text{i.e. } \tilde{\alpha} = \frac{1}{\lambda \hat{h}} + \frac{\hat{\beta}}{\hat{\gamma}} = \tilde{\alpha} + \frac{\hat{\beta}}{\hat{\gamma}}.$$

Also,

$$\tilde{\sigma}^2 = \hat{\omega}' \hat{\Omega} \hat{\omega}$$

$$\tilde{\sigma}^2 = \frac{1}{\lambda^2 \hat{h}} + \frac{1}{\hat{\gamma}}.$$

Thus we notice immediately that the active return $\tilde{\alpha}$ and the $\overline{\text{TE}}^2$ are given by our earlier results plus an additional term. Under the normality assumption we again examine some of the statistical properties of these new estimations. We present the results below, the proofs are straightforward extensions of our earlier results.

$$\begin{aligned} E(\tilde{\alpha}) &= E(\hat{\alpha}) + E(\hat{\beta} / \hat{\gamma}) \\ &= E(\hat{\alpha}) + (\beta / \gamma). \end{aligned}$$

$$\begin{aligned} \text{Var}(\tilde{\alpha}) &= \text{Var}(\hat{\alpha}) + \text{Var}(\hat{\beta} / \hat{\gamma}), \text{ since } \hat{\alpha} \text{ and } \hat{\beta} / \hat{\gamma} \text{ are independent} \\ &= \text{var}(\hat{\alpha}) + \frac{1}{T\gamma} (E(\hat{h}^{-1}) + 1) \end{aligned}$$

$$E(\tilde{\sigma}^2) = E(\hat{\sigma}^2) + \frac{T - N + 1}{T\gamma}.$$

Section 4

We now illustrate the accuracy of these approximations using two contrasting numerical examples. In both cases we have $\lambda = 12.5$ and $h=4$, giving true values of $\alpha = 0.02$, $\text{TE} = 0.04$ and $\text{IR} = 0.5$. These values correspond to the sorts of numbers found in institutional investment for active managers measured on an annualised basis. We now consider two cases.

- i) $T = 180, N = 4$ so that $n = \frac{1}{60} = 0.01667$
- ii) $T = 180, N = 80$, so that $n = \frac{79}{180} = 0.43889$.

The results are given in the following table.

	α	TE	IR
The true values	0.02	0.04	0.05
Case I (T = 180, N = 4)			
Exact Expected Values	0.021726	0.041410	0.517621
Approx	0.021695	0.041660	0.520756
Case II (T = 180, N = 80)			
Exact Expected Values	0.100202	0.089062	1.113272
Approx	0.098218	0.088642	1.108027

What the above results illustrate is the fact that while the estimators $\hat{\alpha}$, $\overline{\text{TE}}$, $\overline{\text{IR}}$ are always biased, the bias is very small when n is small. However, for large n the bias is extremely large being more than four times the true value for α and greater than twice the true value for TE and IR. We also notice that in both cases the approximation is quite accurate.

Keeping our numerical results consistent with those in Scherer (2002, p.165) we will consider two cases

$$\text{i) } \mathbf{Q} = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} = \begin{pmatrix} 0.125 & 6.5 \\ 6.5 & 675.00 \end{pmatrix}$$

$$\text{giving } h = \frac{\gamma}{\alpha\gamma - \beta^2} \approx 16$$

and

$$\text{ii) } \mathbf{Q} = \begin{pmatrix} 0.3 & 6.5 \\ 6.5 & 800 \end{pmatrix} \text{ with } h = 4.$$

In each case, by choosing different values for λ (risk aversion parameter) we can generate a wide set of values for both active returns $\alpha = \mu' \omega = \frac{1}{\lambda h}$ and tracking error $\text{TE} = \frac{1}{\lambda \sqrt{h}}$. The following table highlights this relationship.

Some authors such as Grinold and Kahn (1999) express the units associated with the active return, α and the tracking error, TE, in terms of percent. Others, e.g. Scherer (2002), use the decimal equivalent. However, shifting the units from decimal to percent will alter λ , the risk aversion parameter, by a factor of 100. That is, the λ associated with percent units will be 100th of the value of λ associated with decimal units. Thus, the following constellations of parameter values listed in the two panels below are consistent:

h = 16				h = 4			
λ	α	TE	IR	λ	α	TE	IR
2	0.03125	0.125	0.25	12.5	0.02	0.04	0.5
0.02	3.125	12.5	0.25	0.125	2	4	0.5

In what follows we choose the decimal representation.

