Notes on the Sharpe ratio

Steven E. Pav *

September 21, 2013

Abstract

Herein is an incomplete collection of facts about the Sharpe ratio, and the Sharpe ratio of the Markowitz portfolio. Connections between the Sharpe ratio and the t-test, and between the Markowitz portfolio and the Hotelling T^2 statistic are explored. Many classical results for testing means can be easily translated into tests on assets and portfolios.

Contents

1 The Sharpe ratio **2**
- 1.1 Distribution of the Sharpe ratio 2
- 1.2 Tests involving the Sharpe ratio 3
- 1.3 Moments of the Sharpe Ratio 3
- 1.4 Asymptotics and Confidence Intervals 4

2 Sharpe ratio and portfolio optimization **5**
- 2.1 Asymptotics and Confidence Intervals 6
- 2.2 Inference on SNR 7
- 2.3 Sharpe ratio and simple constrained portfolio optimization 8
- 2.4 Spanning and hedging 8
- 2.5 Optimal Sharpe ratio under positivity constraint 10

3 Miscellanea **11**
- 3.1 Which Returns? 11
- 3.2 Sharpe is nearly leverage invariant 11
- 3.3 The 'haircut' 12

A Asymptotic Efficiency of Sharpe Ratio **18**

B Some Moments **19**

C Square Root F **20**

D Untangling Giri **20**

* shabbychef@gmail.com

1 The Sharpe ratio

In 1966 William Sharpe suggested that the performance of mutual funds be analyzed by the ratio of returns to standard deviation. [22] His eponymous ratio[1], $\hat{\zeta}$, is defined as

$$\hat{\zeta} = \frac{\hat{\mu}}{\hat{\sigma}},$$

where $\hat{\mu}$ is the historical, or sample, mean return of the mutual fund, and $\hat{\sigma}$ is the sample standard deviation. Sharpe admits that one would ideally use *predictions* of return and volatility, but that "the predictions cannot be obtained in any satisfactory manner ... Instead, ex post values must be used." [22]

A most remarkable fact about the Sharpe ratio, of which most practicioners seem entirely unaware, is that it is, up to a scaling, merely the Student t-statistic for testing whether the mean of a random variable is zero.[2] In fact, the Sharpe ratio-test we now use, defined as

$$t =_{\text{df}} \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{n}} = \sqrt{n}\hat{\zeta}, \tag{1}$$

is *not* the form first considered by Gosset (writing as "Student").[7] Gosset originally analyzed the distribution of

$$z = \frac{\hat{\mu}}{s_N} = \frac{\hat{\mu}}{\hat{\sigma}\sqrt{(n-1)/n}} = \hat{\zeta}\sqrt{\frac{n}{n-1}},$$

where s_N is the "standard deviation of the sample," a biased estimate of the population standard deviation that uses n in the denominator instead of $n-1$. The connection to the t-distribution appears in Miller and Gehr's note on the bias of the Sharpe ratio, but has not been well developed. [15]

1. Sharpe guaranteed this ratio would be renamed by giving it the unweildy moniker of 'reward-to-variability,' yet another example of my Law of Implied Eponymy.

2. Sharpe himself seems to not make the connection, even though he quotes t-statistics for a regression fit in his original paper![22]

1.1 Distribution of the Sharpe ratio

Let $x_1, x_2, \ldots, x_n$ be *i.i.d.* draws from a normal distribution $\mathcal{N}(\mu, \sigma)$. Let $\hat{\mu} =_{\text{df}} \sum_i x_i/n$ and $\hat{\sigma}^2 =_{\text{df}} \sum_i (x_i - \hat{\mu})^2/(n-1)$ be the unbiased sample mean and variance, and let

$$t_0 =_{\text{df}} \sqrt{n}\frac{\hat{\mu} - \mu_0}{\hat{\sigma}}. \tag{2}$$

Then t_0 follows a non-central t-distribution with $n-1$ degrees of freedom and non-centrality parameter

$$\delta =_{\text{df}} \sqrt{n}\frac{\mu - \mu_0}{\sigma}.$$

Note the non-centrality parameter, δ, looks like the sample statistic t_0, but defined with population quantities. If $\mu = \mu_0$, then $\delta = 0$, and t_0 follows a central t-distribution. [9, 19]

Recalling that the modern t statistic is related to the Sharpe ratio by only a scaling of $\sqrt{n}$, the distribution of Sharpe ratio assuming normal returns follows a rescaled non-central t-distribution, where the non-centrality parameter depends only on the *signal-to-noise ratio* (hereafter 'SNR'), $\zeta =_{\text{df}} \mu/\sigma$, which is the population analogue of the Sharpe ratio, and the sample size.

Knowing the distribution of the Sharpe ratio is empowering, as interesting facts about the t-distribution or the t-test can be translated into interesting facts about the Sharpe ratio: one can construct hypothesis tests for the SNR, find the power and sample size of those tests, compute confidence intervals of the SNR, correct for deviations from assumptions, *etc.*

1.2 Tests involving the Sharpe ratio

There are a number of statistical tests involving the Sharpe ratio or variants thereupon.

1. The classical one-sample test for mean involves a t-statistic which is like a Sharpe ratio with constant benchmark. Thus to test the null hypothesis:
$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu > \mu_0,$$
we reject if the statistic
$$t_0 = \sqrt{n}\frac{\hat{\mu} - \mu_0}{\hat{\sigma}}$$
is greater than $t_{1-\alpha}\left(n-1\right)$, the $1-\alpha$ quantile of the (central) t-distribution with $n-1$ degrees of freedom.
If $\mu = \mu_1 > \mu_0$, then the power of this test is
$$1 - F_t\left(t_{1-\alpha}\left(n-1\right); \delta_1, n-1\right),$$
where $\delta_1 = \sqrt{n}\left(\mu_1 - \mu_0\right)/\sigma$ and $F_t\left(x; \delta, n-1\right)$ is the cumulative distribution function of the non-central t-distribution with non-centrality parameter δ and $n-1$ degrees of freedom. [19]
2. A one-sample test for signal-to-noise ratio (SNR) involves the t-statistic. To test:
$$H_0 : \zeta = \zeta_0 \quad \text{versus} \quad H_1 : \zeta > \zeta_0,$$
we reject if the statistic $t = \sqrt{n}\hat{\zeta}$ is greater than $t_{1-\alpha}\left(\delta_0, n-1\right)$, the $1-\alpha$ quantile of the non-central t-distribution with $n-1$ degrees of freedom and non-centrality parameter $\delta_0 = \sqrt{n}\zeta_0$.
If $\zeta = \zeta_1 > \zeta_0$, then the power of this test is
$$1 - F_t\left(t_{1-\alpha}\left(\delta_0, n-1\right); \delta_1, n-1\right),$$
where $\delta_1 = \sqrt{n}\zeta_1$ and $F_t\left(x; \delta, n-1\right)$ is the cumulative distribution function of the non-central t-distribution with non-centrality parameter δ and $n-1$ degrees of freedom. [19]

1.3 Moments of the Sharpe Ratio

Based on the moments of the non-central t-distribution, the expected value of the Sharpe ratio is *not* the signal-to-noise ratio (SNR), rather there is a systematic geometric bias. [26, 27] The t-statistic, which follows a non-central t-distribution with parameter δ and $n-1$ degrees of freedom has the following moments:

$$\begin{aligned} \mathrm{E}\left[t\right] &= \delta\sqrt{\frac{n-1}{2}}\frac{\Gamma\left((n-2)/2\right)}{\Gamma\left((n-1)/2\right)} = \delta c_n, \\ \mathrm{Var}\left(t\right) &= \frac{(1+\delta^2)(n-1)}{n-3} - \mathrm{E}\left[t\right]^2. \end{aligned} \tag{3}$$

Here $c_n = \sqrt{\frac{n-1}{2}}\Gamma\left((n-2)/2\right)/\Gamma\left((n-1)/2\right)$, is the 'bias term'. These can be trivially translated into equivalent facts regarding the Sharpe ratio:

$$\begin{aligned} \mathrm{E}\left[\hat{\zeta}\right] &= \zeta c_n, \\ \mathrm{Var}\left(\hat{\zeta}\right) &= \frac{(1+n\zeta^2)(n-1)}{n(n-3)} - \mathrm{E}\left[\hat{\zeta}\right]^2. \end{aligned} \tag{4}$$

The geometric bias term c_n does not equal one, thus the sample t statistic is a *biased* estimator of the non-centrality parameter, δ when $\delta \neq 0$, and the Sharpe ratio is a biased estimator of the signal-to-noise ratio when it is nonzero. [15] The bias term is a function of sample size only, and approaches one fairly quickly. However, there are situations in which it might be unacceptably large.

