Notes on the Sharpe ratio

Steven E. Pav *

March 13, 2016

Abstract

Herein is a hodgepodge 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. A 'unified' framework is described which combines the mean and covariance parameters of a multivariate distribution into the uncentered second moment of a related random variable. This trick streamlines some multivariate computations, and gives the asymptotic distribution of the sample Markowitz portfolio.

Contents

*`shabbychef@gmail.com` This document is a bit of a mess at the moment, and is in a constant state of flux.

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. [61] 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.” [61]

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 t-test we now use, defined as

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

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![61]

is *not* the form first considered by Gosset (writing as "Student").[22] 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. [46]

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 $\hat{\mu} =_{\mathrm{df}} \sum_i x_i/n$ and $\hat{\sigma}^2 =_{\mathrm{df}} \sum_i (x_i - \hat{\mu})^2/(n-1)$ be the unbiased sample mean and variance, and let

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

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

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

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. [26, 58]

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'), $\zeta =_{\mathrm{df}} \mu/\sigma$, 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 $t_{1-\alpha}(n-1)$, 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

$$1 - F_t\left(t_{1-\alpha}\left(n-1\right); n-1, \delta_1\right),$$

where $\delta_1 = \sqrt{n}\left(\mu_1 - \mu_0\right)/\sigma$ and $F_t\left(x; n-1, \delta\right)$ is the cumulative distribution function of the non-central t-distribution with non-centrality parameter δ and $n-1$ degrees of freedom. [58]

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 $t_{1-\alpha}\left(n-1, \delta_0\right)$, 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

$$1 - F_t\left(t_{1-\alpha}\left(n-1, \delta_0\right); n-1, \delta_1\right),$$

where $\delta_1 = \sqrt{n}\zeta_1$ and $F_t\left(x; n-1, \delta\right)$ is the cumulative distribution function of the non-central t-distribution with non-centrality parameter δ and $n-1$ degrees of freedom. [58]

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. [67, 69] The t-statistic, which follows a non-central t-distribution with parameter δ and $n-1$ degrees of freedom has the following moments:

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

Here $d_n = \sqrt{\frac{n-1}{2}}\Gamma\left((n-2)/2\right)/\Gamma\left((n-1)/2\right)$, is the 'bias term'. The geometric bias term is related to the constant c_4 from the statistical control literature via $d_n = \frac{n-1}{n-2}c_4\left(n\right)$. These can be trivially translated into equivalent facts regarding the Sharpe ratio:

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

The geometric bias term d_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. [46] 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: $d_n = 1.08$, *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. [46, 25]

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

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

1.4 Asymptotics and confidence intervals

Lo showed that the Sharpe ratio is asymptotically normal in n with standard deviation [37]

$$se \approx \sqrt{\frac{1 + \frac{\zeta^2}{2}}{n}}. \tag{5}$$

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. [26] 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

$$\hat{\zeta} \pm z_{\alpha/2}\sqrt{\frac{1 + \frac{\hat{\zeta}^2}{2}}{n-1}} \tag{6}$$

has better small sample coverage 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. [67]

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

$$\hat{\zeta}\left(1 - \frac{1}{4(n-1)}\right) \pm z_{\alpha/2}\sqrt{\frac{1}{n} + \frac{\hat{\zeta}^2}{2(n-1)}}. \tag{7}$$

The normality results generally hold for large n, small ζ, and assume normality of x. [26] 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 $[\zeta_l, \zeta_u]$ defined implicitly by

$$1 - \alpha/2 = F_t\left(\hat{\zeta}; n-1, \sqrt{n}\zeta_l\right), \quad \alpha/2 = F_t\left(\hat{\zeta}; n-1, \sqrt{n}\zeta_u\right), \tag{8}$$

where $F_t(x; n-1, \delta)$ 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 [9]), which is slower than approximations based on asymptotic normality.

Mertens gives the form of standard error

$$se \approx \sqrt{\frac{1 + \frac{2+\gamma_4}{4}\zeta^2 - \gamma_3 \zeta}{n}}, \tag{9}$$

where γ_3 is the skew, and γ_4 is the *excess* kurtosis of the returns distribution. [45, 50, 2] These are both zero for normally distributed returns, and so Mertens' form reduces to Lo's form. These are unknown in practice, and have to be estimated from the data, which results in some mis-estimation of the standard error when skew is extreme.

1.5 Asymptotic Distribution of Sharpe ratio

Here I derive the asymptotic distribution of Sharpe ratio, following Jobson and Korkie *inter alia.* [25, 37, 45, 33, 35, 71] Consider the case of p possibly correlated returns streams, with each observation denoted by $\boldsymbol{x}$. Let $\boldsymbol{\mu}$ be the p-vector of population means, and let $\boldsymbol{\kappa}_2$ be the p-vector of the uncentered second moments. Let $\boldsymbol{\zeta}$ be the vector of SNR of the assets. Let r_0 be the 'risk free rate'. We have

$$\boldsymbol{\zeta}_i = \frac{\boldsymbol{\mu}_i - r_0}{\sqrt{\boldsymbol{\kappa}_{2,i} - \boldsymbol{\mu}_i^2}}.$$

Consider the $2p$ vector of $\boldsymbol{x}$, 'stacked' with $\boldsymbol{x}$ (elementwise) squared, $\left[\boldsymbol{x}^\top, {\boldsymbol{x}^2}^\top\right]^\top$. The expected value of this vector is $\left[\boldsymbol{\mu}^\top, {\boldsymbol{\kappa}_2}^\top\right]^\top$; let Ω be the variance of this vector, assuming it exists.

Given n observations of $\boldsymbol{x}$, consider the simple sample estimate

$$\left[\hat{\boldsymbol{\mu}}^\top, {\hat{\boldsymbol{\kappa}}_2}^\top\right]^\top =_{\mathrm{df}} \frac{1}{n}\sum_i^n \left[\boldsymbol{x}^\top, {\boldsymbol{x}^2}^\top\right]^\top.$$

Under the multivariate central limit theorem [68]

$$\sqrt{n}\left(\left[\hat{\boldsymbol{\mu}}^\top, {\hat{\boldsymbol{\kappa}}_2}^\top\right]^\top - \left[\boldsymbol{\mu}^\top, {\boldsymbol{\kappa}_2}^\top\right]^\top\right) \rightsquigarrow \mathcal{N}\left(0, \Omega\right). \tag{10}$$

Let $\hat{\boldsymbol{\zeta}}$ be the sample Sharpe ratio computed from the estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\kappa}}_2$: $\hat{\boldsymbol{\zeta}}_i = \left(\hat{\boldsymbol{\mu}}_i - r_0\right)/\sqrt{\hat{\boldsymbol{\kappa}}_{2,i} - \hat{\boldsymbol{\mu}}_i^2}$. By the multivariate delta method,

$$\sqrt{n}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}\right) \rightsquigarrow \mathcal{N}\left(0, \left(\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, {\boldsymbol{\kappa}_2}^\top]^\top}\right)\Omega\left(\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, {\boldsymbol{\kappa}_2}^\top]^\top}\right)^\top\right). \tag{11}$$

Here the derivative takes the form of two $p \times p$ diagonal matrices pasted together side by side:

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, {\boldsymbol{\kappa}_2}^\top]^\top} &= \left[\begin{array}{cc} \operatorname{diag}\left(\frac{\boldsymbol{\kappa}_2 - \boldsymbol{\mu} r_0}{(\boldsymbol{\kappa}_2 - \boldsymbol{\mu}^2)^{3/2}}\right) & \operatorname{diag}\left(\frac{r_0 - \boldsymbol{\mu}}{2(\boldsymbol{\kappa}_2 - \boldsymbol{\mu}^2)^{3/2}}\right) \end{array}\right], \\ &= \left[\begin{array}{cc} \operatorname{diag}\left(\frac{\boldsymbol{\sigma} + \boldsymbol{\mu}\boldsymbol{\zeta}}{\boldsymbol{\sigma}^2}\right) & \operatorname{diag}\left(\frac{-\boldsymbol{\zeta}}{2\boldsymbol{\sigma}^2}\right) \end{array}\right]. \end{aligned} \tag{12}$$

where $\mathrm{diag}\left(\boldsymbol{z}\right)$ is the matrix with vector $\boldsymbol{z}$ on its diagonal, and where the vector operations above are all performed elementwise.

In practice, the population values, $\boldsymbol{\mu}$, $\boldsymbol{\kappa}_2$, Ω are all unknown, and so the asymptotic variance has to be estimated, using the sample. Letting $\hat{\Omega}$ be some sample estimate of Ω, we have the approximation

$$\hat{\boldsymbol{\zeta}} \approx \mathcal{N}\left(\boldsymbol{\zeta}, \frac{1}{n}\left(\frac{\mathrm{d}\hat{\boldsymbol{\zeta}}}{\mathrm{d}\left[\hat{\boldsymbol{\mu}}^{\top}, \hat{\boldsymbol{\kappa}}_2{}^{\top}\right]^{\top}}\right) \hat{\Omega} \left(\frac{\mathrm{d}\hat{\boldsymbol{\zeta}}}{\mathrm{d}\left[\hat{\boldsymbol{\mu}}^{\top}, \hat{\boldsymbol{\kappa}}_2{}^{\top}\right]^{\top}}\right)^{\top}\right), \tag{13}$$

where the derivatives are formed by plugging in the sample estimates into Equation 12. [37, 45]

1.5.1 Scalar case

For the $p = 1$ case, Ω takes the form

$$\begin{aligned}
\Omega &= \left[\begin{array}{cc} \kappa_2 - \mu^2 & \kappa_3 - \mu\kappa_2 \\ \kappa_3 - \mu\kappa_2 & \kappa_4 - {\kappa_2}^2 \end{array}\right], \\
&= \left[\begin{array}{cc} \sigma^2 & \sigma^2\left(\sigma\gamma_3 + 2\mu\right) \\ \sigma^2\left(\sigma\gamma_3 + 2\mu\right) & \sigma^4\left(\gamma_4 + 2\right) + 4\sigma^3\mu\gamma_3 + 4\sigma^2\mu^2 \end{array}\right], \\
&= \sigma^2 \left[\begin{array}{cc} 1 & \sigma\gamma_3 + 2\mu \\ \sigma\gamma_3 + 2\mu & \sigma^2\left(\gamma_4 + 2\right) + 4\sigma\mu\gamma_3 + 4\mu^2 \end{array}\right].
\end{aligned} \tag{14}$$

where κ_i is the uncentered i^{th} moment of x, γ_3 is the skew, and γ_4 is the excess kurtosis. After much algebraic simplification, the asymptotic variance of Sharpe ratio is given by Mertens' formula, Equation 9:

$$\hat{\zeta} \approx \mathcal{N}\left(\zeta, \frac{1}{n}\left(1 - \hat{\zeta}\gamma_3 + \frac{\gamma_4 + 2}{4}\hat{\zeta}^2\right)\right). \tag{15}$$

Note that Mertens' equation applies even though our definition of Sharpe ratio includes a risk-free rate, r_0.

1.5.2 Tests of equality of multiple Sharpe ratio

Now let $\boldsymbol{g}$ be some vector valued function of the vector $\boldsymbol{\zeta}$. Applying the delta method,

$$\sqrt{n}\left(\boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right) - \boldsymbol{g}\left(\boldsymbol{\zeta}\right)\right) \rightsquigarrow \mathcal{N}\left(0, \left(\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}}\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^{\top}, \boldsymbol{\kappa}_2{}^{\top}]^{\top}}\right) \Omega \left(\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}}\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^{\top}, \boldsymbol{\kappa}_2{}^{\top}]^{\top}}\right)^{\top}\right) \tag{16}$$

To compare whether the Sharpe ratio of p assets are equal, given n contemporaneous observations of their returns, let $\boldsymbol{g}$ be the function which constructs the $p - 1$ contrasts:

$$\boldsymbol{g}\left(\boldsymbol{\zeta}\right) = \left[\boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_{p-1} - \boldsymbol{\zeta}_p\right]^{\top}.$$

One is then testing the null hypothesis $H_0 : \boldsymbol{g}(\boldsymbol{\zeta}) = 0$. Asymptotically, under the null,

$$n\boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right)^\top \left(\left(\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}} \frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, {\boldsymbol{\kappa}_2}^\top]^\top} \right) \Omega \left(\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}} \frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, {\boldsymbol{\kappa}_2}^\top]^\top} \right)^\top \right)^{-1} \boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right) \sim \chi^2\left(p-1\right).$$

For the more general case, where $\boldsymbol{g}$ need not be the vanilla contrasts, the chi-square degrees of freedom is the rank of $\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}}$.

There are a number of modifications of this basic method: Leung and Wong described the basic method. [35] Wright *et al.* suggest that the test statistic be transformed to an approximate F-statistic. [71] Ledoit and Wolf propose using HAC estimators or bootstrapping to construct $\hat{\Omega}$. [33]

For the case of scalar-valued g (*e.g.*, for comparing $p = 2$ strategies), one can construct a two-sided test via an asymptotic t-approximation:

$$\sqrt{n}g\left(\hat{\boldsymbol{\zeta}}\right) \left(\left(\frac{\mathrm{d}g}{\mathrm{d}\boldsymbol{\zeta}} \frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, {\boldsymbol{\kappa}_2}^\top]^\top} \right) \Omega \left(\frac{\mathrm{d}g}{\mathrm{d}\boldsymbol{\zeta}} \frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, {\boldsymbol{\kappa}_2}^\top]^\top} \right)^\top \right)^{-\frac{1}{2}} \sim t\left(n-1\right).$$

In all the above, one can construct asymptotic approximations of the test statistics under the alternative, allowing power analysis or computation of confidence regions on $\boldsymbol{g}(\boldsymbol{\zeta})$.