λ	h = 16			h = 4		
	α	TE	IR	α	TE	IR
2	0.03125	0.125	0.25	0.125	0.25	0.5
4	0.0156	0.0625	0.25	0.0625	0.125	0.5
6	0.0104	0.04167	0.25	0.04167	0.0833	0.5
8	0.0078	0.03125	0.25	0.03125	0.0625	0.5
12.5				0.02	0.04	0.5

We now examine the tracking error optimization and the performance, i.e. relative bias. of the standard estimators for different portfolio sizes, $N = 4$ and $N = 80$ with $T = 180$ in both cases.

θ :	h = 16, $\lambda = 2$			h = 4, $\lambda = 12.5$		
	$\alpha = 0.03125$	TE = 0.125	IR = 0.25	$\alpha = 0.02$	TE = 0.04	IR = 0.5
N=4 $E(\hat{\theta})$	0.0407	0.1378	0.2757	0.0217	0.0414	0.5176
% Rel. Bias	30.24	10.24	10.28	8.50	3.50	3.52
N=80 $E(\hat{\theta})$	0.4558	0.4747	0.9494	0.1002	0.0891	1.1133
% Rel. Bias	1358.56	279.76	279.76	401.00	122.75	122.66

Again, we see evidence of large relative biases in the large N case, pointing to quite staggering inaccuracy.

Section 5. General Linear Restrictions

The results of the previous sections can be readily extended to incorporate general linear restrictions on the relative weights. Here we briefly outline the results; the full derivation is available from the authors upon request. We now consider the maximization of utility subject to a set of K restrictions: $R(\omega - b) = 0$, where R is a $K \times N$ matrix. The Lagrangian and the associated first-order conditions for the relative case are as follows:

$$L = \mu'(\omega - \mathbf{b}) - \frac{\lambda}{2}(\omega - \mathbf{b})'\Omega(\omega - \mathbf{b}) + \theta'\mathbf{R}(\omega - \mathbf{b})$$

$$\frac{\partial L}{\partial \omega} = \mu - \lambda\Omega(\omega - \mathbf{b}) + \mathbf{R}'\theta = 0$$

$$\frac{\partial L}{\partial \theta} = \mathbf{R}(\omega - \mathbf{b}) = 0$$

Solving we find

$$\omega = \mathbf{b} + \frac{1}{\lambda}\Omega^{-1}(\mu - \mathbf{R}'(\mathbf{R}\Omega^{-1}\mathbf{R}')\mathbf{R}\Omega^{-1}\mu)$$

resulting in

$$\begin{aligned} \alpha &= \mu'(\omega - \mathbf{b}) = \frac{1}{\lambda} \left[\mu'\Omega^{-1}\mu - \mu'\Omega^{-1}\mathbf{R}'(\mathbf{R}\Omega^{-1}\mathbf{R}')^{-1}\mathbf{R}\Omega^{-1}\mu \right] \\ &= \frac{1}{\lambda} (1,0)\hat{\mathbf{Q}}_K^{-1}(1,0)' \end{aligned}$$

and

$$\begin{aligned} \sigma^2 &= (\omega - \mathbf{b})'\Omega(\omega - \mathbf{b}) \\ &= \frac{1}{\lambda^2} (1,0)\hat{\mathbf{Q}}_K^{-1}(1,0)' \end{aligned}$$

where $\hat{\mathbf{Q}}_K = (\hat{\mu}, \mathbf{R}')'\hat{\Omega}^{-1}(\hat{\mu}, \mathbf{R}')$

with

$$\hat{\mathbf{Q}}_K^{-1} \sim W_{K+1}(T - N + K, \frac{1}{T}\hat{\mathbf{Q}}_K^{-1}).$$

Following earlier results we now define

$$\hat{\mathbf{h}}_K = (1,0)\hat{\mathbf{Q}}_K^{-1}(1,0)'$$

and we have immediately, corresponding to (7):

$$\hat{\mathbf{h}}_K \sim \frac{\chi^2_{(T-N+K)}}{\chi^2_{(N-K, \frac{1}{\lambda^2})}}$$

where $\mathbf{h}_K = (1,0)'\mathbf{Q}_K^{-1}(1,0)$ with $\mathbf{Q}_K = (\mu, \mathbf{R}')'\Omega^{-1}(\mu, \mathbf{R}')$.

Thus, by a simple substitution into our earlier results we can readily specify the exact distribution and moments of $\hat{\alpha}$, $\overline{\mathbf{IR}}$ and $\overline{\mathbf{TE}}$. That is, we merely replace $N - 1$ by $N - K$ and \mathbf{h}

by h_K . As intuition suggests, increasing the number of restrictions is exactly the same as reducing the number of assets. However, the non-centrality parameter h_K will change as the constraints change.