For example, if one was looking at one year's worth of data with monthly marks, one would have a fairly large bias: $c_n = 1.08$, *i.e.*, almost eight percent. The bias is multiplicative and larger than one, so the Sharpe ratio will overestimate the SNR when the latter is positive, and underestimate it when it is negative. The existence of this bias was first described by Miller and Gehr. [15]

A decent asymptotic approximation [1] to c_n is given by

$$c_{n+1} = 1 + \frac{3}{4n} + \frac{25}{32n^2} + \mathcal{O}\left(n^{-3}\right).$$

1.4 Asymptotics and Confidence Intervals

Lo showed that the Sharpe ratio is asymptotically normal in n with standard deviation $\sqrt{(1+\frac{\hat{\zeta}^2}{2})/n}$. [14] The equivalent result concerning the non-central t-distribution (which, again, is the Sharpe ratio up to scaling by $\sqrt{n}$) was published 60 years prior by Johnson and Welch. [9] Since the SNR, $\hat{\zeta}$, is unknown, Lo suggests approximating it with the Sharpe ratio, giving the following approximate $1-\alpha$ confidence interval on the SNR:

$$\hat{\zeta} \pm z_{\alpha/2}\sqrt{\frac{1+\frac{\hat{\zeta}^2}{2}}{n}},$$

where $z_{\alpha/2}$ is the $\alpha/2$ quantile of the normal distribution. In practice, the asymptotically equivalent form

$$\hat{\zeta} \pm z_{\alpha/2}\sqrt{\frac{1+\frac{\hat{\zeta}^2}{2}}{n-1}}$$

has better small sample coverage, at least for normal returns.

According to Walck,

$$\frac{t(1-\frac{1}{4(n-1)}) - \delta}{\sqrt{1+\frac{t^2}{2(n-1)}}}$$

is asymptotically (in n) a standard normal random variable, where t is the t-statistic, which is the Sharpe ratio up to scaling. [26]

This suggests the following approximate $1-\alpha$ confidence interval on the SNR:

$$\hat{\zeta}\left(1-\frac{1}{4(n-1)}\right) \pm z_{\alpha/2}\sqrt{\frac{1}{n}+\frac{\hat{\zeta}^2}{2(n-1)}}.$$

The normality results generally hold for large n, small ζ, and assume normality of x. [9] We can find confidence intervals on ζ assuming only normality of x (or large n and an appeal to the Central Limit Theorem), by inversion of the cumulative distribution of the non-central t-distribution. A $1-\alpha$ symmetric confidence interval on ζ has endpoints defined implicitly by

$$1-\alpha/2 = F_t\left(\hat{\zeta}; \sqrt{n}\zeta_l, n-1\right), \quad \alpha/2 = F_t\left(\hat{\zeta}; \sqrt{n}\zeta_u, n-1\right),$$

where $F_t\left(x; \delta, n-1\right)$ is the CDF of the non-central t-distribution with non-centrality parameter δ and $n-1$ degrees of freedom. Computationally, this method requires one to invert the CDF (*e.g.*, by Brent's method [4]), which is slower than approximations based on asymptotic normality.

In practice these three confidence interval approximations give very similar coverage, with no appreciable difference when $n > 30$ or so. For small sample sizes, the corrected form of Lo's approximation is slightly liberal (Lo's original formulation is too conservative).

There are approaches to estimating the standard error of the Sharpe ratio taking into account the third and higher moments of the returns. See Opdyke [16] or Baily and Lopez de Prado [2].

2 Sharpe ratio and portfolio optimization

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ be independent draws from a k-variate normal with population mean $\boldsymbol{\mu}$ and population covariance Σ. Let $\hat{\boldsymbol{\mu}}$ be the usual sample estimate of the mean, $\hat{\boldsymbol{\mu}} = \sum_i \boldsymbol{x}_i/n$, and let $\hat{\Sigma}$ be the usual sample estimate of the covariance,

$$\hat{\boldsymbol{\Sigma}} =_{\mathrm{df}} \frac{1}{n-1}\sum_i \left(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}\right)\left(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}\right)^\top.$$

Consider the unconstrained optimization problem

$$\max_{\hat{\boldsymbol{w}} : \hat{\boldsymbol{w}}^\top \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{w}} \le R^2} \frac{\hat{\boldsymbol{w}}^\top \hat{\boldsymbol{\mu}} - r_0}{\sqrt{\hat{\boldsymbol{w}}^\top \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{w}}}}, \tag{5}$$

where r_0 is the risk-free rate, and $R > 0$ is a risk 'budget'.

This problem has solution

$$\hat{\boldsymbol{w}}_* =_{\mathrm{df}} c\,\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}, \tag{6}$$

where the constant c is chosen to maximize return under the given risk budget:

$$c = \frac{R}{\sqrt{\hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\mu}}}}.$$

The Sharpe ratio of this portfolio is

$$\hat{\zeta}_* =_{\mathrm{df}} \frac{\hat{\boldsymbol{w}}_*{}^\top \hat{\boldsymbol{\mu}} - r_0}{\sqrt{\hat{\boldsymbol{w}}_*{}^\top \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{w}}_*}} = \sqrt{\hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\mu}}} - \frac{r_0}{R}. \tag{7}$$

The term $\frac{r_0}{R}$ is deterministic; we can treat it as an annoying additive constant that has to be minded. Define the population analogue of this quantity as

$$\zeta_* =_{\mathrm{df}} \sqrt{\boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu}} - \frac{r_0}{R}. \tag{8}$$

The random term, $n \left(\hat{\boldsymbol{\mu}}^\top \hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \right)^2$, is a Hotelling T^2, which follows a non-central F distribution, up to scaling:

$$\frac{n}{n-1} \frac{n-k}{k} \left(\hat{\zeta}_* + \frac{r_0}{R} \right)^2 \sim F \left(k, n-k, n \left(\zeta_* + \frac{r_0}{R} \right)^2 \right),$$

where $F(v_1, v_2, \delta)$ is the non-central F-distribution with v_1, v_2 degrees of freedom and non-centrality parameter δ. This allows us to make inference about ζ_*.

By using the 'biased' covariance estimate, defined as

$$\tilde{\Sigma} =_{\mathrm{df}} \frac{n-1}{n} \hat{\Sigma} = \frac{1}{n} \sum_i (\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}) (\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})^\top,$$

the above expression can be simplified slightly as

$$\frac{n-k}{k} \hat{\boldsymbol{\mu}}^\top \tilde{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \sim F \left(k, n-k, n \left(\zeta_* + \frac{r_0}{R} \right)^2 \right).$$

2.1 Asymptotics and Confidence Intervals

As noted in Section C, if F is distributed as a non-central F-distribution with v_1 and v_2 degrees of freedom and non-centrality parameter δ, then the mean of $\sqrt{F}$ is approximated by:

$$\mathrm{E}\left[\sqrt{F}\right] \approx \sqrt{\mathrm{E}[F]} - \frac{\frac{{v_2}^2 \left(\delta^2 + (v_1+2)\,(2\,\delta+v_1)\right)}{{v_1}^2\,(v_2-4)\,(v_2-2)} - (\mathrm{E}[F])^2}{8\,(\mathrm{E}[F])^{\frac{3}{2}}}, \tag{9}$$

where $\mathrm{E}[F] = \frac{v_2}{v_1} \frac{v_1+\delta}{v_2-2}$.

Now let $T^2 = n\hat{\zeta}_*^2$ be Hotelling's statistic with n observations of a p-variate vector returns series, and let ζ_* be the maximal SNR of a linear combination of the p populations. We know that

$$\frac{n-p}{p(n-1)} T^2 \sim F(\delta, p, n-p),$$

where the distribution has p and $n-p$ degrees of freedom, and $\delta = n\zeta_*^2$.

Substituting in the p and $n-p$ for v_1 and v_2, letting $p = cn$, and taking the limit as $n \to \infty$, we have

$$\mathrm{E}\left[\hat{\zeta}_*\right] = \sqrt{\frac{(n-1)p}{n(n-p)}}\, \mathrm{E}\left[\sqrt{F}\right] \to \sqrt{\frac{\zeta_*^2 + c}{1-c}},$$

which is approximately, but not exactly, equal to ζ_*. Note that if c becomes arbitrarily small (p is fixed while n grows without bound), then $\hat{\zeta}_*$ is asymptotically unbiased.