1.6 Power and sample size

Consider the t-test for the null hypothesis $H_0 : \mu = 0$. This is equivalent to testing $H_0 : \zeta = 0$. A power rule ties together the (unknown) true effect size (ζ), sample size (n), type I and type II rates. Some example use cases:

1. Suppose you wanted to estimate the mean return of a pairs trade, but the stocks have only existed for two years. Is this enough data assuming the SNR is 2.0 $\mathrm{yr}^{-1/2}$?
2. Suppose investors in a fund you manage want to 'see some returns' within a year otherwise they will withdraw their investment. What SNR should you be hunting for so that, with probability one half, the actual returns will 'look good' over the next year?
3. Suppose you observe three months of a fund's returns, and you fail to reject the null under the one sample t-test. Assuming the SNR of the process is 1.5 $\mathrm{yr}^{-1/2}$, what is the probability of a type II error?

For sufficiently large sample size (say $n \geq 30$), the power law for the t-test is well approximated by

$$n \approx \frac{c}{\zeta^2}, \tag{17}$$

where the constant c depends on the type I rate and the type II rates, and whether one is performing a one- or two-sided test. This relationship was first noted by Johnson and Welch. [26] Unlike the type I rate, which is traditionally set at 0.05, there is no widely accepted traditional value of power.

Values of the coefficient c are given for one and two-sided t-tests at different power levels in Table 1. The case of $\alpha = 0.05, 1 - \beta = 0.80$ is known as "Lehr's rule". [66, 34]

	one.sided	two.sided
power = 0.25	0.96	1.68
power = 0.50	2.72	3.86
power = 0.80	6.20	7.87

Table 1: Scaling of sample size with respect to ζ^2 required to achieve a fixed power in the t-test, at a fixed $\alpha = 0.05$ rate.

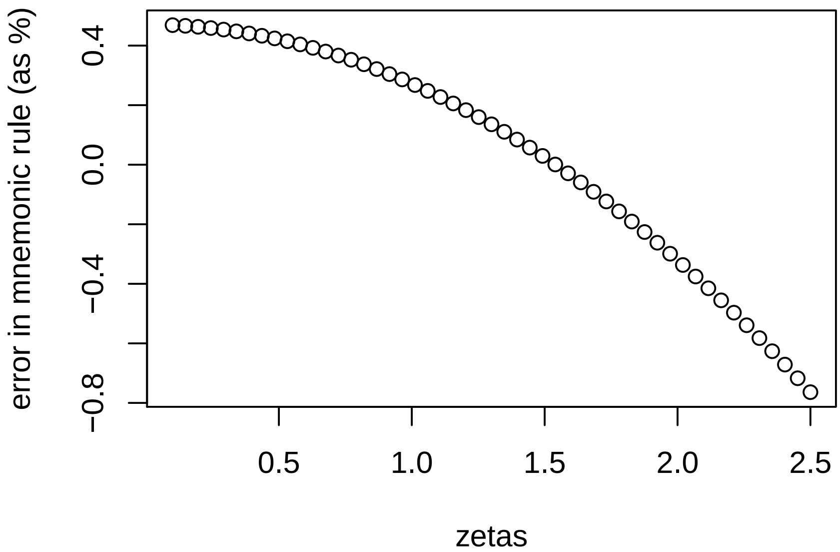

Figure 1: The percent error of the power mnemonic $e \approx n\zeta^2$ is plotted versus ζ.

Consider now the scaling in the rule $n \approx c\zeta^{-2}$. If the SNR ζ is given in daily units, the sample size will be in days. One annualizes ζ by multiplying by the square root of the number of days per year, which downscales n appropriately. That is, if ζ is quoted in annualized terms, this rule of thumb gives the number of *years* of observations required. This is very convenient since we usually think of ζ and $\hat{\zeta}$ in annualized terms.

The following rule of thumb may prove useful:

> The number of years required to reject non-zero mean with power of one half is around $2.7/\zeta^2$.

The mnemonic form of this is "$e = nz^2$". Note that Euler's number appears here coincidentally, as it is nearly equal to $\left[\Phi^{-1}(0.95)\right]^2$. The relative error in this approximation for determining the sample size is shown in Figure 1, as a function of ζ; the error is smaller than one percent in the tested range.

The power rules are sobering indeed. Suppose you were a hedge fund manager whose investors threatened to perform a one-sided t-test after one year. If your strategy's signal-to-noise ratio is less than $1.65\mathrm{yr}^{-1/2}$ (a value which should be considered "very good"), your chances of 'passing' the t-test are less than fifty percent.

1.7 Deviations from assumptions

Van Belle suggests one consider, in priority order, assumptions of independence, heteroskedasticity, and normality in statistical tests. [66]

1.7.1 Sharpe ratio and Autocorrelation

The simplest relaxation of the *i.i.d.* assumption of the returns x_i is to assume the time-series of returns has a fixed autocorrelation. Let ν be the autocorrelation of the series of returns, *i.e.*, the population correlation of x_{i-1} with x_i. [11] In this case the standard error of the mean tends to be an underestimate when $\nu > 0$ and an overestimate when $\nu < 0$. Van Belle [66] notes that, under this formulation, the t statistic (under the null $\mu = 0$) has standard error of approximately $\sqrt{(1+\nu)/(1-\nu)}$. A Monte Carlo study confirms this approximation. In Figure 2 the empirical standard deviation of t-statistics computed on AR(1) series at given values of ν along with the fit value of $\sqrt{(1+\nu)/(1-\nu)}$.

The 'small angle' approximation for this correction is $1+\nu$, which is reasonably accurate for $|\nu| < 0.1$. An alternative expression of this approximation is "a positive autocorrelation of ν inflates the Sharpe ratio by about ν percent."

The corrected t-statistic has the form:

$$t_0{}' = \sqrt{\frac{1-\hat{\nu}}{1+\hat{\nu}}}\sqrt{n}\frac{\hat{\mu}-\mu_0}{\hat{\sigma}} = d\sqrt{n}\hat{\zeta}_0, \tag{18}$$

where d is the correction factor for autocorrelation [66]. The equivalent correction for Sharpe ratio is $\hat{\zeta}_0' = d\hat{\zeta}_0$.

1.7.2 Sharpe ratio and Heteroskedasticity

The term 'heteroskedasticity' typically applies to situations where one is performing inference on the mean effect, and the magnitude of error varies in the sample. This idea does not naturally translate to performing inference on the SNR, since SNR incorporates volatility, and would vary under the traditional definition. Depending on the asset, the SNR might increase or decrease with volatility, an effect further complicated by the risk-free rate, which is assumed to be constant.

Here I will consider the case of asset returns with constant SNR, and fluctuating volatility. That is, both the volatility and expected return are changing over time, with their ratio constant. One can imagine this as some 'latent' return stream which one observes polluted with a varying 'leverage'. So suppose that l_i and x_i are independent random variables with $l_i > 0$. One observes period returns of $l_i x_i$ on period i. We have assumed that the SNR of x is a constant which we are trying to estimate. We have

$$\begin{aligned}
\mathrm{E}\left[lx\right] &= \mathrm{E}\left[l\right]\mathrm{E}\left[x\right],\\
\mathrm{Var}\left(lx\right) &= \mathrm{E}\left[l^2\right]\mathrm{E}\left[x^2\right] - \mathrm{E}\left[l\right]^2\mathrm{E}\left[x\right]^2 = \mathrm{E}\left[x^2\right]\mathrm{Var}\left(l\right) + \mathrm{Var}\left(x\right)\mathrm{E}\left[l\right]^2,
\end{aligned} \tag{19}$$

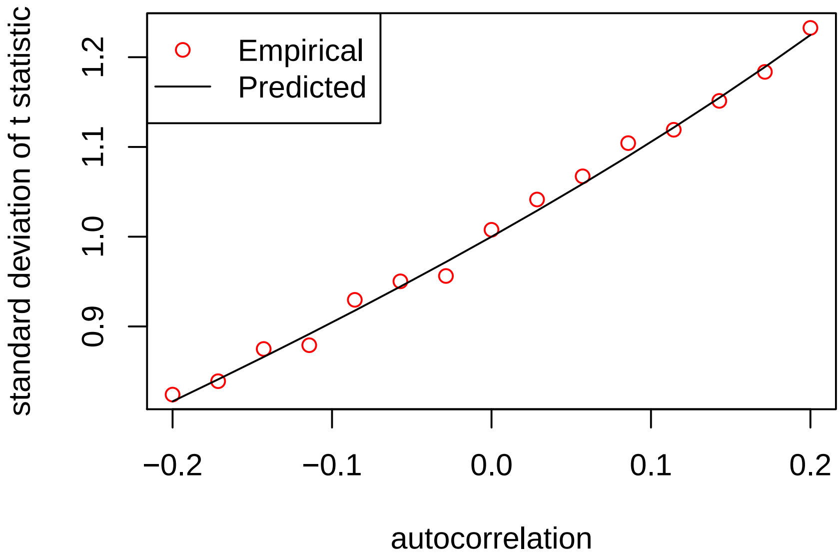

Figure 2: The empirical standard deviation for the t-statistic is shown at different values of the autocorrelation, ν. Each point represents 8000 series of approximately 3 years of daily data, with each series generated by an AR(1) process with normal innovations. Each series has actual SNR of zero. The fit line is that suggested by Van Belle's correction for autocorrelation, namely $\sqrt{(1+\nu)/(1-\nu)}$.

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 heteroskedasticity tends to give underestimates of the 'true' Sharpe ratio of the returns series, x. However, the deflation is typically modest, on the order of 10%. The shrinkage of Sharpe ratio will also typically lead to slight deflation of the estimated standard error, but for large n and daily returns, this will not lead to inflated type I rate.

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 $\mathrm{E}[x]^2 \approx 0.01\,\mathrm{Var}(x)$, and so $\mathrm{E}\left[x^2\right] \approx 1.01\,\mathrm{Var}(x)$. Thus it suffices to estimate the ratio $\mathrm{Var}(l) / \mathrm{E}[l]^2$, 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

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

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

1.7.3 Sharpe ratio and Non-normality

The distribution of the Sharpe ratio given in Section 1.1 is only valid when the returns of the fund are normally distributed. If not, the Central Limit Theorem guarantees that the sample mean is *asymptotically* normal (assuming the variance exists!), but the convergence to a normal can require a large sample size. In practice, the tests described in Section 1.2 work fairly well for returns from kurtotic distributions, but can be easily fooled by skewed returns.

There should be no real surprise in this statement. Suppose one is analyzing the returns of a hedge fund which is secretly writing insurance policies, and has had no claims in the past 5 years. The true expected mean return of the fund might be zero or negative, but the historical data does not contain a 'Black Swan' type event. [65] We need not make any fabulous conjectures about the 'non-stationarity' of our return series, or the failure of models or our ability to predict: skew is a property of the distribution, and we do not have enough evidence to detect the skew.

To demonstrate this fact, I look at the empirical type I rate for the hypothesis test: $H_0 : \zeta = 1.0$ versus the alternative $H_1 : \zeta > 1.0$ for different distributions of returns: I sample from a Gaussian (as the benchmark); a t-distribution with 10 degrees of freedom; a Tukey h-distribution, with different values of h; a 'lottery' process which is a shifted, rescaled Bernoulli random variable; and a 'Lambert W x Gaussian' distribution, with different values of the skew parameter. [20, 19, 21] I also draw samples from the daily log returns of the S & P 500 over the period from January 05, 1970 to December 31, 2012, affinely transformed to have $\zeta = 1.0\mathrm{yr}^{-1/2}$. I also draw from a symmetrized S & P 500 returns series.

The t- and Tukey distributions are fairly kurtotic, but have zero skew, while the lottery and Lambert W x Gaussian distributions are skewed and (therefore) kurtotic. All distributions have been rescaled to have $\zeta = 1.0\text{yr}^{-1/2}$; that is, I am estimating the empirical type I rate under the null. At the nominal $\alpha = 0.05$ level, we expect to get a reject rate around five percent.

I test the empirical type I rate of the test implied by the confidence intervals in Equation 8. I also employ Mertens' standard errors, Equation 9, estimating the skew and kurtosis empirically, then comparing to quantiles of the normal distribution. The tests are one-sided tests, against the alternative $H_a : \zeta < 1.0\text{yr}^{-1/2}$.

distribution	param	skew	kurtosis	typeI	cor.typeI
Gaussian		0	0	0.048	0.048
Student's t	df = 10	0	1	0.048	0.049
SP500		-1	26	0.057	0.058
symmetric SP500		0	25	0.057	0.06
Tukey h	h = 0.1	0	5.5	0.052	0.054
Tukey h	h = 0.2	0	1.3e+03	0.052	0.058
Tukey h	h = 0.4		Inf	0.14	0.17
Lottery	p = 0.020	6.9	45	0.0071	0.054
Lottery	p = 0.010	9.8	95	0.002	0.046
Lottery	p = 0.005	14	2e+02	0.00024	0.04
Lambert x Gaussian	delta = 0.4	2.7	18	0.028	0.053
Lambert x Gaussian	delta = 0.2	1.2	5.7	0.039	0.053
Lambert x Gaussian	delta = -0.2	-1.2	5.7	0.063	0.051
Lambert x Gaussian	delta = -0.4	-2.7	18	0.072	0.046
Lambert x Gaussian	delta = -0.8	-8.5	2.1e+02	0.14	0.073

Table 2: Empirical type I rates of the test for $\zeta = 1.0\text{yr}^{-1/2}$ via distribution of the Sharpe ratio are given for various distributions of returns. The empirical rates are based on 8192 simulations of three years of daily returns, with a nominal rate of $\alpha = 0.05$. The 'corrected' type I rates refer to a normal approximation using Mertens' correction. Skew appears to have a much more adverse effect than kurtosis alone.