This is clear from the following. If we let

$$\mathbf{R}_{K \times N} = \begin{pmatrix} \mathbf{r}'_1 \\ \mathbf{r}'_2 \\ \vdots \\ \mathbf{r}'_K \end{pmatrix} \text{ where } \mathbf{r}'_i \text{ is a } 1 \times N \text{ vector.}$$

Then

$$\begin{aligned} \mathbf{Q}_K &= \begin{pmatrix} \boldsymbol{\mu}' \boldsymbol{\Omega}^{-1} \boldsymbol{\mu} & \boldsymbol{\mu}' \boldsymbol{\Omega}^{-1} \mathbf{R}' \\ \mathbf{R} \boldsymbol{\Omega}^{-1} \boldsymbol{\mu} & \mathbf{R} \boldsymbol{\Omega}^{-1} \mathbf{R}' \\ 1 & \mathbf{K} \end{pmatrix} \begin{matrix} 1 \\ \mathbf{K} \end{matrix} \\ &= \begin{pmatrix} \mathbf{a} \vdots \beta_1 & \beta_1 \cdots \beta_K \\ \boldsymbol{\beta} \vdots \gamma_1 & \gamma_{12} \cdots \gamma_{1K} \\ \vdots & \vdots \\ \beta_K \vdots \gamma_K & \cdots \cdots \gamma_{KK} \end{pmatrix} = \begin{pmatrix} \mathbf{a} & \boldsymbol{\beta}' \\ \boldsymbol{\beta} & \boldsymbol{\Gamma} \end{pmatrix} \end{aligned}$$

and

h_K will be the (1,1) element of the inverse of the $(K + 1)^{\text{th}}$ principal minor of \mathbf{Q}_K i.e.

$$h_K = (\mathbf{a} - \boldsymbol{\beta}' \boldsymbol{\Gamma}^{-1} \boldsymbol{\beta})^{-1}.$$

That is for $K = 1$ $h_K = h_1 = (\mathbf{a} - \beta_1^2 / \gamma_{11})^{-1}$ and when

$$K = 2, h_K = h_2 = \left[\mathbf{a} - (\beta_1 \beta_2) \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{12} & \gamma_{22} \end{bmatrix}^{-1} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \right]^{-1} \text{ etc.}$$

When the restrictions are orthogonal in the sense that $\boldsymbol{\Gamma} = \mathbf{R} \boldsymbol{\Omega}^{-1} \mathbf{R}'$ is a diagonal matrix, $\boldsymbol{\Gamma}^{-1}$ will have a simple representation along with h_K . In this case

$$h_K = (\mathbf{a} - \sum_{i=1}^K \beta_i^2 / \gamma_{ii})^{-1}$$

which illustrates, quite clearly that as K increases h_K also increases. To see the effect on the estimators consider the bias in $\hat{\alpha}$. From our results in Section 3 we have

$$E(\hat{\alpha}) - \alpha = \frac{N - K}{\lambda(T - N + K)} + \frac{1}{\lambda h_K} \left(\frac{T}{(T - N + K)} - 1 \right)$$

and thus as K increases the bias will tend to be zero. In the more general case the same argument applies as long as h_K is bounded from below.

In the case of inequality constraints, the problem is more complex. This problem has been discussed in Jagannathan and Ma (2002), although they consider upper and lower constraints on the portfolio proportions only (see equations 1-4, pg. 6, 2002).

To convert a realistic optimization into an exact problem, we consider the Kuhn-Tucker conditions, appropriate to quadratic utility.

Our problem now becomes $\max L = \mu' \omega - \frac{\lambda}{2} \omega' \Omega \omega$ as before, but now, we consider K constraints of the form $A\omega \leq b$ and also no short-sales $\omega \geq 0$.

Our Kuhn-Tucker conditions are now

$$A\omega + v = b$$

$$\lambda \Omega \omega + A'u - y = \mu$$

and $\omega \geq 0$, $u \geq 0$, $y \geq 0$, and $v \geq 0$ plus the complementary constraint $\omega'y + u'v = 0$.

Because of the concavity of the objective function and linearity of the constraints, Kuhn-Tucker conditions apply and ω will be optimal iff we can find u , y and v such that all four vectors together satisfy the above constraints. If we wished to capture explicitly the fact that some of the inequality constraints are actually equality constraints, then further refinements are necessary.

To simplify this problem but to consider the impact of constraints, we shall consider our calculations when $\omega_1 \geq 0$ but otherwise the problem is as in (11).

Suppose that we constrained ω_1 to $\tilde{\omega}_1$ such that $\tilde{\omega}_1 \geq 0$. If the original $\hat{\omega}_1 \geq 0$, $\tilde{\omega}_1 = \hat{\omega}_1$, if $\hat{\omega}_1 < 0$, $\tilde{\omega}_1 = 0$. The distribution of $\tilde{\pi}$ and $\tilde{\sigma}^2$ will be as before if $\hat{\omega}_1 \geq 0$, but with N reduced by one and all parameters approximately adjusted. As we increase the number of constraints to K , say, such that $\omega \geq 0$ we get 2^K regions corresponding to all cases where constraints bite or not. In each of these regions the distribution may differ.