The asymptotic variance appears to be

$$\operatorname{Var}\left(\hat{\zeta}_*\right) \to \frac{\zeta_*^4 + 2\zeta_*^2 + c}{2n(1-c)^2(\zeta_*^2 + c)} \approx \frac{1+2c}{2n}\left(1 + \frac{1}{1 + c/\zeta_*^2}\right).$$

Consider as an example, the case where $p = 30$, $n = 1000\,\text{days}$, and $\zeta_* = 1.5\,\text{years}^{-0.5}$. Assuming 253 days per year, the expected value of $\hat{\zeta}_*$ is approximately $3.19\,\text{years}^{-0.5}$, with standard error around 0.41. This is a very serious bias. The problem is that the 'aspect ratio,' $c = p/n$, is quite a bit larger than ζ_*^2, and so it dominates the expectation. For real-world portfolios one expects ζ_*^2 to be no bigger than around $0.02\,\text{days}^{-1}$, and thus one should aim to have $n \gg 150p$, as a bare minimum (to achieve $\zeta_*^2 > 3c$, say). A more reasonable rule of thumb would be $n \geq 253p$, *i.e.*, at least one year of data per degree of freedom.

Using the asymptotic first moments of the Sharpe ratio gives only very rough approximate confidence intervals on ζ_*. The following are passable when $\zeta_*^2 \gg c$:

$$\hat{\zeta}_*\sqrt{1-c} - \frac{c}{2\hat{\zeta}_*} \pm z_\alpha \sqrt{\frac{2\hat{\zeta}_*^2 + c}{2n(1-c)(\hat{\zeta}_*^2 + c)}} \approx \hat{\zeta}_*\sqrt{1-c} - \frac{c}{2\hat{\zeta}_*} \pm z_\alpha \sqrt{\frac{1}{2n(1-c)}}$$

A better way to find confidence intervals is implicitly, by solving

$$\begin{aligned} 1 - \alpha/2 &= F_f\left(\left(\frac{n(n-p)}{p(n-1)}\right)\hat{\zeta}_*^2; n\zeta_l^2, p, n-p\right), \\ \alpha/2 &= F_f\left(\left(\frac{n(n-p)}{p(n-1)}\right)\hat{\zeta}_*^2; n\zeta_u^2, p, n-p\right), \end{aligned} \tag{10}$$

where $F_f\left(x; \delta, p, n-p\right)$ is the CDF of the non-central F-distribution with non-centrality parameter δ and p and $n-p$ degrees of freedom. This method requires computational inversion of the CDF function. Also, there may not be ζ_l or ζ_u such that the above hold with equality, so one is forced to use the limiting forms:

$$\begin{aligned} \zeta_l &= \min\left\{z \,\middle|\, z \geq 0,\ 1 - \alpha/2 \geq F_f\left(\left(\frac{n(n-p)}{p(n-1)}\right)\hat{\zeta}_*^2; nz^2, p, n-p\right)\right\}, \\ \zeta_u &= \min\left\{z \,\middle|\, z \geq 0,\ \alpha/2 \geq F_f\left(\left(\frac{n(n-p)}{p(n-1)}\right)\hat{\zeta}_*^2; nz^2, p, n-p\right)\right\}. \end{aligned} \tag{11}$$

Since $F_f\left(\cdot; nz^2, p, n-p\right)$ is a decreasing function of z^2, and approaches zero in the limit, the above confidence intervals are well defined.

2.2 Inference on SNR

Spruill gives a sufficient condition for the MLE of the non-centrality parameter to be zero, given a number of observations of random variables taking a non-central F distribution. [25] For the case of a single observation, the condition is particularly simple: if the random variable is no greater than one, the MLE of the non-centrality parameter is equal to zero. The equivalent fact about the optimal Sharpe ratio is that if $\hat{\zeta}_*^2 \leq \frac{c}{1-c}$, then the MLE of ζ_* is zero, where, again, $c = p/n$ is the 'aspect ratio.'

Using the expectation of the non-central F distribution, we can find an unbiased estimator of ζ_*^2. It is given by $(1-c)\hat{\zeta}_*^2 - c$. While this is unbiased for ζ_*^2, there is no guarantee that it is positive! Thus in practice, one should probably use the MLE of ζ_*^2, which is guaranteed to be non-negative, then take its square root to estimate ζ_*.

Kubokawa, Robert and Saleh give an improved method ('KRS'!) for estimating the non-centrality parameter given an observation of a non-central F statistic. [11]

2.3 Sharpe ratio and simple constrained portfolio optimization

Let G be an $k_g \times k$ matrix of rank $k_g \le k$. Let G^C be the matrix whose rows span the null space of the rows of G, *i.e.*, $\mathsf{G}^C\mathsf{G}^\top = 0$. Consider the constrained optimization problem

$$\max_{\hat{\boldsymbol{w}}:\mathsf{G}^C\hat{\boldsymbol{w}}=0,\,\hat{\boldsymbol{w}}^\top\hat{\Sigma}\hat{\boldsymbol{w}}\le R^2} \frac{\hat{\boldsymbol{w}}^\top\hat{\boldsymbol{\mu}} - r_0}{\sqrt{\hat{\boldsymbol{w}}^\top\hat{\Sigma}\hat{\boldsymbol{w}}}}, \tag{12}$$

where, as previously, $\hat{\boldsymbol{\mu}}$, $\hat{\Sigma}$ are the sample mean vector and covariance matrix, r_0 is the risk-free rate, and $R > 0$ is a risk 'budget'.

The gist of this constraint is that feasible portfolios must be some linear combination of the rows of G, or $\hat{\boldsymbol{w}} = \mathsf{G}^\top\hat{\boldsymbol{w}}_g$, for some unknown vector $\hat{\boldsymbol{w}}_g$. When viewed in this light, the constrained problem reduces to that of optimizing the portfolio on k_g assets with sample mean $\mathsf{G}\hat{\boldsymbol{\mu}}$ and sample covariance $\mathsf{G}\hat{\Sigma}\mathsf{G}^\top$. This problem has solution

$$\hat{\boldsymbol{w}}_{*,\mathsf{G}} =_{\mathrm{df}} c\,\mathsf{G}^\top\left(\mathsf{G}\hat{\Sigma}\mathsf{G}^\top\right)^{-1}\mathsf{G}\hat{\boldsymbol{\mu}}, \tag{13}$$

where the constant c is chosen to maximize return under the given risk budget, as in the unconstrained case. The Sharpe ratio of this portfolio is

$$\hat{\zeta}_{*,\mathsf{G}} =_{\mathrm{df}} \frac{{\hat{\boldsymbol{w}}_{*,\mathsf{G}}}^\top\hat{\boldsymbol{\mu}} - r_0}{\sqrt{{\hat{\boldsymbol{w}}_{*,\mathsf{G}}}^\top\hat{\Sigma}\hat{\boldsymbol{w}}_{*,\mathsf{G}}}} = \sqrt{(\mathsf{G}\hat{\boldsymbol{\mu}})^\top\left(\mathsf{G}\hat{\Sigma}\mathsf{G}^\top\right)^{-1}(\mathsf{G}\hat{\boldsymbol{\mu}})} - \frac{r_0}{R}. \tag{14}$$

Again, for purposes of estimating the population analogue, we can largely ignore, for simplicity of exposition, the deterministic 'drag' term r_0/R. As in the unconstrained case, the random term is a T^2 statistic, which can be transformed to a non-central F as

$$\frac{n}{n-1}\frac{n-k_g}{k_g}\left(\hat{\zeta}_{*,\mathsf{G}} + \frac{r_0}{R}\right)^2 \sim F\left(k_g, n-k_g, n\left(\zeta_{*,\mathsf{G}} + \frac{r_0}{R}\right)^2\right).$$

This allows us to make inference about $\zeta_{*,\mathsf{G}}$, the population analogue of $\hat{\zeta}_{*,\mathsf{G}}$.