The results are given in Table 2, and indicate that skew is a more serious problem than kurtosis. The results from the S & P 500 data are fairly encouraging: broad market returns are perhaps not as skewed as the Lambert W x Gaussian distributions that we investigate, and the type I rate is near nominal when using three years of data. Lumley *et al.* found similar results for the t-test when looking at sample sizes greater than 500. [38] Since the t-test statistic is equivalent to Sharpe ratio (up to scaling), this result carries over to the test for SNR.

The Mertens' correction appears to be less liberal for the highly skewed Lambert distributions, but perhaps more liberal for the Tukey and S & P 500 distributions.

However, skew is a serious problem when using the Sharpe ratio. A practicioner must be reasonably satisfied that the return stream under examination is not seriously skewed to use the Sharpe ratio. Moreover, one can *not* use historical data to detect skew, for the same reason that skew causes the distribution of Sharpe ratio to degrade.

1.8 Linear attribution models

The Sharpe ratio and t-test as described previously can be more generally described in terms of linear regression. Namely one models the returns of interest as $x_t = \beta_0 1 + \epsilon_t$, where ϵ_t are modeled as *i.i.d.* zero-mean innovations with standard deviation σ. Performing a linear regression, one gets the estimates $\hat{\beta}_0$ and $\hat{\sigma}$, and can test the null hypothesis $H_0 : \beta_0/\sigma = 0$ via a t-test. To see that this is the case, one only need recognize that the sample mean is indeed the least-squares estimator, *i.e.*, $\hat{\mu} = \operatorname{argmin}_a \sum_t (a - x_t)^2$.

More generally, we might want to model returns as the linear combination of l factor returns:

$$x_t = \beta_0 1 + \sum_i^{l-1} \beta_i f_{i,t} + \epsilon_t, \tag{20}$$

where $f_{i,t}$ are the returns of some i^{th} 'factor' at time t. There are numerous candidates for the factors, and their choice should depend on the return series being modeled. For example, one would choose different factors when modeling the returns of a single company versus those of a broad-market mutual fund versus those of a market-neutral hedge fund, *etc.*. Moreover, the choice of factors might depend on the type of analysis being performed. For example, one might be trying to 'explain away' the returns of one investment as the returns of another investment (presumably one with smaller fees) plus noise. Alternatively, one might be trying to establish that a given investment has idiosyncratic 'alpha' (*i.e.*, β_0) without significant exposure to other factors.

1.8.1 Examples of linear attribution models

- As noted above, the Sharpe ratio implies a trivial factor model, namely $x_t = \beta_0 1 + \epsilon_t$. This simple model is generally a poor one for describing stock returns; one is more likely to see it applied to the returns of mutual funds, hedge funds, *etc.*
- The simple model does not take into account the influence of 'the market' on the returns of stocks. This suggests a factor model equivalent to the Capital Asset Pricing Model (CAPM), namely $x_t = \beta_0 1 + \beta_M f_{M,t} + \epsilon_t$, where $f_{M,t}$ is the return of 'the market' at time t. [6] (Note that the term 'CAPM' usually encompasses a number of assumptions used to justify the validity of this model for stock returns.)
  This is clearly a superior model for stocks and portfolios with a long bias (*e.g.*, typical mutual funds), but might seem inappropriate for a long-short balanced hedge fund, say. In this case, however, the loss of power in including a market term is typically very small, while the possibility of reducing type I errors is quite valuable. For example, one might discover that a seemingly long-short balanced fund actually has some market exposure, but no significant idiosyncratic returns (one cannot reject $H_0 : \beta_0 = 0$, say); this is valuable information, since a hedge-fund investor might balk at paying high fees for a return stream that replicates a (much less expensive) ETF plus noise.
- Generalizations of the CAPM factor model abound. For example, the Fama-French 3-factor model (I drop the risk-free rate for simplicity):

$$x_t = \beta_0 1 + \beta_{\mathrm{M}} f_{\mathrm{M},t} + \beta_{\mathrm{SMB}} f_{\mathrm{SMB},t} + \beta_{\mathrm{HML}} f_{\mathrm{HML},t} + \epsilon_t,$$

where $f_{\mathrm{M},t}$ is the return of 'the market', $f_{\mathrm{SMB},t}$ is the return of 'small minus big cap' stocks (the difference in returns of these two groups), and $f_{\mathrm{HML},t}$ is the return of 'high minus low book value' stocks. [16] Carhart adds a momentum factor:

$$x_t = \beta_0 1 + \beta_{\mathrm{M}} f_{\mathrm{M},t} + \beta_{\mathrm{SMB}} f_{\mathrm{SMB},t} + \beta_{\mathrm{HML}} f_{\mathrm{HML},t} + \beta_{\mathrm{UMD}} f_{\mathrm{UMD},t} + \epsilon_t,$$

where $f_{\mathrm{UMD},t}$ is the return of 'ups minus downs', *i.e.*, the returns of the previous period winners minus the returns of previous period losers. [12]

- Henriksson and Merton describe a technique for detecting market-timing ability in a portfolio. One can cast this model as

$$x_t = \beta_0 1 + \beta_{\mathrm{M}} f_{\mathrm{M},t} + \beta_{\mathrm{HM}} \left(-f_{\mathrm{M},t}\right)^{+} + \epsilon_t,$$

where $f_{\mathrm{M},t}$ are the returns of 'the market' the portfolio is putatively timing, and x^{+} is the positive part of x. [23] Actually, one or several factor timing terms can be added to any factor model. Note that unlike the factor returns in models discussed above, one expects $\left(-f_{\mathrm{M}}\right)^{+}$ to have significantly non-zero mean. This will cause some decrease in power when testing β_0 for significance. Also note that while Henriksson and Merton intend this model as a positive test for β_{HM}, one could treat the timing component as a factor which one seeks to ignore entirely, or downweight its importance.

- Often the linear factor model is used with a 'benchmark' (mutual fund, index, ETF, *etc.*) used as the factor returns. In this case, the process generating x_t may or may not be posited to have zero exposure to the benchmark, but usually one is testing for significant idiosyncratic term.
- Any of the above models can be augmented by splitting the idiosyncratic term into a constant term and some time-based term. For example, it is often argued that a certain strategy 'worked in the past' but does no longer. This implies a splitting of the constant term as

$$x_t = \beta_0 1 + {\beta_0}' f_{0,t} + \sum_i \beta_i f_{i,t} + \epsilon_t,$$

where $f_{0,t} = (n-t)/n$, given n observations. In this case the idiosyncratic part is an affine function of time, and one can test for β_0 independently of the time-based trend (one can also test whether ${\beta_0}' > 0$ to see if the 'alpha' is truly decaying). One can also imagine time-based factors which attempt to address seasonality or 'regimes'.

1.8.2 Tests involving the linear attribution model

Given n observations of the returns and the factors, let $\boldsymbol{x}$ be the vector of returns and let F be the $n \times l$ matrix consisting of the returns of the l factors and a column of all ones. The ordinary least squares estimator for the regression coefficients is expressed by the 'normal equations':

$$\hat{\boldsymbol{\beta}} = \left(\mathsf{F}^{\top}\mathsf{F}\right)^{-1}\mathsf{F}^{\top}\boldsymbol{x}.$$

The estimated variance of the error term is $\hat{\sigma}^2 = \left(\boldsymbol{x} - \mathsf{F}\hat{\boldsymbol{\beta}}\right)^{\top}\left(\boldsymbol{x} - \mathsf{F}\hat{\boldsymbol{\beta}}\right)/(n-l)$.

1. The classical t-test for regression coefficients tests the null hypothesis:

$$H_0 : \beta^{\top}\boldsymbol{v} = c \quad \text{versus} \quad H_1 : \beta^{\top}\boldsymbol{v} > c,$$

for some conformable vector $\boldsymbol{v}$ and constant c. To perform this test, we construct the regression t-statistic

$$t = \frac{\hat{\beta}^{\top}\boldsymbol{v} - c}{\hat{\sigma}\sqrt{\boldsymbol{v}^{\top}\left(\mathsf{F}^{\top}\mathsf{F}\right)^{-1}\boldsymbol{v}}}. \tag{21}$$

This statistic should be distributed as a non-central t-distribution with non-centrality parameter

$$\delta = \frac{\beta^{\top}\boldsymbol{v} - c}{\sigma\sqrt{\boldsymbol{v}^{\top}\left(\mathsf{F}^{\top}\mathsf{F}\right)^{-1}\boldsymbol{v}}},$$

and $n - l$ degrees of freedom. Thus we reject the null if t is greater than $t_{1-\alpha}\left(n - l\right)$, the $1 - \alpha$ quantile of the (central) t-distribution with $n - l$ degrees of freedom.

2. To test the null hypothesis:

$$H_0 : \beta^{\top}\boldsymbol{v} = \sigma c \quad \text{versus} \quad H_1 : \beta^{\top}\boldsymbol{v} > \sigma c,$$

for given $\boldsymbol{v}$ and c, one constructs the t-statistic

$$t = \frac{\hat{\beta}^{\top}\boldsymbol{v}}{\hat{\sigma}\sqrt{\boldsymbol{v}^{\top}\left(\mathsf{F}^{\top}\mathsf{F}\right)^{-1}\boldsymbol{v}}}. \tag{22}$$

This statistic should be distributed as a non-central t-distribution with non-centrality parameter

$$\delta = \frac{c}{\sqrt{\boldsymbol{v}^{\top}\left(\mathsf{F}^{\top}\mathsf{F}\right)^{-1}\boldsymbol{v}}},$$

and $n - l$ degrees of freedom. Thus we reject the null if t is greater than $t_{1-\alpha}\left(n - l, \delta\right)$, the $1 - \alpha$ quantile of the non-central t-distribution with $n - l$ degrees of freedom and non-centrality parameter δ.
Note that the statistic $\hat{\beta}_0/\hat{\sigma}$ is the equivalent to the Sharpe ratio in the general factor model (and β_0/σ is the population analogue).
2FIX: 2 sample test for SNR of independent groups?

1.8.3 Deviations from the model

The advantage of viewing Sharpe ratio as a least squares regression problem (or of using the more general factor model for attribution), is that regression is a well-studied problem. Indeed, numerous books and articles have been written about the topic and how to test for, and deal with, deviations from the model: autocorrelation, heteroskedasticity, non-normality, outliers, *etc.* [56, 31, 5, 28]

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 $\boldsymbol{\Sigma}$. Let $\hat{\boldsymbol{\mu}}$ be the usual sample estimate of the mean, $\hat{\boldsymbol{\mu}} = \sum_i \boldsymbol{x}_i/n$, and let $\hat{\boldsymbol{\Sigma}}$ 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

$$\max_{\hat{\boldsymbol{\nu}}:\hat{\boldsymbol{\nu}}^{\top}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\nu}} \le R^2} \frac{\hat{\boldsymbol{\nu}}^{\top}\hat{\boldsymbol{\mu}} - r_0}{\sqrt{\hat{\boldsymbol{\nu}}^{\top}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\nu}}}}, \tag{23}$$

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

This problem has solution

$$\hat{\boldsymbol{\nu}}_* =_{\mathrm{df}} c\, \hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}, \tag{24}$$

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

$$\hat{\zeta}_* =_{\mathrm{df}} \frac{\hat{\boldsymbol{\nu}}_*{}^{\top}\hat{\boldsymbol{\mu}} - r_0}{\sqrt{\hat{\boldsymbol{\nu}}_*{}^{\top}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\nu}}_*}} = \sqrt{\hat{\boldsymbol{\mu}}^{\top}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}} - \frac{r_0}{R}. \tag{25}$$

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

$$\zeta_* =_{\mathrm{df}} \sqrt{\boldsymbol{\mu}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}} - \frac{r_0}{R}. \tag{26}$$

The random term, $n\left(\hat{\boldsymbol{\mu}}^{\top}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}\right)^2$, 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 $F\left(v_1, v_2, \delta\right)$ is the non-central F-distribution with v_1, v_2 degrees of freedom and non-centrality parameter δ. This allows us to make inference about ζ_*.