Sharpe(1970), Best(2000), Best and Grauer(1991), and no doubt many others, mention that a description of the constrained frontier consists essentially of solving for the corner portfolios, see e.g. Sharpe(1970, pg 66); these being the set of efficient portfolios where the set of active constraints change. Ordering these portfolios by expected return, and considering any two adjacent portfolios, two fund separation will apply to all the funds in between, that is, they can be treated as linear combinations of the two adjacent corner portfolios. Unfortunately, in the context of our problem, the stochastic nature of the means and covariances implies that the corner portfolios become stochastic, and this gives rise to different numbers of constraints holding and consequent mixtures of distributions for alpha and tracking error.

Section 6. Conclusion

Our paper has collected together and extended a number of results on exact properties of portfolio measures, some of which already exist in the literature. We extend these results to include relative and absolute return utility and relative and absolute mean-variance frontiers.

We compute biases for the optimal portfolios, alpha, volatility and the information ratio. We detect significant biases for the case when the number of assets increases with the sample size, a case of great practical relevance. We further show that when the problem is constrained these biases are reduced. This sheds some light on the practitioner approach to Mean-Variance optimization of imposing large numbers of constraints. Not only does this control the optimization; but, if the constraints are valid, it reduces the bias as well. Finally, simulating optimization can be seen as a satisfactory procedure if N is small relative to T , or if N is large and K is large relative to T , or if the average optimal portfolio or its moments are bias-corrected. We have not investigated the impact of N , K and T on the width of the simulated confidence intervals for the key parameters nor have we considered how we might extend our analysis to the Kuhn-Tucker problem discussed in section 5; these remain topics for future research.

Bibliography

Best, M.J. (2000) An Algorithm for the Solution of the Parametric Quadratic Programming Problem, in *Applied Mathematics and Parallel Computing – Festschrift for Klaus Ritter*, H. Fischer, B. Riedmuller and S. Schaffler (editors), Heidelberg: Physica-Verlag

Best, M.J. and R.B. Grauer(1991) On the Sensitivity of Mean-Variance-Efficient Portfolios: Some Analytical and Computational Results, *The Review of Financial Studies*,4,315-342.

Britten-Jones, M. (1999) The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights, *Journal of Finance* 54, 655-671.

Green, R.C. and B. Hollifield (1992). When will Mean-Variance Efficient Portfolios be Well Diversified? *Journal of Finance*, 47, 1785-1809.

Grinold, R.C. and R.N. Kahn (1999) *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Selecting Superior Returns and Controlling Risk*, McGraw Hill Inc.

Harvey, C., J.C. Leichty, M.W. Leichty, P. Muller, (2004) *Portfolio Selection with Higher Moments*

Hillier, G. and S. Satchell (2003) Some Exact Results for Efficient Portfolios with Given Returns, *New Advances in Portfolio Construction and Implementation*, Butterworth-Heinemann, London.

Jagannathan, R. and T. Ma (2002). Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraint Helps. NBER Working Paper 8922.

Jagannathan, R. and T. Ma (2003). Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraint Helps. *Journal of Finance*, 58, 1651-1683.

Jobson, J.D. (1991) Confidence Region for the Mean-Variance Efficient Set: An Alternative Approach to Estimation Risk, *Review of Quantitative Finance and Accounting*, 1, 235-237.

Jobson, J.D. and R. Korkie (1989). A performance interpretation of multivariate tests of asset set intersection, spanning, and mean-variance efficiency. *Journal of Financial and Quantitative Analysis* 24, 185-204.

Johnson, N.L. and S.C. Kotz (1972) *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley and Sons. Inc.

Michaud, R. (1998) *Efficient Asst Management* (Harvard Business School Press).

Muirhead, R.J. (1982) *Aspects of Multivariate Statistical Theory*, John Wiley and Sons Inc.

Satchell, S.E. and A. Scowcroft (2003). *New Advances in Portfolio Construction and Implementation*, Butterworth and Heinemann, London.

Scherer, B. (2002) *Portfolio Construction and Risk Budgeting*, Risk Books.

Sharpe, W.F. (1970) *Portfolio Theory and Capital Markets*, McGraw-Hill Book Company.

Sharpe, W.F. (1981) Decentralised Investment Management, *Journal of Finance*, 36, 217-234.

Stein, J. (2002) *Portfolio Theory and Market Fluctuations*, MSc Dissertation in Mathematical Finance, Oxford University.