2.4 Spanning and hedging

Consider the constrained portfolio optimization problem on k assets,

$$\max_{\hat{\boldsymbol{w}}:\mathsf{G}\hat{\Sigma}\hat{\boldsymbol{w}}=\boldsymbol{g},\,\hat{\boldsymbol{w}}^\top\hat{\Sigma}\hat{\boldsymbol{w}}\le R^2} \frac{\hat{\boldsymbol{w}}^\top\hat{\boldsymbol{\mu}} - r_0}{\sqrt{\hat{\boldsymbol{w}}^\top\hat{\Sigma}\hat{\boldsymbol{w}}}}, \tag{15}$$

where G is an $k_g \times k$ matrix of rank k_g, and, as previously, $\hat{\boldsymbol{\mu}}$, $\hat{\Sigma}$ are sample mean vector and covariance matrix, r_0 is the risk-free rate, and $R > 0$ is a risk 'budget'. We can interpret the G constraint as stating that the covariance of the returns of a feasible portfolio with the returns of a portfolio whose weights are in a given row of G shall equal the corresponding element of $\boldsymbol{g}$. In the garden variety application of this problem, G consists of k_g rows of the identity matrix, and $\boldsymbol{g}$ is the zero vector; in this case, feasible portfolios are 'hedged' with respect to the k_g assets selected by G (although they may hold some position in the hedged assets).

Assuming that the G constraint and risk budget can be simultaneously satisfied, the solution to this problem, via the Lagrange multiplier technique, is

$$\begin{aligned} \hat{\boldsymbol{w}}_* &= c\left(\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}} - \mathsf{G}^\top\left(\mathsf{G}\hat{\Sigma}\mathsf{G}^\top\right)^{-1}\mathsf{G}\hat{\boldsymbol{\mu}}\right) + \mathsf{G}^\top\left(\mathsf{G}\hat{\Sigma}\mathsf{G}^\top\right)^{-1}\boldsymbol{g}, \\ c^2 &= \frac{R^2 - \boldsymbol{g}^\top\left(\mathsf{G}\hat{\Sigma}\mathsf{G}^\top\right)\boldsymbol{g}}{\hat{\boldsymbol{\mu}}^\top\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}} - \left(\mathsf{G}\hat{\boldsymbol{\mu}}\right)^\top\left(\mathsf{G}\hat{\Sigma}\mathsf{G}^\top\right)^{-1}\left(\mathsf{G}\hat{\boldsymbol{\mu}}\right)}, \end{aligned} \tag{16}$$

where the numerator in the last equation need be positive for the problem to be feasible.

The case where $\boldsymbol{g} \neq 0$ is 'pathological', as it requires a fixed non-zero covariance of the target portfolio with some other portfolio's returns. Setting $\boldsymbol{g} = 0$ ensures the problem is feasible, and I will make this assumption hereafter. Under this assumption, the optimal portfolio is

$$\hat{\boldsymbol{w}}_* = c\left(\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}} - \mathsf{G}^\top\left(\mathsf{G}\hat{\Sigma}\mathsf{G}^\top\right)^{-1}\mathsf{G}\hat{\boldsymbol{\mu}}\right) = c_1\hat{\boldsymbol{w}}_{*,\mathsf{I}} - c_2\hat{\boldsymbol{w}}_{*,\mathsf{G}},$$

using the notation from Section 2.3. Note that, up to scaling, $\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}$ is the unconstrained optimal portfolio, and thus the imposition of the G constraint only changes the unconstrained portfolio in assets corresponding to columns of G containing non-zero elements. In the garden variety application where G is a single row of the identity matrix, the imposition of the constraint only changes the holdings in the asset to be hedged (modulo changes in the leading constant to satisfy the risk budget).

The squared Sharpe ratio of the optimal portfolio is

$$\hat{\zeta}_*^2 = \hat{\boldsymbol{\mu}}^\top\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}} - \left(\mathsf{G}\hat{\boldsymbol{\mu}}\right)^\top\left(\mathsf{G}\hat{\Sigma}\mathsf{G}^\top\right)^{-1}\left(\mathsf{G}\hat{\boldsymbol{\mu}}\right) = \hat{\zeta}_{*,\mathsf{I}}^2 - \hat{\zeta}_{*,\mathsf{G}}^2, \tag{17}$$

using the notation from Section 2.3, and setting $r_0 = 0$.

Some natural questions to ask are

1. Does the imposition of the G constraint cause a material decrease in Sharpe ratio? Can we estimate the size of the drop?

   Performing the same computations on the population analogues (*i.e.*, $\boldsymbol{\mu}$, Σ), we have $\zeta_*^2 = \zeta_{*,\mathsf{I}}^2 - \zeta_{*,\mathsf{G}}^2$, and thus the drop in squared Signal-noise ratio by imposing the G hedge constraint is equal to $\zeta_{*,\mathsf{G}}^2$. We can perform inference on this quantity by considering the statistic $\hat{\zeta}_{*,\mathsf{G}}^2$, as in the previous section.

2. Is the constrained portfolio 'good'? Formally we can test the hypothesis $H_0 : \zeta^2_{*,\mathsf{I}} - \zeta^2_{*,\mathsf{G}} = 0$, or find point or interval estimates of $\zeta^2_{*,\mathsf{I}} - \zeta^2_{*,\mathsf{G}}$.

   This generalizes the known tests of *portfolio spanning.* [10, 8] A spanning test considers whether the optimal portfolio on a pre-fixed subset of k_g assets has the same Sharpe ratio as the optimal portfolio on all k assets, *i.e.*, whether those k_g assets 'span' the set of all assets.

   If you let G be the $k_g \times k$ matrix consisting of the k_g rows of the identity matrix corresponding to the k_g assets to be tested for spanning, then the term

$$\hat{\zeta}^2_{*,\mathsf{G}} = \left(\mathsf{G}\hat{\boldsymbol{\mu}}\right)^\top \left(\mathsf{G}\hat{\Sigma}\mathsf{G}^\top\right)^{-1} \left(\mathsf{G}\hat{\boldsymbol{\mu}}\right)$$

   is the squared Sharpe ratio of the optimal portfolio on only the k_g spanning assets. A spanning test is then a test of the hypothesis

$$H_0 : \zeta^2_{*,\mathsf{I}} = \zeta^2_{*,\mathsf{G}}.$$

   The test statistic

$$F_{\mathsf{G}} = \frac{n-k}{k-k_g} \frac{\hat{\zeta}^2_{*,\mathsf{I}} - \hat{\zeta}^2_{*,\mathsf{G}}}{\frac{n-1}{n} + \hat{\zeta}^2_{*,\mathsf{G}}} \tag{18}$$

   was shown by Rao to follow an F distribution under the null hypothesis. [18] Giri showed that, under the alternative, and conditional on observing $\hat{\zeta}^2_{*,\mathsf{G}}$,

$$F_{\mathsf{G}} \sim F\left(k - k_g, n-k, \frac{n}{1 + \frac{n}{n-1}\hat{\zeta}^2_{*,\mathsf{G}}}\left(\zeta^2_{*,\mathsf{I}} - \zeta^2_{*,\mathsf{G}}\right)\right), \tag{19}$$

   where $F\left(v_1, v_2, \delta\right)$ is the non-central F-distribution with v_1, v_2 degrees of freedom and non-centrality parameter δ. See Section D. [6]

2.5 Optimal Sharpe ratio under positivity constraint

Consider the following portfolio optimization problem:

$$\max_{\hat{\boldsymbol{w}}:\hat{\boldsymbol{w}}\geq 0,\, \hat{\boldsymbol{w}}^\top\hat{\Sigma}\hat{\boldsymbol{w}}\leq R^2} \frac{\hat{\boldsymbol{w}}^\top\hat{\boldsymbol{\mu}} - r_0}{\sqrt{\hat{\boldsymbol{w}}^\top\hat{\Sigma}\hat{\boldsymbol{w}}}}, \tag{20}$$

where the constraint $\hat{\boldsymbol{w}} \geq 0$ is to be interpreted element-wise. In general, the optimal portfolio, call it $\hat{\boldsymbol{w}}_{*,+}$, must be found numerically.[3]

The squared Sharpe ratio of the portfolio $\hat{\boldsymbol{w}}_{*,+}$ has value

$$\hat{\zeta}^2_{*,+} = \frac{\left({\hat{\boldsymbol{w}}_{*,+}}^\top\hat{\boldsymbol{\mu}}\right)^2}{{\hat{\boldsymbol{w}}_{*,+}}^\top\hat{\Sigma}\hat{\boldsymbol{w}}_{*,+}}.$$

The statistic $n\hat{\zeta}^2_{*,+}$, which is a constrained Hotelling T^2, has been studied to test the hypothesis of zero multivariate mean against an inequality-constrained alternative hypothesis. [23, 21]

Unfortunately, $\hat{\zeta}^2_{*,+}$ is not a *similar* statistic. That is, its distribution depends on the population analogue, $\zeta^2_{*,+}$, but also on the uknown nuisance parameter, Σ. And so using $\hat{\zeta}^2_{*,+}$ to test the hypothesis $H_0 : \zeta^2_{*,+} = 0$ only yields

3. Unless, by some miracle, the unconstrained optimal portfolio happens to satisfy the positivity constraint.

a conservative test, with a maximal type I rate. Intuitively, the Hotelling T^2, which is invariant with respect to an invertible transform, should not mix well with the positive-orthant constraint, which is not invariant.