By using the 'biased' covariance estimate, defined as

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

the above expression can be simplified slightly as

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

2.1 Tests involving Hotelling's Statistic

Here I list the classical multivariate analogues to the tests described in Section 1.2:

1. The classical one-sample test for mean of a multivariate random variable uses Hotelling's statistic, just as the univariate test uses the t-statistic. Unlike the univariate case, we cannot perform a one-sided test (because $p > 1$ makes one-sidedness an odd concept), and thus we have the two-sided test:
$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0,$$
we reject at the α level if
$$T_0^2 = n(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)^\top \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0) \geq \frac{p(n-1)}{n-p} f_{1-\alpha}(p, n-p),$$
where $f_{1-\alpha}(p, n-p)$ is the $1-\alpha$ quantile of the (central) F-distribution with p and $n-p$ degrees of freedom.
If $\boldsymbol{\mu} = \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_0$, then the power of this test is
$$1 - \beta = 1 - F_f(f_{1-\alpha}(p, n-p); p, n-p, \delta_1),$$
where
$$\delta_1 = n(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$$
is the noncentrality parameter, and $F_f(x; p, n-p, \delta)$ is the cumulative distribution function of the non-central F-distribution with non-centrality parameter δ and $p, n-p$ degrees of freedom. [4]
Note that the non-centrality parameter is equal to the population analogue of the Hotelling statistic itself. One should take care that some references (and perhaps statistical packages) have different ideas about how the non-centrality parameter should be communicated. The above formulation matches the convention used in the R statistical package and in Matlab's statistics toolbox. It is, however, off by a factor of two with respect to the convention used by Bilodeau and Brenner. [4]
2. A one-sample test for optimal signal-to-noise ratio (SNR) involves the Hotelling statistic as follows. To test
$$H_0 : \zeta_* = \zeta_0 \quad \text{versus} \quad H_1 : \zeta_* > \zeta_0,$$
we reject if
$$T_0^2 > \frac{p(n-1)}{n-p} f_{1-\alpha}(p, n-p, \delta_0),$$
where T_0^2 and δ_0 are defined as above, and where $f_{1-\alpha}(p, n-p, \delta_0)$ is the $1-\alpha$ quantile of the non-central F-distribution with non-centrality parameter δ_0 and p and $n-p$ degrees of freedom.
If $\zeta_* > \zeta_0$, then the power of this test is
$$1 - \beta = 1 - F_f(f_{1-\alpha}(p, n-p, \delta_0); p, n-p, \delta_*),$$
where $\delta_* = n\zeta_*^2$, is the noncentrality parameter, and $F_f(x; p, n-p, \delta)$ is the the cumulative distribution function of the non-central F-distribution with non-centrality parameter δ and $p, n-p$ degrees of freedom.

2.1.1 Power and Sample Size

In Section 1.6 I outlined the relationship of sample size and effect size for the one-sample t-test, or equivalently, the one-sample test for SNR. Here I extend those results to the Hotelling test for zero optimal population SNR, *i.e.*, the null $\zeta_0 = 0$. As noted in Section 2.1, the power of this test is $1 - \beta = 1 - F_f\left(f_{1-\alpha}\left(p, n-p, 0\right); p, n-p, \delta_*\right).$

This equation implicitly defines a sample size, n given α, β, p and δ_*. As it happens, for fixed values of α, β and p, the sample size relationship is similar to that found for the t-test:

$$n \approx \frac{c}{\zeta_*^2},$$

where the constant c depends on α, β and p. For $\alpha = 0.05$, an approximate value of the numerator c is given in Table 3 for a few different values of the power. Note that for $p = 1$, we should recover the same sample-size relationship as shown in Table 1 for the two-sided test. This is simply because Hotelling's statistic for $p = 1$ is Student's t-statistic squared (and thus side information is lost).

	numerator
power = 0.25	$1.66p^{0.438+0.006\log p}$
power = 0.50	$3.86p^{0.351+0.012\log p}$
power = 0.80	$7.87p^{0.277+0.017\log p}$

Table 3: The numerator in the sample size relationship required to achieve a fixed power in Hotelling's test is shown. The type I rate is 0.05.

2.2 Asymptotics and Confidence Intervals

As noted in Section C.1, 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:

$$\mathrm{E}\left[\sqrt{F}\right] \approx \sqrt{\mathrm{E}\left[F\right]} - \frac{\frac{{v_2}^2\left(\delta^2+(v_1+2)\left(2\,\delta+v_1\right)\right)}{{v_1}^2\left(v_2-4\right)\left(v_2-2\right)} - \left(\mathrm{E}\left[F\right]\right)^2}{8\left(\mathrm{E}\left[F\right]\right)^{\frac{3}{2}}}, \tag{27}$$

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

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

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

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 $p = c_a n$, and taking the limit as $n \to \infty$, we have

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

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

The asymptotic variance appears to be

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

Consider as an example, the case where $p = 30$, $n = 1000\,\text{days}$, and $\zeta_* = 1.5\,\text{yr}^{-1/2}$. Assuming 253 days per year, the expected value of $\hat{\zeta}_*$ is approximately $3.19\,\text{yr}^{-1/2}$, with standard error around 0.41. This is a very serious bias. The problem is that the 'aspect ratio,' $c_a = p/n$, 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 $\zeta_*^2 > 3c_a$, 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 $\zeta_*^2 \gg c_a$:

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

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

$$\begin{aligned} 1-\alpha/2 &= F_f\left(\left(\frac{n(n-p)}{p(n-1)}\right)\hat{\zeta}_*^2; p, n-p, n\zeta_l^2\right), \\ \alpha/2 &= F_f\left(\left(\frac{n(n-p)}{p(n-1)}\right)\hat{\zeta}_*^2; p, n-p, n\zeta_u^2\right), \end{aligned} \tag{28}$$

where $F_f(x; p, n-p, \delta)$ 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:

$$\begin{aligned} \zeta_l &= \min\left\{ z \,\middle|\, z \geq 0,\ 1-\alpha/2 \geq F_f\left(\left(\frac{n(n-p)}{p(n-1)}\right)\hat{\zeta}_*^2; p, n-p, nz^2\right)\right\}, \\ \zeta_u &= \min\left\{ z \,\middle|\, z \geq 0,\ \alpha/2 \geq F_f\left(\left(\frac{n(n-p)}{p(n-1)}\right)\hat{\zeta}_*^2; p, n-p, nz^2\right)\right\}. \end{aligned} \tag{29}$$

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

2.3 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. [64] 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 $\hat{\zeta}_*^2 \le \frac{c_a}{1-c_a}$, then the MLE of ζ_* is zero, where, again, $c_a = p/n$ is the 'aspect ratio.'

Using the expectation of the non-central F distribution, we can find an unbiased estimator of ζ_*^2. [67] It is given by

$$\mathrm{E}\left[\left(1 - c_a\right)\hat{\zeta}_*^2 - c_a\right] = \zeta_*^2. \tag{30}$$

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. [30]

2.4 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, $\hat{\boldsymbol{\nu}}_*$. 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, the achieved SNR is defined as

$$\zeta_{s,*} =_{\mathrm{df}} \frac{\boldsymbol{\mu}^{\top}\hat{\boldsymbol{\nu}}_*}{\sqrt{\hat{\boldsymbol{\nu}}_*{}^{\top}\Sigma\hat{\boldsymbol{\nu}}_*}}. \tag{31}$$

The haircut is then the quantity,

$$h =_{\mathrm{df}} 1 - \frac{\zeta_{s,*}}{\zeta_*} = 1 - \left(\frac{\hat{\boldsymbol{\nu}}_*{}^{\top}\boldsymbol{\mu}}{\boldsymbol{\nu}_*{}^{\top}\boldsymbol{\mu}}\right)\left(\frac{\sqrt{\boldsymbol{\nu}_*{}^{\top}\Sigma\boldsymbol{\nu}_*}}{\sqrt{\hat{\boldsymbol{\nu}}_*{}^{\top}\Sigma\hat{\boldsymbol{\nu}}_*}}\right), \tag{32}$$

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 $[0, 2]$. When the haircut is larger than 1, the portfolio $\hat{\boldsymbol{\nu}}_*$ has *negative* expected returns.

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 $\hat{\boldsymbol{\nu}}_*$, which is random.

To analyze the haircut, first consider the effects of a rotation of the returns vector. Let P be some invertible square matrix, and let $\boldsymbol{y} = \mathsf{P}^{\top}\boldsymbol{x}$. The population mean and covariance of $\boldsymbol{y}$ are $\mathsf{P}^{\top}\boldsymbol{\mu}$ and $\mathsf{P}^{\top}\Sigma\mathsf{P}$, thus the Markowitz portfolio is $\mathsf{P}^{-1}\Sigma^{-1}\boldsymbol{\mu} = \mathsf{P}^{-1}\hat{\boldsymbol{\nu}}_*$. These hold for the sample analogues as well. Rotation does not change the maximum SNR, since $\left(\mathsf{P}^{\top}\boldsymbol{\mu}\right)^{\top}\left(\mathsf{P}^{\top}\Sigma\mathsf{P}\right)^{-1}\left(\mathsf{P}^{\top}\boldsymbol{\mu}\right) =$

$\boldsymbol{\mu}^{\top}\Sigma^{-1}\boldsymbol{\mu} = \zeta_*$. Rotation does not change the achieved SNR of the sample Markowitz portfolio, since this is

$$\frac{\left(\mathsf{P}^{\top}\boldsymbol{\mu}\right)^{\top}\mathsf{P}^{-1}\hat{\boldsymbol{\nu}}_*}{\sqrt{\left(\mathsf{P}^{-1}\hat{\boldsymbol{\nu}}_*\right)^{\top}\left(\mathsf{P}^{\top}\Sigma\mathsf{P}\right)\left(\mathsf{P}^{-1}\hat{\boldsymbol{\nu}}_*\right)}} = \frac{\boldsymbol{\mu}^{\top}\hat{\boldsymbol{\nu}}_*}{\sqrt{\hat{\boldsymbol{\nu}}_*{}^{\top}\Sigma\hat{\boldsymbol{\nu}}_*}} = \zeta_{s,*}.$$

Thus the haircut is not changed under a rotation. Now choose P to be a square root of Σ^{-1} that rotates $\boldsymbol{\mu}$ onto the first coordinate.[3] That is, pick P such that $\Sigma^{-1} = \mathsf{P}\mathsf{P}^{\top}$ and $\mathsf{P}^{\top}\boldsymbol{\mu} = \left\|\mathsf{P}^{\top}\boldsymbol{\mu}\right\|_2 \boldsymbol{e}_1$. Note that $\left\|\mathsf{P}^{\top}\boldsymbol{\mu}\right\|_2 = \sqrt{\left(\mathsf{P}^{\top}\boldsymbol{\mu}\right)^{\top}\left(\mathsf{P}^{\top}\boldsymbol{\mu}\right)} = \sqrt{\boldsymbol{\mu}^{\top}\Sigma^{-1}\boldsymbol{\mu}} = \zeta_*$. Thus the mean and covariance of $\boldsymbol{y}$ are $\zeta_*\boldsymbol{e}_1$ and $\mathsf{P}^{\top}\Sigma\mathsf{P} = \mathsf{P}^{\top}\left(\mathsf{P}\mathsf{P}^{\top}\right)^{-1}\mathsf{P} = \mathsf{I}$.

3. That is, let P be the orthogonal rotation of a lower Cholesky factor of Σ^{-1}.

So without loss of generality, it suffices to study the case where one observes $\boldsymbol{y}$, forms the Markowitz portfolio and experiences some haircut. But the population parameters associated with $\boldsymbol{y}$ are simpler to deal with, a fact abused in the section.

2.4.1 Approximate haircut under Gaussian returns

A simple approximation to the haircut can be had by supposing that $\hat{\boldsymbol{\nu}}_{y,*} \approx \hat{\boldsymbol{\mu}}_y$. That is, since the population covariance of $\boldsymbol{y}$ is the identity, ignore the contribution of the sample covariance to the Markowitz portfolio. Thus we are treating the elements of $\hat{\boldsymbol{\nu}}_{y,*}$ as independent Gaussians, each zero mean except the first element which has mean ζ_*, and each with variance $\frac{1}{n}$. We can then untangle the contribution of the first element of $\hat{\boldsymbol{\nu}}_{y,*}$ from the denominator by making some trigonometric transforms:

$$\begin{aligned}\tan\left(\arcsin\left(1-h\right)\right) &\sim \frac{\mathcal{N}\left(\zeta_*, 1/n\right)}{\sqrt{\chi^2\left(p-1\right)/n}} \sim \frac{\mathcal{N}\left(\sqrt{n}\zeta_*, 1\right)}{\sqrt{\chi^2\left(p-1\right)}} \\ &\sim \frac{1}{\sqrt{p-1}} t\left(\sqrt{n}\zeta_*, p-1\right). \end{aligned} \quad (33)$$

Here $t\left(\delta, \nu\right)$ is a non-central t-distribution with non-centrality parameter δ and ν degrees of freedom.

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 suspect 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)
```

```r
    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)
    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
n.sim <- 4096
n.stok <- 6
n.yr <- 4
n.obs <- ceiling(ope * n.yr)
zeta.s <- 1.2/sqrt(ope)  # optimal SNR, in daily units

# run some experiments
experiments <- replicate(n.sim, sample.haircut(n.obs,
    n.stok, zeta.s))
hcuts <- experiments[1, ]
```

```r
print(summary(hcuts))

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.01    0.15    0.25    0.30    0.39    1.74

# haircut approximation in the equation above
```

```r
qhcut <- function(p, df1, df2, zeta.s, lower.tail = TRUE) {
    atant <- atan((1/sqrt(df1 - 1)) * qt(p, df = df1 -
        1, ncp = sqrt(df2) * zeta.s, lower.tail = !lower.tail))
    # a slightly better approximation is: retval <- 1 -
    # sin(atant - 0.0184 * zeta.s * sqrt(df1 - 1))
    retval <- 1 - sin(atant)
}
# 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 33 by a Q-Q plot in Figure 3. 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 25%, meaning that the median population SNR of the sample portfolios is around $0.9\text{yr}^{-1/2}$. The maximum value of the haircut over the 4096 simulations, however is 1.74, which is larger than one; this happens if and only if the sample portfolio has negative expected return: $\hat{\boldsymbol{\nu}}_*{}^{\top}\boldsymbol{\mu} < 0$. 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 33 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 $\hat{\boldsymbol{\nu}}^{\top}\boldsymbol{\mu}/\sqrt{\hat{\boldsymbol{\nu}}^{\top}\Sigma^{-1}\hat{\boldsymbol{\nu}}}$ 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 33. The reason is that the error in the approximation of ζ_* is not independent of the modeling error that causes the haircut.