One consequence of non-similarity is that using in-sample Sharpe ratio as a yardstick of the quality of so constrained portfolio is unwise. For one can imagine universe A, containing of two zero-mean assets, and universe B with two assets with positive mean, where the different covariances in the two universes implies that the sample optimal constrained Sharpe ratio is likely to be larger in universe A than in universe B.

3 Miscellanea

3.1 Which Returns?

There is often some confusion regarding the form of returns (*i.e.*, log returns or 'relative' returns) to be used in computation of the Sharpe ratio. Usually log returns are recommended because they aggregate over time by summation (*e.g.*, the sum of a week's worth of daily log returns is the weekly log return), and thus taking the mean of them is considered sensible. For this reason, adjusting the time frame (*e.g.*, annualizing) of log returns is trivial.

However, relative returns have the property that they are additive 'laterally': the relative return of a portfolio on a given day is the dollar-weighted mean of the relative returns of each position. This property is important when one considers more general attribution models, or Hotelling's distribution. To make sense of the sums of relative returns one can think of a fund manager who always invests a fixed amount of capital, siphoning off excess returns into cash, or borrowing[4] cash to purchase stock. Under this formulation, the returns aggregate over time by summation just like log returns.

4. at no interest!

One reason fund managers might use relative returns when reporting Sharpe ratio is because it inflates the results! The 'boost' from computing Sharpe using relative returns is approximately:

$$\frac{\hat{\zeta}_r - \hat{\zeta}}{\hat{\zeta}} \approx \frac{1}{2} \frac{\sum_i x^2}{\sum_i x}, \tag{21}$$

where $\hat{\zeta}_r$ is the Sharpe measured using relative returns and $\hat{\zeta}$ uses log returns. This approximation is most accurate for daily returns, and for the modest values of Sharpe ratio one expects to see for real funds.

3.2 Sharpe is nearly leverage invariant

Suppose that you observe the returns of a strategy, but the fund manager is changing the leverage from period to period. Suppose the fund manager's decisions are completely uninformed[5], and so that changes in leverage are completely independent from the future performance of the underlying strategy. Can one compute the Sharpe ratio on the observed returns without adjusting for leverage (which may be unknown)?

5. I know this is a stretch...

Given some modest conditions, one can indeed. Let l_i be the leverage on period i, and let $l_i x_i$ be the observed levered returns[6]. Suppose that l_i and x_i

6. Here one must use relative returns instead of log returns.

are independent random variables and $l_i > 0$. We have

$$\begin{aligned} \mathrm{E}\left[lx\right] &= \mathrm{E}\left[l\right]\mathrm{E}\left[x\right], \\ \mathrm{Var}\left(lx\right) &= \mathrm{E}\left[l^2\right]\mathrm{E}\left[x^2\right] - \mathrm{E}\left[l\right]^2\mathrm{E}\left[x\right]^2 = \mathrm{E}\left[x^2\right]\mathrm{Var}\left(l\right) + \mathrm{Var}\left(x\right)\mathrm{E}\left[l\right]^2, \end{aligned} \tag{22}$$

And thus, with some rearrangement,

$$\zeta_{lx} = \frac{\zeta_x}{\sqrt{1 + \frac{\mathrm{E}[x^2]}{\mathrm{Var}(x)}\frac{\mathrm{Var}(l)}{\mathrm{E}[l]^2}}}.$$

Thus measuring Sharpe ratio without adjusting for leverage tends to give underestimates of the 'true' Sharpe ratio of the returns series. However, the deflation is probably very modest indeed.

Note that when looking at *e.g.*, daily returns, the (non-annualized) Sharpe ratio on the given mark frequency is usually on the order of 0.1 or less, thus $\mathrm{E}\left[x\right]^2 \approx 0.01\mathrm{Var}\left(x\right)$, and so $\mathrm{E}\left[x^2\right] \approx 1.01\mathrm{Var}\left(x\right)$. Thus it suffices to estimate the ratio $\mathrm{Var}\left(l\right)/\mathrm{E}\left[l\right]^2$, the squared *coefficient of variation* of l, to compute the correction factor.

Consider, for example, the case where l is the `VIX` index. Empirically the `VIX` has a coefficient of variation around 0.4. Assuming the daily Sharpe ratio is 0.1, we have

$$\sqrt{1 + \frac{\mathrm{E}\left[x^2\right]}{\mathrm{Var}\left(x\right)}\frac{\mathrm{Var}\left(l\right)}{\mathrm{E}\left[l\right]^2}} \approx 1.08.$$

In this case the correction factor for leverage is fairly small.

3.3 The 'haircut'

Care must be taken interpreting the confidence intervals and the estimated optimal SNR of a portfolio. This is because ζ_* is the *maximal* population SNR achieved by any portfolio; it is at least equal to, and potentially much larger than, the SNR achieved by the portfolio based on sample statistics, $\hat{\boldsymbol{w}}_*$. There is a gap or 'haircut' due to mis-estimation of the optimal portfolio. One would suspect that this gap is worse when the true effect size (*i.e.*, ζ_*) is smaller, when there are fewer observations (n), and when there are more assets (p).

Assuming $\boldsymbol{\mu}$ is not all zeros, define the haircut as the quantity

$$h =_{\mathrm{df}} 1 - \frac{1}{\zeta_*}\frac{{\hat{\boldsymbol{w}}_*}^{\top}\boldsymbol{\mu}}{\sqrt{{\hat{\boldsymbol{w}}_*}^{\top}\Sigma\hat{\boldsymbol{w}}_*}} = 1 - \left(\frac{{\hat{\boldsymbol{w}}_*}^{\top}\boldsymbol{\mu}}{{\boldsymbol{\nu}_*}^{\top}\boldsymbol{\mu}}\right)\left(\frac{\sqrt{{\boldsymbol{\nu}_*}^{\top}\Sigma\boldsymbol{\nu}_*}}{\sqrt{{\hat{\boldsymbol{w}}_*}^{\top}\Sigma\hat{\boldsymbol{w}}_*}}\right), \tag{23}$$

where $\boldsymbol{\nu}_*$ is the population optimal portfolio, positively proportional to $\Sigma^{-1}\boldsymbol{\mu}$. Thus the haircut is one minus the ratio of population SNR achieved by the sample Markowitz portfolio to the optimal population SNR (which is achieved by the population Markowitz portfolio). A smaller value means that the sample portfolio achieves a larger proportion of possible SNR, or, equivalently, a larger value of the haircut means greater mis-estimation of the optimal portfolio. The haircut takes values in $[0, 2]$.

Modeling the haircut is not straightforward because it is a random quantity which is not observed. That is, it mixes the unknown population parameters Σ and $\boldsymbol{\mu}$ with the sample quantity $\hat{\boldsymbol{w}}_*$, which is random.

When n/p is large, the following is a reasonable approximation to the distribution of h:

$$\sqrt{p-1}\tan\left(\arcsin\left(1-h\right)\right) \approx t\left(\sqrt{n}\zeta_{*}, p-1\right), \tag{24}$$

where $t(x, y)$ is a non-central t-distribution with non-centrality parameter x and y degrees of freedom. This approximation can be found by ignoring all variability in the sample estimate of the covariance matrix, that is by assuming that the sample optimal portfolio was computed with the *population* covariance: $\hat{\boldsymbol{w}}_{*} \propto \boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}$. Because mis-estimation of the covariance matrix should contribute some error, I expect that this approximation is a 'stochastic lower bound' on the true haircut. Numerical simulations, however, suggest it is a fairly tight bound for large n/p. (I would be willing to guess that the true distribution involves a non-central F-distribution, but the proof is beyond me at the moment.)

Here I look at the haircut via Monte Carlo simulations:

```
require(MASS)
# simple markowitz.
simple.marko <- function(rets) {
    mu.hat <- as.vector(apply(rets, MARGIN = 2, mean,
        na.rm = TRUE))
    Sig.hat <- cov(rets)
    w.opt <- solve(Sig.hat, mu.hat)
    retval <- list(mu = mu.hat, sig = Sig.hat, w = w.opt)
    return(retval)
}
# make multivariate pop. & sample w/ given
# zeta.star
gen.pop <- function(n, p, zeta.s = 0) {
    true.mu <- matrix(rnorm(p), ncol = p)
    # generate an SPD population covariance. a hack.
    xser <- matrix(rnorm(p * (p + 100)), ncol = p)
    true.Sig <- t(xser) %*% xser
    pre.sr <- sqrt(true.mu %*% solve(true.Sig, t(true.mu)))
    # scale down the sample mean to match the zeta.s
    true.mu <- (zeta.s/pre.sr[1]) * true.mu
    X <- mvrnorm(n = n, mu = true.mu, Sigma = true.Sig)
    retval = list(X = X, mu = true.mu, sig = true.Sig,
        SNR = zeta.s)
    return(retval)
}
# a single simulation
sample.haircut <- function(n, p, ...) {
    popX <- gen.pop(n, p, ...)
    smeas <- simple.marko(popX$X)
    # I have got to figure out how to deal with
    # vectors...
    ssnr <- (t(smeas$w) %*% t(popX$mu))/sqrt(t(smeas$w) %*%
        popX$sig %*% smeas$w)
    hcut <- 1 - (ssnr/popX$SNR)
    # for plugin estimator, estimate zeta.star
    asro <- sropt(z.s = sqrt(t(smeas$w) %*% smeas$mu),
        df1 = p, df2 = n)
```

```r
    zeta.hat.s <- inference(asro, type = "KRS")  # or 'MLE', 'unbiased'
    return(c(hcut, zeta.hat.s))
}
# set everything up
set.seed(as.integer(charToRaw("496509a9-dd90-4347-aee2-1de6d3635724")))
ope <- 253
LONG.FORM <- FALSE
n.sim <- if (LONG.FORM) 2048 else 512
n.stok <- if (LONG.FORM) 8 else 6
n.yr <- 4
n.obs <- ceiling(ope * n.yr)
zeta.s <- 1.2/sqrt(ope)  # optimal SNR, in daily units
# run some experiments
system.time(experiments <- replicate(n.sim, sample.haircut(n.obs,
    n.stok, zeta.s)))

##    user  system elapsed
##    0.87    0.00    0.87

hcuts <- experiments[1, ]
print(summary(hcuts))

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.01    0.16    0.24    0.29    0.39    1.14

# haircut approximation in the equation above
qhcut <- function(p, df1, df2, zeta.s, lower.tail = TRUE) {
    1 - sin(atan((1/sqrt(df1 - 1)) * qt(p, df = df1 -
        1, ncp = sqrt(df2) * zeta.s, lower.tail = !lower.tail)))
}
# if you wanted to look at how bad the plug-in
# estimator is, then uncomment the following (you
# are warned): zeta.hat.s <- experiments[2,];
# qqplot(qhcut(ppoints(length(hcuts)),n.stok,n.obs,zeta.hat.s),hcuts,
# xlab = 'Theoretical Approximate Quantiles', ylab
# = 'Sample Quantiles');
# qqline(hcuts,datax=FALSE,distribution =
# function(p) { qhcut(p,n.stok,n.obs,zeta.hat.s) },
# col=2)
# qqplot;
qqplot(qhcut(ppoints(length(hcuts)), n.stok, n.obs,
    zeta.s), hcuts, xlab = "Theoretical Approximate Quantiles",
    ylab = "Sample Quantiles")
qqline(hcuts, datax = FALSE, distribution = function(p) {
    qhcut(p, n.stok, n.obs, zeta.s)
}, col = 2)
```

I check the quality of the approximation given in Equation 24 by a Q-Q plot in Figure 1. For the case where $n = 1012$ (4 years of daily observations), $p = 6$ and $\zeta_* = 1.2\text{yr}^{-1/2}$, the t-approximation is very good indeed.

The median value of the haircut is on the order of 24%, meaning that the median population SNR of the sample portfolios is around $0.91\text{yr}^{-1/2}$. The maximum value of the haircut over the 512 simulations, however is 1.14, which is larger than one; this happens if and only if the sample portfolio has negative

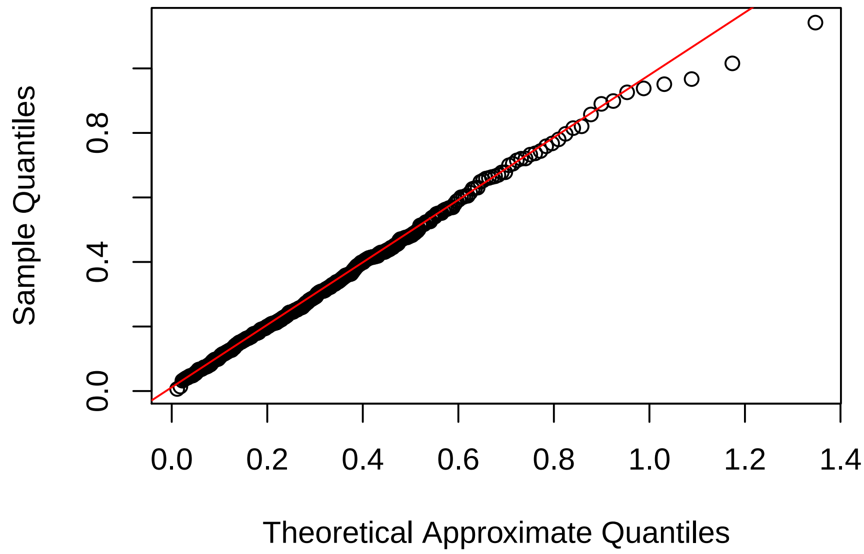

Figure 1: Q-Q plot of 512 simulated haircut values versus the approximation given by Equation 24 is shown.

expected return: $\hat{\boldsymbol{w}}_*{}^\top \boldsymbol{\mu} < 0$. In this case the Markowitz portfolio is actually *destroying value* because of modeling error: the mean return of the selected portfolio is negative, even though positive mean is achievable.

The approximation in Equation 24 involves the unknown population parameters $\boldsymbol{\mu}$ and Σ, but does not make use of the observed quantities $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$. It seems mostly of theoretical interest, perhaps for producing prediction intervals on h when planning a trading strategy (*i.e.*, balancing n and p). A more practical problem is that of estimating confidence intervals on $\hat{\boldsymbol{w}}^\top \boldsymbol{\mu} / \sqrt{\hat{\boldsymbol{w}}^\top \Sigma^{-1} \hat{\boldsymbol{w}}}$ having observed $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$. In this case one *cannot* simply plug-in some estimate of ζ_* computed from $\hat{\zeta}_*$ (via MLE, KRS, *etc.*) into Equation 24. The reason is that the error in the approximation of ζ_* is not independent of the modeling error that causes the haircut.

References

[1] Digital library of mathematical functions. http://dlmf.nist.gov/5.11#iii, 2011. URL http://dlmf.nist.gov/5.11#iii.

[2] David~H. Bailey and Marcos Lopez~de Prado. The Sharpe Ratio Efficient Frontier. *Social Science Research Network Working Paper Series*, April 2011. URL http://ssrn.com/abstract=1821643.

[3] Martin Bilodeau and David Brenner. *Theory of Multivariate Statistics (Springer Texts in Statistics)*. Springer, 1 edition, August 1999. ISBN 0387987398. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.172.3290.

[4] Richard~P. Brent. *Algorithms for Minimization without Derivatives*. Prentice-Hall, Englewood Cliffs, N.J., 1973.

[5] Mark Britten-Jones. The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance*, 54(2):655–671, 1999. URL http://www.jstor.org/stable/2697722.

[6] N~Giri. On the likelihood ratio test of a normal multivariate testing problem. *The Annals of Mathematical Statistics*, 35(1):181–189, 1964. doi: 10.1214/aoms/1177703740. URL http://projecteuclid.org/euclid.aoms/1177703740.

[7] William~Sealy Gosset. The probable error of a mean. *Biometrika*, 6(1):1–25, March 1908. URL http://dx.doi.org/10.2307/2331554. Originally published under the pseudonym "Student".

[8] Gur Huberman and Shmuel Kandel. Mean-variance spanning. *The Journal of Finance*, 42(4):pp. 873–888, 1987. ISSN 00221082. URL http://www.jstor.org/stable/2328296.

[9] N.~L. Johnson and B.~L. Welch. Applications of the non-central t-distribution. *Biometrika*, 31(3-4):362–389, March 1940. doi: 10.1093/biomet/31.3-4.362. URL http://dx.doi.org/10.1093/biomet/31.3-4.362.