2.4.2 Empirical approximations under Gaussian returns

For 'sane' ranges of n, p, and ζ_*, Monte Carlo studies using Gaussian returns support the following approximations for the haircut, which you should take with a grain of salt:

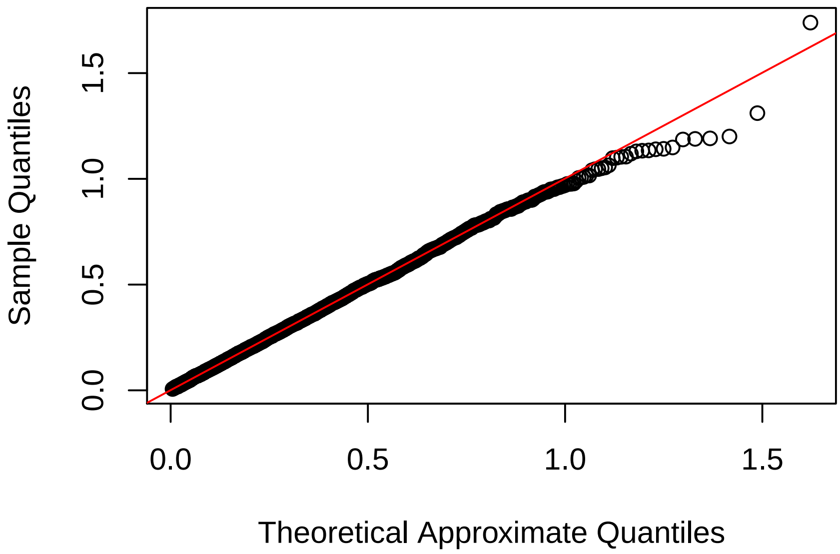

Figure 3: Q-Q plot of 4096 simulated haircut values versus the approximation given by Equation 33 is shown.

$$
\begin{aligned}
h &\approx 1 - \sin\left(\arctan\left(\frac{t}{\sqrt{p-1}}\right) - 0.0184\zeta_*\sqrt{p-1}\right),\\
&\qquad \text{where } t \sim t\left(\sqrt{n}\zeta_*, p-1\right),\\
\operatorname{median}(h) &\approx 1 - \sin\left(\arctan\left(\frac{\sqrt{n}\zeta_*}{\sqrt{p-1}}\right)\right),\\
\operatorname{E}[h] &\approx 1 - \sqrt{\frac{n\zeta_*^2}{p + n\zeta_*^2}},\\
\operatorname{Var}(h) &\approx \frac{p}{\left(p + \left[n\zeta_*^2\right]^{1.08}\right)^2}.
\end{aligned}
\tag{34}
$$

The first of these is a slight modification of the approximation given in Equation 33, which captures some of the SNR loss due to mis-estimation of $\boldsymbol{\Sigma}$. Note that each of these approximations uses the unknown maximal SNR, ζ_*; plugging in the sample estimate $\hat{\zeta}_*$ will give poor approximations because $\hat{\zeta}_*$ is biased. (See Section 2.2 and Section 2.3.)

These approximations are compared to empirical values from the the 4096 Monte Carlo simulations reported above, in Table 4.

	Monte.Carlo	approximation
median	0.25	0.27
mean	0.30	0.30
standard deviation	0.20	0.19

Table 4: Empirical approximate values of the median, mean, and standard deviation of the haircut distribution are given for 4096 Monte Carlo simulations of 1012 days of Gaussian data for 6 assets with $\zeta_* = 1.2\mathrm{yr}^{-1/2}$. The approximations from Equation 34 are also reported.

3 Sharpe ratio and constrained portfolio optimization

3.1 Basic subspace constraint

Let G be an $k_g \times k$ matrix of rank $k_g \leq k$. 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

$$
\max_{\hat{\boldsymbol{\nu}}: \mathsf{G}^C\hat{\boldsymbol{\nu}} = 0,\, \hat{\boldsymbol{\nu}}^\top\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\nu}} \leq R^2} \frac{\hat{\boldsymbol{\nu}}^\top\hat{\boldsymbol{\mu}} - r_0}{\sqrt{\hat{\boldsymbol{\nu}}^\top\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\nu}}}}, \tag{35}
$$

where, as previously, $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Sigma}}$ 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 $\hat{\boldsymbol{\nu}} = \mathsf{G}^\top\hat{\boldsymbol{\nu}}_g$, for some unknown vector $\hat{\boldsymbol{\nu}}_g$. 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

$$\hat{\boldsymbol{\nu}}_{*,\mathsf{G}} =_{\mathrm{df}} c\,\mathsf{G}^{\top}\left(\mathsf{G}\hat{\Sigma}\mathsf{G}^{\top}\right)^{-1}\mathsf{G}\hat{\boldsymbol{\mu}}, \tag{36}$$

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

$$\hat{\zeta}_{*,\mathsf{G}} =_{\mathrm{df}} \frac{\hat{\boldsymbol{\nu}}_{*,\mathsf{G}}{}^{\top}\hat{\boldsymbol{\mu}} - r_0}{\sqrt{\hat{\boldsymbol{\nu}}_{*,\mathsf{G}}{}^{\top}\hat{\Sigma}\hat{\boldsymbol{\nu}}_{*,\mathsf{G}}}} = \sqrt{\left(\mathsf{G}\hat{\boldsymbol{\mu}}\right)^{\top}\left(\mathsf{G}\hat{\Sigma}\mathsf{G}^{\top}\right)^{-1}\left(\mathsf{G}\hat{\boldsymbol{\mu}}\right)} - \frac{r_0}{R}. \tag{37}$$

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 $\hat{\zeta}_{*,\mathsf{G}}$.

3.2 Spanning and hedging

Consider the constrained portfolio optimization problem on k assets,

$$\max_{\hat{\boldsymbol{\nu}}:\mathsf{G}\hat{\Sigma}\hat{\boldsymbol{\nu}}=\boldsymbol{g},\,\hat{\boldsymbol{\nu}}^{\top}\hat{\Sigma}\hat{\boldsymbol{\nu}}\le R^2} \frac{\hat{\boldsymbol{\nu}}^{\top}\hat{\boldsymbol{\mu}} - r_0}{\sqrt{\hat{\boldsymbol{\nu}}^{\top}\hat{\Sigma}\hat{\boldsymbol{\nu}}}}, \tag{38}$$

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

$$\begin{aligned}\hat{\boldsymbol{\nu}}_* &= c\left(\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}} - \mathsf{G}^{\top}\left(\mathsf{G}\hat{\Sigma}\mathsf{G}^{\top}\right)^{-1}\mathsf{G}\hat{\boldsymbol{\mu}}\right) + \mathsf{G}^{\top}\left(\mathsf{G}\hat{\Sigma}\mathsf{G}^{\top}\right)^{-1}\boldsymbol{g},\\ c^2 &= \frac{R^2 - \boldsymbol{g}^{\top}\left(\mathsf{G}\hat{\Sigma}\mathsf{G}^{\top}\right)\boldsymbol{g}}{\hat{\boldsymbol{\mu}}^{\top}\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}} - \left(\mathsf{G}\hat{\boldsymbol{\mu}}\right)^{\top}\left(\mathsf{G}\hat{\Sigma}\mathsf{G}^{\top}\right)^{-1}\left(\mathsf{G}\hat{\boldsymbol{\mu}}\right)},\end{aligned} \tag{39}$$

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

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

using the notation from Section 3.1. 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

$$\hat{\zeta}_{*}^{2} = \hat{\boldsymbol{\mu}}^{\top}\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}} - \left(\mathsf{G}\hat{\boldsymbol{\mu}}\right)^{\top}\left(\mathsf{G}\hat{\Sigma}\mathsf{G}^{\top}\right)^{-1}\left(\mathsf{G}\hat{\boldsymbol{\mu}}\right) = \hat{\zeta}_{*,\mathsf{I}}^{2} - \hat{\zeta}_{*,\mathsf{G}}^{2}, \tag{40}$$

using the notation from Section 3.1, 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 $H_0 : \zeta_{*,\mathsf{I}}^{2} - \zeta_{*,\mathsf{G}}^{2} = 0$, or find point or interval estimates of $\zeta_{*,\mathsf{I}}^{2} - \zeta_{*,\mathsf{G}}^{2}$.

   This generalizes the known tests of *portfolio spanning.* [27, 24] 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

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

   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

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

   The test statistic

   $$F_{\mathsf{G}} = \frac{n-k}{k-k_g}\,\frac{\hat{\zeta}_{*,\mathsf{I}}^{2} - \hat{\zeta}_{*,\mathsf{G}}^{2}}{\frac{n-1}{n} + \hat{\zeta}_{*,\mathsf{G}}^{2}} \tag{41}$$

was shown by Rao to follow an F distribution under the null hypothesis. [55] Giri showed that, under the alternative, and conditional on observing $\hat{\zeta}^2_{*,\mathsf{G}}$,

$$F_{\mathsf{G}} \sim F\left(k - k_g, n - k, \frac{n}{1 + \frac{n}{n-1}\hat{\zeta}^2_{*,\mathsf{G}}}\left(\zeta^2_{*,\mathsf{I}} - \zeta^2_{*,\mathsf{G}}\right)\right), \tag{42}$$

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. See Section D. [18]

3.3 Portfolio optimization with an ℓ_2 constraint

Consider the constrained portfolio optimization problem on k assets,

$$\max_{\hat{\boldsymbol{\nu}}:\hat{\boldsymbol{\nu}}^\top \mathsf{\Gamma}\hat{\boldsymbol{\nu}} \le R^2} \frac{\hat{\boldsymbol{\nu}}^\top \hat{\boldsymbol{\mu}} - r_0}{\sqrt{\hat{\boldsymbol{\nu}}^\top \hat{\mathsf{\Sigma}}\hat{\boldsymbol{\nu}}}}, \tag{43}$$

where R is an ℓ_2 constraint, and $\mathsf{\Gamma}$ is a fixed, symmetric positive definite matrix. This corresponds to the case where one is maximizing Sharpe ratio subject to a volatility constraint imposed by a covariance different from the one used to estimate Sharpe ratio. This can result from *e.g.*, using a longer history to compute $\mathsf{\Gamma}$, or from having an insane risk-manager, *etc.*

Let P be the matrix whose rows are the generalized eigenvalues of $\hat{\mathsf{\Sigma}}, \mathsf{\Gamma}$, and let $\mathsf{\Lambda}$ be the diagonal matrix whose elements are the generalized eigenvalues. That is, we have

$$\hat{\mathsf{\Sigma}}\mathsf{P} = \mathsf{\Gamma}\mathsf{P}\mathsf{\Lambda}, \quad \mathsf{P}^\top\mathsf{\Gamma}\mathsf{P} = \mathsf{I}.$$

Now let $\hat{\boldsymbol{\nu}} = \mathsf{P}\hat{\boldsymbol{v}}$. We can re-frame the original problem, Equation 43, in terms of $\hat{\boldsymbol{v}}$ as follows:

$$\max_{\hat{\boldsymbol{v}}:\hat{\boldsymbol{v}}^\top \hat{\boldsymbol{v}} \le R^2} \frac{\hat{\boldsymbol{v}}^\top \mathsf{P}\hat{\boldsymbol{\mu}} - r_0}{\sqrt{\hat{\boldsymbol{v}}^\top \mathsf{\Lambda}\hat{\boldsymbol{v}}}}, \tag{44}$$

Employing the Lagrange multiplier technique, this optimization problem is solved by

$$\hat{\boldsymbol{v}}_* = c(\mathsf{\Lambda} + \gamma\mathsf{I})^{-1}\mathsf{P}\hat{\boldsymbol{\mu}}, \tag{45}$$

where c is set to satisfy the risk cap, and γ comes from the Lagrange multiplier. To satisfy the risk cap, we should have

$$c = \frac{R}{\sqrt{\hat{\boldsymbol{\mu}}^\top \mathsf{P}^\top (\mathsf{\Lambda} + \gamma\mathsf{I})^{-2}\mathsf{P}\hat{\boldsymbol{\mu}}}}.$$

The problem is reduced to a one-dimensional optimization of γ:

$$\max_{\gamma} \frac{\hat{\boldsymbol{\mu}}^\top \mathsf{P}^\top (\mathsf{\Lambda} + \gamma\mathsf{I})^{-1}\mathsf{P}\hat{\boldsymbol{\mu}} - \frac{r_0}{R}\sqrt{\hat{\boldsymbol{\mu}}^\top \mathsf{P}^\top (\mathsf{\Lambda} + \gamma\mathsf{I})^{-2}\mathsf{P}\hat{\boldsymbol{\mu}}}}{\sqrt{\hat{\boldsymbol{\mu}}^\top \mathsf{P}^\top \mathsf{\Lambda} (\mathsf{\Lambda} + \gamma\mathsf{I})^{-2}\mathsf{P}\hat{\boldsymbol{\mu}}}}. \tag{46}$$

Unfortunately, this problem has to be solved numerically in γ. Moreover, the statistical properties of the resultant optimum are not, to my knowledge, well understood.