[10] Raymond Kan and GuoFu Zhou. Tests of mean-variance spanning. *Annals of Economics and Finance*, 13(1), 2012. URL `http://www.aeconf.net/Articles/May2012/aef130105.pdf`.

[11] Tatsuya Kubokawa, C.~P. Robert, and A.~K. Md.~E. Saleh. Estimation of noncentrality parameters. *Canadian Journal of Statistics*, 21(1):45–57, 1993.

[12] Bruno Lecoutre. Another look at confidence intervals for the noncentral T distribution. *Journal of Modern Applied Statistical Methods*, 6(1):107–116, 2007. URL `http://www.univ-rouen.fr/LMRS/Persopage/Lecoutre/telechargements/Lecoutre_Another_look-JMSAM2007_6(1).pdf`.

[13] Olivier Ledoit and Michael Wolf. Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859, Dec 2008. ISSN 0927-5398. doi: http://dx.doi.org/10.1016/j.jempfin.2008.03.002. URL `http://www.ledoit.net/jef2008_abstract.htm`.

[14] Andrew~W. Lo. The Statistics of Sharpe Ratios. *Financial Analysts Journal*, 58(4), July/August 2002. URL `http://ssrn.com/paper=377260`.

[15] Robert~E. Miller and Adam~K. Gehr. Sample size bias and Sharpe's performance measure: A note. *Journal of Financial and Quantitative Analysis*, 13(05):943–946, 1978. doi: 10.2307/2330636. URL `http://dx.doi.org/10.1017/S0022109000014216`.

[16] J.~D. Opdyke. Comparing sharpe ratios: So where are the p-values? *Journal of Asset Management*, 8(5), 2007. URL `http://ssrn.com/abstract=886728`.

[17] M.~D. Perlman. One-sided testing problems in multivariate analysis. *The Annals of Mathematical Statistics*, 40(2):549–567, 1969. URL `http://www.jstor.org/stable/2239474`.

[18] C.~Radhakrishna Rao. *Advanced Statistical Methods in Biometric Research*. John Wiley and Sons, 1952.

[19] Fritz Scholz. Applications of the noncentral t-distribution. April 2007. URL `http://www.stat.washington.edu/fritz/DATAFILES498B2008/NoncentralT.pdf`.

[20] Pranab~Kumar Sen and Mervyn~J. Silvapulle. An appraisal of some aspects of statistical inference under inequality constraints. *Journal of Statistical Planning and Inference*, 107(1 - 2):3–43, 2002. ISSN 0378-3758. doi: 10.1016/S0378-3758(02)00242-2. URL `http://www.sciencedirect.com/science/article/pii/S0378375802002422`.

[21] Pranab~Kumar Sen and Ming-Tien Tsai. Two-stage likelihood ratio and union-intersection tests for one-sided alternatives multivariate mean with nuisance dispersion matrix. *Journal of Multivariate Analysis*, 68(2):264–282, 1999. ISSN 0047-259X. doi: 10.1006/jmva.1998.1791. URL `http://www.sciencedirect.com/science/article/pii/S0047259X98917912`.

[22] William~F. Sharpe. Mutual fund performance. *Journal of Business*, 39:119, 1965. URL `http://ideas.repec.org/a/ucp/jnlbus/v39y1965p119.html`.

[23] Mervyn~J. Silvapulle. A Hotelling's T^2-type statistic for testing against one-sided hypotheses. *J. Multivar. Anal.*, 55(2):312–319, November 1995. ISSN 0047-259X. doi: 10.1006/jmva.1995.1081. URL `http://dx.doi.org/10.1006/jmva.1995.1081`.

[24] Mervyn~J. Silvapulle and Pranab~Kumar Sen. *Constrained statistical inference : inequality, order, and shape restrictions*. Wiley-Interscience, Hoboken, N.J., 2005. ISBN 0471208272.

[25] M.~C. Spruill. Computation of the maximum likelihood estimate of a noncentrality parameter. *Journal of Multivariate Analysis*, 18(2):216–224, 1986. ISSN 0047-259X. doi: 10.1016/0047-259X(86)90070-9. URL `http://www.sciencedirect.com/science/article/pii/0047259X86900709`.

[26] Christian Walck. *Hand-book on STATISTICAL DISTRIBUTIONS for experimentalists*. December 1996. URL `http://www.stat.rice.edu/~dobelman/textfiles/DistributionsHandbook.pdf`.

[27] Wikipedia. Noncentral t-distribution — wikipedia, the free encyclopedia. `http://en.wikipedia.org/w/index.php?title=Noncentral_t-distribution&oldid=423475834`, 2011. URL `http://en.wikipedia.org/w/index.php?title=Noncentral_t-distribution&oldid=423475834`. [Online; accessed 18-April-2011].

A Asymptotic Efficiency of Sharpe Ratio

Suppose that $x_1, x_2, \ldots, x_n$ are drawn *i.i.d.* from a normal distribution with unknown SNR and variance. Suppose one has an (vector) estimator of the SNR and the variance. The Fisher information matrix can easily be shown to be:

$$\mathcal{I}\left(\zeta, \sigma\right) = n \left(\begin{array}{cc} 1 & \frac{\zeta}{2\sigma^2} \\ \frac{\zeta}{2\sigma^2} & \frac{2+\zeta^2}{4\sigma^4} \end{array}\right) \tag{25}$$

Inverting the Fisher information matrix gives the Cramer-Rao lower bound for an unbiased vector estimator of SNR and variance:

$$\mathcal{I}^{-1}\left(\zeta, \sigma\right) = \frac{1}{n} \left(\begin{array}{cc} 1 + \zeta^2/2 & -\zeta\sigma^2 \\ -\zeta\sigma^2 & 2\sigma^4 \end{array}\right) \tag{26}$$

Now consider the estimator $\left[\tilde{\zeta}, \hat{\sigma}^2\right]^\top$. This is an unbiased estimator for $\left[\zeta, \sigma^2\right]^\top$. One can show that the variance of this estimator is

$$\operatorname{Var}\left(\left[\tilde{\zeta}, \hat{\sigma}^2\right]^\top\right) = \left(\begin{array}{cc} \frac{(1+n\zeta^2)(n-1)}{{c_n}^2 n(n-3)} - \zeta^2 & \zeta\sigma^2\left(\frac{1}{c_n} - 1\right) \\ \zeta\sigma^2\left(\frac{1}{c_n} - 1\right) & \frac{2\sigma^4}{n-1} \end{array}\right). \tag{27}$$

The variance of $\tilde{\zeta}$ follows from Equation 4. The cross terms follow from the independence of the sample mean and variance, and from the unbiasedness of the two estimators. The variance of $\hat{\sigma}^2$ is well known.

Since $c_n = 1+\frac{3}{4(n-1)}+\mathcal{O}\left(n^{-2}\right)$, the asymptotic variance of $\tilde{\zeta}$ is $\frac{(n-1)+\frac{n}{2}\zeta^2}{(n+(3/2))(n-3)}+\mathcal{O}\left(n^{-2}\right)$, and the covariance of $\tilde{\zeta}$ and $\hat{\sigma}^2$ is $-\zeta\hat{\sigma}^2\frac{3}{4n}+\mathcal{O}\left(n^{-2}\right)$. Thus the estimator $\left[\tilde{\zeta}, \hat{\sigma}^2\right]^\top$ is asymptotically *efficient*, *i.e.*, it achieves the Cramer-Rao lower bound asymptotically.

B Some Moments

It is convenient to have the first two moments of some common distributions.