3.4 Optimal Sharpe ratio under positivity constraint

Consider the following portfolio optimization problem:

$$\max_{\hat{\boldsymbol{\nu}}:\hat{\boldsymbol{\nu}}\geq 0,\,\hat{\boldsymbol{\nu}}^{\top}\hat{\Sigma}\hat{\boldsymbol{\nu}}\leq R^2} \frac{\hat{\boldsymbol{\nu}}^{\top}\hat{\boldsymbol{\mu}} - r_0}{\sqrt{\hat{\boldsymbol{\nu}}^{\top}\hat{\Sigma}\hat{\boldsymbol{\nu}}}}, \tag{47}$$

where the constraint $\hat{\boldsymbol{\nu}} \geq 0$ is to be interpreted element-wise. In general, the optimal portfolio, call it $\hat{\boldsymbol{\nu}}_{*,+}$, must be found numerically.[4]

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

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

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. [62, 60]

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, $\boldsymbol{\Sigma}$. And so using $\hat{\zeta}^2_{*,+}$ to test the hypothesis $H_0 : \zeta^2_{*,+} = 0$ only yields 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.

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

4 Multivariate inference in unified form

Here I describe a way to think about multivariate distributions that eliminates, to some degree, the distinction between mean and covariance, in order to simplify calculations and exposition. The basic idea is to prepend a deterministic 1 to the random vector, then perform inference on the *uncentered* second moment matrix. A longer form of this chapter is available elsewhere. [51]

Let $\tilde{\boldsymbol{x}}$ be the p-variate vector $\boldsymbol{x}$ prepended with a 1: $\tilde{\boldsymbol{x}} = \left[1, \boldsymbol{x}^{\top}\right]^{\top}$. Consider the second moment of $\tilde{\boldsymbol{x}}$:

$$\Theta =_{\mathrm{df}} \mathrm{E}\left[\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\top}\right] = \left[\begin{array}{cc} 1 & \boldsymbol{\mu}^{\top} \\ \boldsymbol{\mu} & \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^{\top} \end{array}\right]. \tag{48}$$

By inspection one can confirm that the inverse of Θ is

$$\Theta^{-1} = \left[\begin{array}{cc} 1 + \boldsymbol{\mu}^{\top}\Sigma^{-1}\boldsymbol{\mu} & -\boldsymbol{\mu}^{\top}\Sigma^{-1} \\ -\Sigma^{-1}\boldsymbol{\mu} & \Sigma^{-1} \end{array}\right] = \left[\begin{array}{cc} 1 + \zeta^2_{*} & -\boldsymbol{\nu}_{*}{}^{\top} \\ -\boldsymbol{\nu}_{*} & \Sigma^{-1} \end{array}\right].$$

The (upper) Cholesky factor of Θ is

$$\Theta^{1/2} = \left[\begin{array}{cc} 1 & \boldsymbol{\mu}^{\top} \\ 0 & \Sigma^{1/2} \end{array}\right].$$

In some situations, the Cholesky factor of Θ^{-1} might be of interest. In this situation, one can *append* a 1 to $\boldsymbol{x}$ instead of prepending it. When Θ is defined in this way, the Cholesky factor of Θ^{-1} (but not that of Θ) has a nice form:

$$\begin{bmatrix} \Sigma^{-1/2} & -\Sigma^{-1/2}\boldsymbol{\mu} \\ 0 & 1 \end{bmatrix}^{\top} \begin{bmatrix} \Sigma^{-1/2} & -\Sigma^{-1/2}\boldsymbol{\mu} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \Sigma^{-1} & -\boldsymbol{\nu}_* \\ -\boldsymbol{\nu}_*{}^{\top} & 1+\zeta_*^2 \end{bmatrix}, \tag{49}$$

where the latter is Θ^{-1} when defined by appending a 1.

The relationships above are merely facts of linear algebra, and so hold for the sample estimates as well:

$$\begin{bmatrix} 1+\hat{\zeta}_*^2 & -\hat{\boldsymbol{\nu}}_*{}^{\top} \\ -\hat{\boldsymbol{\nu}}_* & \hat{\Sigma}^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} 1 & \hat{\boldsymbol{\mu}}^{\top} \\ \hat{\boldsymbol{\mu}} & \hat{\Sigma}+\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^{\top} \end{bmatrix} = \begin{bmatrix} 1 & \hat{\boldsymbol{\mu}}^{\top} \\ 0 & \hat{\Sigma}^{1/2} \end{bmatrix}^{\top} \begin{bmatrix} 1 & \hat{\boldsymbol{\mu}}^{\top} \\ 0 & \hat{\Sigma}^{1/2} \end{bmatrix}.$$

Given n *i.i.d.* observations of $\boldsymbol{x}$, let $\tilde{\mathsf{X}}$ be the matrix whose rows are the vectors $\tilde{\boldsymbol{x}}_i{}^{\top}$. The naïve sample estimator

$$\hat{\Theta} =_{\mathrm{df}} \frac{1}{n}\tilde{\mathsf{X}}^{\top}\tilde{\mathsf{X}} \tag{50}$$

is an unbiased estimator since $\Theta = \mathrm{E}\left[\tilde{\boldsymbol{x}}^{\top}\tilde{\boldsymbol{x}}\right]$.

4.1 Asymptotic distribution of the Markowitz portfolio

Collecting the mean and covariance into the second moment matrix as we have done gives the asymptotic distribution of the sample Markowitz portfolio without much work. This computation generalizes the 'standard' asymptotic analysis of Sharpe ratio of multiple assets, as in Section 1.5.

Let $\mathrm{vec}\left(\mathsf{A}\right)$, and $\mathrm{vech}\left(\mathsf{A}\right)$ be the vector and half-space vector operators. The former turns an $p \times p$ matrix into an p^2 vector of its columns stacked on top of each other; the latter vectorizes a symmetric (or lower triangular) matrix into a vector of the non-redundant elements. [40]

Define, as we have above, $\hat{\Theta}$ to be the unbiased sample estimate of Θ, based on n *i.i.d.* samples of $\tilde{\boldsymbol{x}}$. Under the multivariate central limit theorem [68]

$$\sqrt{n}\left(\mathrm{vech}\left(\hat{\Theta}\right) - \mathrm{vech}\left(\Theta\right)\right) \rightsquigarrow \mathcal{N}\left(0, \Omega\right), \tag{51}$$

where Ω is the variance of $\mathrm{vech}\left(\hat{\Theta}\right)$, which, in general, is unknown. For the case where $\boldsymbol{x}$ is multivariate Gaussian, Ω is known; see Section **??**.

The Markowitz portfolio appears in $-\hat{\Theta}^{-1}$. Let L be the 'Elimination Matrix,' a matrix of zeros and ones with the property that $\mathrm{vech}\left(\mathsf{A}\right) = \mathsf{L}\,\mathrm{vec}\left(\mathsf{A}\right)$. [40] Let D be the duplication matrix, which has the property that $\mathrm{vec}\left(\mathsf{A}\right) = \mathsf{D}\,\mathrm{vech}\left(\mathsf{A}\right)$. We can find the asymptotic distribution of $\hat{\Theta}^{-1}$ via the delta method. The derivative of the matrix inverse is given by

$$\frac{\mathrm{d}\mathrm{vec}\left(\mathsf{A}^{-1}\right)}{\mathrm{d}\,\mathrm{vec}\left(\mathsf{A}\right)} = -\mathsf{A}^{-1} \otimes \mathsf{A}^{-1}, \tag{52}$$

for symmetric A. [40, 15] We can reduce this to the non-redundant parts via the Elimination matrix:

$$\frac{\mathrm{d}\mathrm{vech}\left(\mathsf{A}^{-1}\right)}{\mathrm{d}\,\mathrm{vech}\left(\mathsf{A}\right)} = \mathsf{L}\frac{\mathrm{d}\mathrm{vec}\left(\mathsf{A}^{-1}\right)}{\mathrm{d}\,\mathrm{vec}\left(\mathsf{A}\right)}\mathsf{D} = -\mathsf{L}\left(\mathsf{A}^{-1} \otimes \mathsf{A}^{-1}\right)\mathsf{D}. \tag{53}$$

Then we have, via the delta method,

$$\sqrt{n}\left(\operatorname{vech}\left(\hat{\Theta}^{-1}\right) - \operatorname{vech}\left(\Theta^{-1}\right)\right) \rightsquigarrow \\ \mathcal{N}\left(0, \left[\mathsf{L}\left(\Theta^{-1} \otimes \Theta^{-1}\right)\mathsf{D}\right]^{\top} \Omega \left[\mathsf{L}\left(\Theta^{-1} \otimes \Theta^{-1}\right)\mathsf{D}\right]\right). \quad (54)$$

To estimate the covariance of $\operatorname{vech}\left(\hat{\Theta}^{-1}\right) - \operatorname{vech}\left(\Theta^{-1}\right)$, plug in $\hat{\Theta}$ for Θ in the covariance computation, and use some consistent estimator for Ω, call it $\hat{\Omega}$. The simple sample estimate can be had by computing the sample covariance of the vectors $\operatorname{vech}\left(\tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i{}^{\top}\right) = \left[1, \boldsymbol{x}_i{}^{\top}, \operatorname{vech}\left(\boldsymbol{x}_i \boldsymbol{x}_i{}^{\top}\right)^{\top}\right]^{\top}$. More elaborate covariance estimators can be used, for example, to deal with violations of the *i.i.d.* assumptions.

Empirically, the marginal Wald test for zero weighting in the Markowitz portfolio based on this approximation are nearly identical to the t-statistics produced by the procedure of Britten-Jones, as shown below. [10]

```
nday <- 1024
nstk <- 5

# under the null: all returns are zero mean;
set.seed(as.integer(charToRaw("7fbb2a84-aa4c-4977-8301-539e48355a35")))
rets <- matrix(rnorm(nday * nstk), nrow = nday)

# t-stat via Britten-Jones procedure
bjones.ts <- function(rets) {
    ones.vec <- matrix(1, nrow = dim(rets)[1], ncol = 1)
    bjones.mod <- lm(ones.vec ~ rets - 1)
    bjones.sum <- summary(bjones.mod)
    retval <- bjones.sum$coefficients[, 3]
}
# wald stat via inverse second moment trick
ism.ws <- function(rets) {
    # flipping the sign on returns is idiomatic,
    asymv <- ism_vcov(-rets)
    asymv.mu <- asymv$mu[1:asymv$p]
    asymv.Sg <- asymv$Ohat[1:asymv$p, 1:asymv$p]
    retval <- asymv.mu/sqrt(diag(asymv.Sg))
}

bjones.tstat <- bjones.ts(rets)
ism.wald <- ism.ws(rets)

# compare them:
print(bjones.tstat)

##  rets1  rets2  rets3  rets4  rets5
##  0.495  0.048  1.208 -0.454 -1.464

print(ism.wald)

## asset_001 asset_002 asset_003 asset_004 asset_005
##     0.496     0.048     1.211    -0.457    -1.464
```

```r
# repeat under the alternative;
set.seed(as.integer(charToRaw("a5f17b28-436b-4d01-a883-85b3e5b7c218")))
zero.rets <- t(matrix(rnorm(nday * nstk), nrow = nday))
mu.vals <- (1/sqrt(253)) * seq(-1, 1, length.out = nstk)
rets <- t(zero.rets + mu.vals)

bjones.tstat <- bjones.ts(rets)
ism.wald <- ism.ws(rets)

# compare them:
print(bjones.tstat)

## rets1 rets2 rets3 rets4 rets5
## -3.74 -1.76 -0.03  2.90  2.54

print(ism.wald)

## asset_001 asset_002 asset_003 asset_004 asset_005
##     -3.69     -1.75     -0.03      2.90      2.54
```

4.2 Unified Multivariate Gaussian

Note that

$$\left(\boldsymbol{x} - \boldsymbol{\mu}\right)^{\top} \Sigma^{-1} \left(\boldsymbol{x} - \boldsymbol{\mu}\right) = \tilde{\boldsymbol{x}}^{\top} \Theta^{-1} \tilde{\boldsymbol{x}} - 1.$$

Using the block determinant formula, we find that Θ has the same determinant as Σ, that is $\left|\Theta\right| = \left|\Sigma\right|$. These relationships hold without assuming a particular distribution for $\boldsymbol{x}$.

Assume, now, that $\boldsymbol{x}$ is multivariate Gaussian. Then the density of $\boldsymbol{x}$ can be expressed more simply as

$$\begin{aligned}
f_{\mathcal{N}}\left(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma\right) &= \frac{1}{\sqrt{\left(2\pi\right)^{p} \left|\Sigma\right|}} \exp\left(-\frac{1}{2} \left(\boldsymbol{x} - \boldsymbol{\mu}\right)^{\top} \Sigma^{-1} \left(\boldsymbol{x} - \boldsymbol{\mu}\right)\right), \\
&= \frac{\left(\left|\Sigma\right|\right)^{-\frac{1}{2}}}{\left(2\pi\right)^{p/2}} \exp\left(-\frac{1}{2}\left(\tilde{\boldsymbol{x}}^{\top} \Theta^{-1} \tilde{\boldsymbol{x}} - 1\right)\right), \\
&= \left(2\pi\right)^{-p/2} \left(\left|\Theta\right|\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\tilde{\boldsymbol{x}}^{\top} \Theta^{-1} \tilde{\boldsymbol{x}} - 1\right)\right), \\
&= \left(2\pi\right)^{-p/2} \exp\left(\frac{1}{2} - \frac{1}{2} \log\left|\Theta\right| - \frac{1}{2} \operatorname{tr}\left(\Theta^{-1} \tilde{\boldsymbol{x}} \tilde{\boldsymbol{x}}^{\top}\right)\right), \\
\therefore -\log f_{\mathcal{N}}\left(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma\right) &= c_p + \frac{1}{2} \log\left|\Theta\right| + \frac{1}{2} \operatorname{tr}\left(\Theta^{-1} \tilde{\boldsymbol{x}} \tilde{\boldsymbol{x}}^{\top}\right),
\end{aligned}$$

for the constant $c_p = e^{\frac{1}{2}} - \frac{p}{2} \log\left(2\pi\right)$.

Given n *i.i.d.* observations of $\boldsymbol{x}$, let $\tilde{\mathsf{X}}$ be the matrix whose rows are the vectors $\boldsymbol{x_i}^{\top}$. Then the negative log density of $\tilde{\mathsf{X}}$ is

$$-\log f_{\mathcal{N}}\left(\tilde{\mathsf{X}}; \Theta\right) = n c_p + \frac{n}{2} \log\left|\Theta\right| + \frac{1}{2} \operatorname{tr}\left(\Theta^{-1} \tilde{\mathsf{X}}^{\top} \tilde{\mathsf{X}}\right).$$

Again let $\hat{\Theta} = \tilde{X}^\top \tilde{X}/n$, the unbiased sample estimate of Θ. Then

$$\frac{-2 \log f_{\mathcal{N}}\left(\tilde{X}; \Theta\right)}{n} = c_p + \log\left|\Theta\right| + \operatorname{tr}\left(\Theta^{-1}\hat{\Theta}\right).$$

By Lemma (5.1.1) of Press [54], this can be expressed as a density on $\hat{\Theta}$, which is a sufficient statistic:

$$\begin{aligned}
\frac{-2 \log f\left(\hat{\Theta}; \Theta\right)}{n} &= \frac{-2 \log f_{\mathcal{N}}\left(\tilde{X}; \Theta\right)}{n} - \frac{2}{n}\left(\frac{n-p-2}{2}\log\left|\hat{\Theta}\right|\right) \\
&\quad - \frac{2}{n}\left(\frac{p+1}{2}\left(n - \frac{p}{2}\right)\log\pi - \sum_{j=1}^{p+1}\log\Gamma\left(\frac{n+1-j}{2}\right)\right), \\
&= c_p - \frac{p+1}{n}\left(n - \frac{p}{2}\right)\log\pi - \frac{2}{n}\sum_{j=1}^{p+1}\log\Gamma\left(\frac{n+1-j}{2}\right) \\
&\quad + \log\left|\Theta\right| - \frac{n-p-2}{n}\log\left|\hat{\Theta}\right| + \operatorname{tr}\left(\Theta^{-1}\hat{\Theta}\right), \\
&= c'_{n,p} + \log\left|\Theta\right| - \frac{n-p-2}{n}\log\left|\hat{\Theta}\right| + \operatorname{tr}\left(\Theta^{-1}\hat{\Theta}\right), \\
&= c'_{n,p} - \log\left|\Theta^{-1}\right| - \frac{n-p-2}{n}\log\left|\hat{\Theta}\right| + \operatorname{tr}\left(\Theta^{-1}\hat{\Theta}\right).
\end{aligned}$$

The density of $\hat{\Theta}$ is thus

$$f\left(\hat{\Theta}; \Theta\right) = c''_{n,p}\frac{\left|\hat{\Theta}\right|^{\frac{n-p-2}{2}}}{\left|\Theta\right|^{\frac{n}{2}}}\exp\left(-\frac{n}{2}\operatorname{tr}\left(\Theta^{-1}\hat{\Theta}\right)\right). \tag{55}$$

Thus $n\hat{\Theta}$ has the same density, up to the leading constant, as a $p+1$-dimensional Wishart random variable with n degrees of freedom and scale matrix Θ. In fact, $n\hat{\Theta}$ is a *conditional* Wishart, conditional on $\hat{\Theta}_{1,1} = 1$.

4.3 Maximum Likelihood Estimator

The maximum likelihood estimator of Θ is found by taking the derivative of the (log) likelihood with respect to Θ and finding a root. However, the derivative of log likelihood with respect to Θ is mildly unpleasant:

$$\begin{aligned}
\frac{\mathrm{d}\log f\left(\hat{\Theta}; \Theta\right)}{\mathrm{d}\Theta} &= -\frac{n}{2}\frac{\mathrm{d}\log\left|\Theta\right|}{\mathrm{d}\Theta} - \frac{n}{2}\frac{\mathrm{d}\operatorname{tr}\left(\Theta^{-1}\hat{\Theta}\right)}{\mathrm{d}\Theta}, \\
&= -\frac{n}{2}\Theta^{-1} + \frac{n}{2}\Theta^{-1}\hat{\Theta}\Theta^{-1},
\end{aligned} \tag{56}$$

However, the derivative with respect to Θ^{-1} is a bit simpler:

$$\begin{aligned}
\frac{\mathrm{d}\log f\left(\hat{\Theta}; \Theta\right)}{\mathrm{d}\Theta^{-1}} &= \frac{n}{2}\frac{\mathrm{d}\log\left|\Theta^{-1}\right|}{\mathrm{d}\Theta^{-1}} - \frac{n}{2}\frac{\mathrm{d}\operatorname{tr}\left(\Theta^{-1}\hat{\Theta}\right)}{\mathrm{d}\Theta^{-1}}, \\
&= \frac{n}{2}\left(\Theta - \hat{\Theta}\right).
\end{aligned} \tag{57}$$

(See Magnus and Neudecker or the Matrix Cookbook for a refresher on matrix derivatives. [41, 53]) Thus the likelihood is maximized by $\Theta_{\mathrm{MLE}} = \hat{\Theta}$, *i.e.*, the unbiased sample estimator is also the MLE. Note that this is also a root of Equation 56.

Since $\Theta_{\mathrm{MLE}} = \hat{\Theta}$, the log likelihood of the MLE is

$$\begin{aligned} \log f\left(\Theta_{\mathrm{MLE}} \middle| \hat{\Theta}\right) &= -\frac{n}{2} c'_{n,p} - \frac{n}{2} \log\left|\Theta_{\mathrm{MLE}}\right| + \frac{n-p-2}{2} \log\left|\hat{\Theta}\right| \\ &\quad + \operatorname{tr}\left({\Theta_{\mathrm{MLE}}}^{-1}\hat{\Theta}\right), \\ &= -\frac{n}{2} c'_{n,p} - \frac{p+2}{2} \log\left|\hat{\Theta}\right| + (p+1). \end{aligned} \tag{58}$$

4.4 Likelihood Ratio Test

Suppose that Θ_0 is the maximum likelihood estimate of Θ under some null hypothesis under consideration. The likelihood ratio test statistic is

$$\begin{aligned} -2\log\Lambda &=_{\mathrm{df}} -2\log\left(\frac{f\left(\Theta_0 \middle| \hat{\Theta}\right)}{f\left(\Theta_{\mathrm{MLE}} \middle| \hat{\Theta}\right)}\right), \\ &= n\left(\log\left|\Theta_0 {\Theta_{\mathrm{MLE}}}^{-1}\right| + \operatorname{tr}\left(\left[{\Theta_0}^{-1} - {\Theta_{\mathrm{MLE}}}^{-1}\right]\hat{\Theta}\right)\right), \\ &= n\left(\log\left|\Theta_0\hat{\Theta}^{-1}\right| + \operatorname{tr}\left({\Theta_0}^{-1}\hat{\Theta}\right) - [p+1]\right). \end{aligned} \tag{59}$$

4.4.1 Tests on the Precision and Markowitz Portfolio

For some conformable symmetric matrices A_i, and given scalars a_i, consider the null hypothesis

$$H_0 : \operatorname{tr}\left(\mathsf{A}_i\Theta^{-1}\right) = a_i,\ i = 1, \ldots, m. \tag{60}$$

The constraints have to be sensible. For example, they cannot violate the positive definiteness of Θ^{-1}, *etc.* Without loss of generality, we can assume that the A_i are symmetric, since Θ is symmetric, and for symmetric G and square H, $\operatorname{tr}(\mathsf{GH}) = \operatorname{tr}\left(\mathsf{G}\frac{1}{2}\left(\mathsf{H} + \mathsf{H}^{\top}\right)\right)$, and so we could replace any non-symmetric A_i with $\frac{1}{2}\left(\mathsf{A}_i + {\mathsf{A}_i}^{\top}\right)$.

Employing the Lagrange multiplier technique, the maximum likelihood estimator under the null hypothesis satisfies

$$\begin{aligned} 0 &= \frac{\mathrm{d}\log f\left(\hat{\Theta}; \Theta\right)}{\mathrm{d}\Theta^{-1}} - \sum_i \lambda_i \frac{\mathrm{d}\operatorname{tr}\left(\mathsf{A}_i\Theta^{-1}\right)}{\mathrm{d}\Theta^{-1}}, \\ &= -\Theta + \hat{\Theta} - \sum_i \lambda_i \mathsf{A}_i, \\ \therefore \Theta_{\mathrm{MLE}} &= \hat{\Theta} - \sum_i \lambda_i \mathsf{A}_i. \end{aligned}$$

The maximum likelihood estimator under the constraints has to be found numerically by solving for the λ_i, subject to the constraints in Equation 60.

This framework slightly generalizes Dempster's "Covariance Selection." [14] Covariance selection reduces to the case where each a_i is zero, and each A_i

is a matrix of all zeros except two (symmetric) ones somewhere in the lower right $p \times p$ sub matrix. In all other respects, however, the solution here follows Dempster.

An iterative technique for finding the MLE based on a Newton step would proceed as follow. [47] Let $\boldsymbol{\lambda}^{(0)}$ be some initial estimate of the vector of λ_i. (A good initial estimate can likely be had by abusing the asymptotic normality result from Section 4.1.) The residual of the k^{th} estimate, $\boldsymbol{\lambda}^{(k)}$ is

$$\boldsymbol{\epsilon}_i^{(k)} =_{\text{df}} \operatorname{tr}\left(\mathsf{A}_i \left[\hat{\Theta} - \sum_j \lambda_j^{(k)} \mathsf{A}_j\right]^{-1}\right) - a_i. \tag{61}$$

The Jacobian of this residual with respect to the l^{th} element of $\boldsymbol{\lambda}_i^{(k)}$ s

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{\epsilon}_i^{(k)}}{\mathrm{d}\lambda_l^{(k)}} &= \operatorname{tr}\left(\mathsf{A}_i \left[\hat{\Theta} - \sum_j \lambda_j^{(k)} \mathsf{A}_j\right]^{-1} \mathsf{A}_l \left[\hat{\Theta} - \sum_j \lambda_j^{(k)} \mathsf{A}_j\right]^{-1}\right), \\ &= \operatorname{vec}\left(\mathsf{A}_i\right)^{\top} \left(\left[\hat{\Theta} - \sum_j \lambda_j^{(k)} \mathsf{A}_j\right]^{-1} \otimes \left[\hat{\Theta} - \sum_j \lambda_j^{(k)} \mathsf{A}_j\right]^{-1}\right) \operatorname{vec}\left(\mathsf{A}_l\right).\end{aligned} \tag{62}$$

Newton's method is then the iterative scheme

$$\boldsymbol{\lambda}^{(k+1)} \leftarrow \boldsymbol{\lambda}^{(k)} - \left(\frac{\mathrm{d}\boldsymbol{\epsilon}^{(k)}}{\mathrm{d}\boldsymbol{\lambda}^{(k)}}\right)^{-1} \boldsymbol{\epsilon}^{(k)}_{.} \tag{63}$$

When (if?) the iterative scheme converges on the optimum, one can compute the likelihood ratio statistic $-2 \log \Lambda$, as defined in Equation 59. By Wilks' Theorem, under the null hypothesis, $-2 \log \Lambda$ is, asymptotically in n, distributed as a chi-square with m degrees of freedom. [70]

5 Miscellanea

5.1 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[5]

5. at no interest!

cash to purchase stock. Under this formulation, the returns aggregate over time by summation just like log returns.

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:

$$\frac{\hat{\zeta}_r - \hat{\zeta}}{\hat{\zeta}} \approx \frac{1}{2} \frac{\sum_i x^2}{\sum_i x}, \tag{64}$$

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.

5.2 Sharpe tricks

5.2.1 Sharpe ratio bounds probability of a loss

Suppose the SNR of a return series is positive. Then, by Cantelli's inequality:

$$\Pr\left\{x < 0\right\} = \Pr\left\{\mu - x > \mu\right\} = \Pr\left\{\mu - x > \zeta\sigma\right\} \le \frac{1}{1 + \zeta^2}.$$

This is a *very* loose upper bound on the probability of a loss, and is fairly useless on any timescale for which the SNR is less than one.