Suppose F is distributed as a non-central F-distribution with v_1 and v_2 degrees of freedom and non-centrality parameter δ, then the mean and variance of F are [26]:

$$\begin{aligned} \mathrm{E}\left[F\right] &= \frac{v_2}{v_1}\frac{v_1+\delta}{v_2-2},\\ \mathrm{Var}\left(F\right) &= \left(\frac{v_2}{v_1}\right)^2 \frac{2}{(v_2-2)(v_2-4)}\left(\frac{(\delta+v_1)^2}{v_2-2}+2\delta+v_1\right). \end{aligned} \tag{28}$$

Suppose T^2 is distributed as a (non-central) Hotelling's statistic for n observations on p assets, with non-centrality parameter δ. Then [3]

$$\frac{n-p}{p(n-1)}T^2 = F$$

takes a non-central F-distribution with $v_1 = p$ and $v_2 = n-p$ degrees of freedom. Then we have the following moments:

$$\begin{aligned} \mathrm{E}\left[T^2\right] &= \frac{(n-1)\left(p+\delta\right)}{n-p-2},\\ \mathrm{Var}\left(T^2\right) &= \frac{2\left(n-1\right)^2}{(n-p-2)(n-p-4)}\left(\frac{(\delta+p)^2}{n-p-2}+2\delta+p\right). \end{aligned} \tag{29}$$

Suppose $\hat{\zeta}_*^2$ is the maximal Sharpe ratio on a basket of p assets with n observations, assuming *i.i.d.* Gaussian errors. Then $n\hat{\zeta}_*^2$ is distributed as a non-central Hotelling statistic, and we have the following moments:

$$\begin{aligned} \mathrm{E}\left[\hat{\zeta}_*^2\right] &= \frac{n-1}{n}\frac{\left(p+n\zeta_*^2\right)}{n-p-2} = \left(1-\frac{1}{n}\right)\frac{\left(c+\zeta_*^2\right)}{1-c-\frac{2}{n}},\\ \mathrm{Var}\left(\hat{\zeta}_*^2\right) &= \left(\frac{n-1}{n}\right)^2\frac{2}{(n-p-2)(n-p-4)}\left(\frac{(n\zeta_*^2+p)^2}{n-p-2}+2n\zeta_*^2+p\right),\\ &= \left(1-\frac{1}{n}\right)^2\frac{1}{n}\frac{2}{(1-c-\frac{2}{n})(1-c-\frac{4}{n})}\left(\frac{(\zeta_*^2+c)^2}{1-c-\frac{2}{n}}+2\zeta_*^2+c\right), \end{aligned} \tag{30}$$

where $c = p/n$ is the aspect ratio, and ζ_*^2 is the maximal SNR achievable on a basket of the assets: $\zeta_*^2 = \boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$.

C Square Root F

If F is distributed as a non-central F-distribution with v_1 and v_2 degrees of freedom and non-centrality parameter δ, then the mean and variance of F are [26]:

$$\begin{aligned} \mathrm{E}\left[F\right] &= \frac{v_2}{v_1}\frac{v_1+\delta}{v_2-2}, \\ \mathrm{Var}\left(F\right) &= \left(\frac{v_2}{v_1}\right)^2 \frac{2}{(v_2-2)(v_2-4)}\left(\frac{(\delta+v_1)^2}{v_2-2}+2\delta+v_1\right). \end{aligned} \tag{31}$$

Using the Taylor series expansion of the square root gives the approximate mean of $\sqrt{F}$:

$$\mathrm{E}\left[\sqrt{F}\right] \approx \sqrt{\mathrm{E}\left[F\right]} - \frac{\frac{{v_2}^2\left(\delta^2+(v_1+2)\,(2\,\delta+v_1)\right)}{{v_1}^2\,(v_2-4)\,(v_2-2)} - \left(\mathrm{E}\left[F\right]\right)^2}{8\,\left(\mathrm{E}\left[F\right]\right)^{\frac{3}{2}}}. \tag{32}$$

D Untangling Giri

Here I translate Giri's work on Rao's LRT into the terminology used in the rest of this note. [6] In equation (1.9), Giri defines the LRT statistic Z by

$$Z =_{\mathrm{df}} \frac{1 - N\bar{X}_{[2]}^{\top}\left(S_{22} + N\bar{X}_{[2]}\bar{X}_{[2]}^{\top}\right)^{-1}\bar{X}_{[2]}}{1 - N\bar{X}_{[1]}^{\top}\left(S_{11} + N\bar{X}_{[1]}\bar{X}_{[1]}^{\top}\right)^{-1}\bar{X}_{[1]}}. \tag{33}$$

Simply applying the Woodbury formula, we have

$$\begin{aligned} \left(S_{11} + N\bar{X}_{[1]}\bar{X}_{[1]}^{\top}\right)^{-1} &= {S_{11}}^{-1} - N\left({S_{11}}^{-1}\bar{X}_{[1]}\right)\left(1 + N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}\right)^{-1}\left({S_{11}}^{-1}\bar{X}_{[1]}\right)^{\top}, \\ &= {S_{11}}^{-1} - \frac{N\left({S_{11}}^{-1}\bar{X}_{[1]}\right)\left({S_{11}}^{-1}\bar{X}_{[1]}\right)^{\top}}{1 + N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}} \end{aligned}$$

And thus

$$\begin{aligned} N\bar{X}_{[1]}^{\top}\left(S_{11} + N\bar{X}_{[1]}\bar{X}_{[1]}^{\top}\right)^{-1}\bar{X}_{[1]} &= N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]} - \frac{\left(N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}\right)^2}{1 + N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}}, \\ &= \frac{N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}}{1 + N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}}, \\ 1 - N\bar{X}_{[1]}^{\top}\left(S_{11} + N\bar{X}_{[1]}\bar{X}_{[1]}^{\top}\right)^{-1}\bar{X}_{[1]} &= \frac{1}{1 + N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}}. \end{aligned}$$

Thus the Z statistic can be more simply defined as

$$Z = \frac{1 + N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}}{1 + N\bar{X}_{[2]}^{\top}{S_{22}}^{-1}\bar{X}_{[2]}}. \tag{34}$$

In section 3, Giri notes that, conditional on observing R_1, Z takes a (noncentral) beta distribution with $\frac{1}{2}(N-p)$ and $\frac{1}{2}(p-q)$ degrees of freedom and non-centrality parameter $\delta_2(1-R_1)$. From inspection, it is a 'type II' noncentral beta, which can be transformed into a noncentral F:

$$\frac{N-p}{p-q}\frac{1-Z}{Z} = \frac{N-p}{p-q}\frac{N\bar{X}_{[2]}^{\top}{S_{22}}^{-1}\bar{X}_{[2]} - N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}}{1+N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}}. \tag{35}$$

Giri defines R_1 in equation (2.2). It is equivalent to

$$1-R_1 = \frac{1}{1+N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}}.$$

Giri defines δ_1, δ_2 in equation (2.3). We have

$$\delta_2 = N\xi^{\top}\Sigma^{-1}\xi - N{\xi_{[1]}}^{\top}{\Sigma_{11}}^{-1}\xi_{[1]}.$$

Taking this all together, we have, conditional on observing $\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}$,

$$\frac{N-p}{p-q}\frac{N\bar{X}_{[2]}^{\top}{S_{22}}^{-1}\bar{X}_{[2]} - N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}}{1+N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}} \sim F\left(p-q, N-p, \frac{N\left(\xi^{\top}\Sigma^{-1}\xi - {\xi_{[1]}}^{\top}{\Sigma_{11}}^{-1}\xi_{[1]}\right)}{1+N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}}\right). \tag{36}$$

Now note that S_{11} refers to the sample Gram matrix, and thus S_{11}/N is the biased covariance estimate, $\tilde{\hat{\Sigma}}$ on the subset of q assets, while $\bar{X}_{[1]}$ is the mean of the subset of q assets. Giri's terminology translates into the terminology of spanning tests used in Section 2.4 as follows:

$$\begin{aligned}
N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]} &= \frac{n}{n-1}\hat{\zeta}_{*,\mathsf{G}}^{2},\\
N\bar{X}_{[2]}^{\top}{S_{22}}^{-1}\bar{X}_{[2]} &= \frac{n}{n-1}\hat{\zeta}_{*,\mathsf{I}}^{2},\\
{\xi_{[1]}}^{\top}{\Sigma_{11}}^{-1}\xi_{[1]} &= \zeta_{*,\mathsf{G}}^{2},\\
\xi^{\top}\Sigma^{-1}\xi &= \zeta_{*,\mathsf{I}}^{2},\\
N &= n,\\
p-q &= k-k_g.
\end{aligned}$$

Thus, conditional on observing $\hat{\zeta}_{*,\mathsf{G}}^{2}$, we have

$$\frac{n-k}{k-k_g}\frac{\hat{\zeta}_{*,\mathsf{I}}^{2} - \hat{\zeta}_{*,\mathsf{G}}^{2}}{(n-1)/n + \hat{\zeta}_{*,\mathsf{G}}^{2}} \sim F\left(k-k_g, n-k, \frac{n}{1+\frac{n}{n-1}\hat{\zeta}_{*,\mathsf{G}}^{2}}\left(\zeta_{*,\mathsf{I}}^{2} - \zeta_{*,\mathsf{G}}^{2}\right)\right). \tag{37}$$