5.3 Sharpe ratio and drawdowns

Drawdowns are the quant's bugbear. Though a fund may have a reasonably high signal-noise ratio, it will likely face redemptions and widespread managerial panic if it experiences a large drawdown. Moreover, drawdowns are a statistically nebulous beast: the sample maximum drawdown does not correspond in an obvious way to some population parameter; the variance of sample maximum drawdown is typically very high; traded strategies are typically cherry-picked to not have a large maximum drawdown in backtests; the distribution of maximum drawdowns is certainly affected by skew and kurtosis, heteroskedasticity, omitted variable bias and autocorrelation. Even assuming *i.i.d.* Gaussian returns, modeling drawdowns is non-trivial. [39, 3]

However, it may be helpful to have a simple model of drawdowns, and there is a connection with the Sharpe ratio. Given n observations of the mark to market of a single asset, p_i, the maximum drawdown is defined as

$$D_n =_{\text{df}} \max_{1 \le i < j \le n} \log\left(\frac{p_i}{p_j}\right). \tag{65}$$

The drawdown is negative the most extreme peak to point log return, and is always non-negative. The maximum drawdown can be expressed as a a percent loss as $100\left(e^{D_n} - 1\right)\%$.

Let x_i be the *log* returns: $x_i = \log \frac{p_i}{p_{i-1}}$, assumed to be *i.i.d.* Let μ and σ be the population mean and standard deviation of the log returns x_i. Now note that

$$\log\left(\frac{p_i}{p_j}\right) = -\sum_{i<k\leq j} x_k = -\left([j-i-1]\,\mu + \sigma\sum_{i<k\leq j} y_k\right),$$
$$= -\sigma\left([j-i-1]\,\zeta + \sum_{i<k\leq j} y_k\right),$$

where y_i is a zero-mean, unit-variance random variable that is a linear function of x_i.

Now re-express the maximum drawdown in units of the volatility of log returns at the sampling frequency:

$$\frac{D_n}{\sigma} = -\min_{1\leq i<j\leq n}\left([j-i-1]\,\zeta + \sum_{i<k\leq j} y_k\right). \tag{66}$$

The volatility is a natural numeraire: one expects an asset with a larger volatility to have larger drawdowns. Moreover, the quantity on the righthand side is a random variable drawn from a one parameter (the signal-noise ratio) family, rather than a two parameter (location and scale) family.

5.3.1 VaR-like constraint

One reasonable way a portfolio manager might approach drawdowns is to define a 'knockout' drawdown from which she will certainly not recover[6] and a maximum probability of hitting that knockout in a given epoch (*i.e.*, n). For example, the desired property might be *"the probability of a 40% drawdown in one year is less than 0.1%."* These constrain the acceptable signal-noise ratio and volatility of the fund.

6. This is certainly a function of the fund's clients, or the PM's boss.

As a risk constraint, this condition shares the hallmark limitation of the value-at-risk (VaR) measure, namely that it may limit the probability of a certain sized drawdown, but not its expected magnitude. For example, underwriting catastrophe insurance may satisfy this drawdown constraint, but may suffer from enormous losses when a drawdown does occur. Nevertheless, this VaR-like constraint is simple to model, and may be more useful than harmful.

Fix the one-parameter family of distributions on y. Then, for given ϵ, δ, and n, the acceptable funds are defined by the set

$$\mathcal{C}\left(\epsilon,\delta,n\right) =_{\mathrm{df}} \left\{(\zeta,\sigma)\,|\,\sigma>0,\,\Pr\left\{D_n\geq\sigma\epsilon\right\}\leq\delta\right\}. \tag{67}$$

It is obvious that the set $\mathcal{C}\left(\epsilon,\delta,n\right)$ is 'lower right monotonic', *i.e.*, a fund with lower volatility or higher signal-noise ratio than a fund in the set is also in the set[7]. That is, if $(\zeta_1,\sigma_1)\in\mathcal{C}\left(\epsilon,\delta,n\right)$ and $\zeta_1\leq\zeta_2$ and $\sigma_2\leq\sigma_1$ then $(\zeta_2,\sigma_2)\in\mathcal{C}\left(\epsilon,\delta,n\right)$.

7. Ignoring the obvious discontinuity at the origin.

When the x are daily returns, the range of signal-noise ratio one may reasonably expect for portfolios of equities is fairly modest. In this case, the lower boundary of $\mathcal{C}\left(\epsilon,\delta,n\right)$ can be approximated by a half space:

$$\left\{(\zeta,\sigma)\in\mathcal{C}\left(\epsilon,\delta,n\right)\,|\,|\zeta|\leq\zeta_{big}\right\}\approx\left\{(\zeta,\sigma)\,|\,\sigma\leq\sigma_0+b\zeta,\,|\zeta|\leq\zeta_{big}\right\},$$

where σ_0 and b are functions of ϵ, δ, n, and the family of distributions on y. The minimum acceptable signal-noise ratio is $-\sigma_0/b$. It may be the case that σ_0 is negative.

References

[1] Digital library of mathematical functions, 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 Becker. Exact simulation of final, minimal and maximal values of Brownian motion and jump-diffusions with applications to option pricing. *Computational Management Science*, 7(1):1–17, 2010. URL http://www.oekonometrie.uni-saarland.de/papers/SimBrownWP.pdf.

[4] 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.

[5] David Birkes and Yadolah Dodge. *Alternative methods of regression*. Wiley series in probability and mathematical statistics. Wiley, New York [u.a.], [2. print.] edition, 1993. URL http://books.google.com/books?id=CIedErj0HKcC.

[6] Fisher Black, M C Jensen, and Myron Scholes. The capital asset pricing model: Some empirical tests. *Studies in the theory of capital markets*, 81(3):79–121, 1972. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=908569&rec=1&srcabs=350100.

[7] Taras Bodnar and Yarema Okhrin. On the product of inverse Wishart and normal distributions with applications to discriminant analysis and portfolio theory. *Scandinavian Journal of Statistics*, 38(2):311–331, 2011. ISSN 1467-9469. doi: 10.1111/j.1467-9469.2011.00729.x. URL http://dx.doi.org/10.1111/j.1467-9469.2011.00729.x.

[8] Taras Bodnar and Wolfgang Schmid. Econometrical analysis of the sample efficient frontier. *The European Journal of Finance*, 15(3):317–335, 2009. doi: 10.1080/13518470802423478. URL http://www.tandfonline.com/doi/abs/10.1080/13518470802423478.

[9] Richard P. Brent. *Algorithms for Minimization without Derivatives*. Prentice-Hall, Englewood Cliffs, N.J., 1973. URL http://books.google.com/books?id=FR_RgSsC42EC.

[10] 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.

[11] Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting*. Springer, 2nd edition, March 2002. ISBN 9780387953519. URL http://books.google.com/books?vid=ISBN0387953515.

[12] Mark M Carhart. On persistence in mutual fund performance. *Journal of Finance*, 52(1):57, 1997. URL http://www.jstor.org/stable/2329556?origin=crossref.

[13] John Howland Cochrane. *Asset pricing*. Princeton Univ. Press, Princeton [u.a.], 2001. ISBN 0691074984. URL http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+322224764&sourceid=fbw_bibsonomy.

[14] A. P. Dempster. Covariance selection. *Biometrics*, 28(1):pp. 157–175, 1972. ISSN 0006341X. URL http://www.jstor.org/stable/2528966.

[15] Paul L. Fackler. Notes on matrix calculus. Privately Published, 2005. URL http://www4.ncsu.edu/~pfackler/MatCalc.pdf.

[16] Eugene F Fama and Kenneth R French. The cross-section of expected stock returns. *Journal of Finance*, 47(2):427, 1992. URL http://www.jstor.org/stable/2329112?origin=crossref.

[17] R. C. Geary. The distribution of "Student's" ratio for non-normal samples. *Supplement to the Journal of the Royal Statistical Society*, 3(2):pp. 178–184, 1936. ISSN 14666162. URL http://www.jstor.org/stable/2983669.

[18] Narayan C. 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.

[19] G. M. Goerg. Lambert W Random Variables - A New Family of Generalized Skewed Distributions with Applications to Risk Estimation. *ArXiv e-prints*, December 2009. URL http://arxiv.org/abs/0912.4554.

[20] G. M. Goerg. The Lambert Way to Gaussianize skewed, heavy tailed data with the inverse of Tukey's h transformation as a special case. *ArXiv e-prints*, October 2010. URL http://arxiv.org/abs/1010.2265.

[21] Georg M. Goerg. *LambertW: Analyze and Gaussianize skewed, heavy-tailed data*, 2012. URL http://CRAN.R-project.org/package=LambertW. R package version 0.2.9.9.

[22] 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".

[23] Roy D Henriksson and Robert C Merton. On market timing and investment performance. II. statistical procedures for evaluating forecasting skills. *The Journal of Business*, 54(4):513–533, 1981. URL http://www.jstor.org/stable/2352722.

[24] 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.

[25] J. D. Jobson and Bob M. Korkie. Performance hypothesis testing with the Sharpe and Treynor measures. *The Journal of Finance*, 36(4):pp. 889–908, 1981. ISSN 00221082. URL http://www.jstor.org/stable/2327554.

[26] 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.

[27] 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.

[28] Peter Kennedy. *A guide to econometrics*. Blackwell Pub., 6 edition, 2008. ISBN 9781405182577. URL http://books.google.com/books?id=ax1QcAAACAAJ.

[29] Andreas Krause. An overview of asset pricing models. 2001. URL http://people.bath.ac.uk/mnsak/Research/Asset_pricing.pdf.

[30] Tatsuya Kubokawa, C. P. Robert, and A. K. Md. E. Saleh. Estimation of noncentrality parameters. *Canadian Journal of Statistics*, 21(1):45–57, 1993. URL http://www.jstor.org/stable/3315657.

[31] Michael H. Kutner, Christopher J. Nachtsheim, and John Neter. *Applied Linear Regression Models*. McGraw-Hill/Irwin, fourth international edition, September 2004. ISBN 0072955678. URL http://books.google.com/books?vid=ISBN0072955678.

[32] 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.

[33] 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.

[34] Robert Lehr. Sixteen S-squared over D-squared: a relation for crude sample size estimates. *Statistics in Medicine*, 11(8):1099–102, 1992. ISSN 0277-6715. doi: 10.1002/sim.4780110811. URL http://www.biomedsearch.com/nih/Sixteen-S-squared-over-D/1496197.html.

[35] Pui-Lam Leung and Wing-Keung Wong. On testing the equality of multiple Sharpe ratios, with application on the evaluation of iShares. *Journal of Risk*, 10(3):15–30, 2008. URL http://www.risk.net/digital_assets/4760/v10n3a2.pdf.

[36] Ying Liu, Marie Rekkas, and Augustine Wong. Inference for the Sharpe ratio using a likelihood-based approach. *Journal of Probability and Statistics*, 2012:24, 2012. doi: 10.1155/2012/878561. URL http://summit.sfu.ca/system/files/iritems1/13253/878561.pdf.

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

[38] Thomas Lumley, Paula Diehr, Scott Emerson, and Lu Chen. The importance of the normality assumption in large public health data sets. *Annu Rev Public Health*, 23:151–69+, 2002. ISSN 0163-7525. doi: 10.1146/annurev.publheath.23.100901.140546. URL http://arjournals.annualreviews.org/doi/pdf/10.1146/annurev.publhealth.23.100901.140546.

[39] Malik Magdon-Ismail, Amir F Atiya, Amrit Pratap, and Yaser S Abu-Mostafa. The sharpe ratio, range, and maximal drawdown of a brownian motion. Technical report, Technical Report TR 02-13, RPI Computer Science, 2002. URL http://intelligenthedgefundinvesting.com/pubs/rb-mapa1.pdf.

[40] Jan R. Magnus and H. Neudecker. The elimination matrix: some lemmas and applications. *SIAM Journal on Algebraic Discrete Methods*, 1(4):422–449, 1980. URL http://www.janmagnus.nl/papers/JRM008.pdf.

[41] Jan R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics: Texts and References Section. Wiley, 3rd edition, 2007. ISBN 9780471986331. URL http://www.janmagnus.nl/misc/mdc2007-3rdedition.

[42] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):pp. 77–91, 1952. ISSN 00221082. URL http://www.jstor.org/stable/2975974.

[43] Harry Markowitz. The early history of portfolio theory: 1600-1960. *Financial Analysts Journal*, pages 5–16, 1999. URL http://www.jstor.org/stable/10.2307/4480178.

[44] Harry Markowitz. Foundations of portfolio theory. *The Journal of Finance*, 46(2):469–477, 2012. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.1991.tb02669.x/abstract.

[45] Elmar Mertens. Comments on variance of the IID estimator in Lo (2002). Technical report, Working Paper University of Basel, Wirtschaftswissenschaftliches Zentrum, Department of Finance, 2002. URL http://www.elmarmertens.com/research/discussion/soprano01.pdf.

[46] 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.

[47] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer series in operations research and financial engineering. Springer, 2006. ISBN 9780387400655. URL http://books.google.com/books?id=VbHYoSyelFcC.

[48] Yarema Okhrin and Wolfgang Schmid. Distributional properties of portfolio weights. *Journal of Econometrics*, 134(1):235–256, 2006. URL http://www.sciencedirect.com/science/article/pii/S0304407605001442.

[49] Yarema Okhrin and Wolfgang Schmid. Estimation of optimal portfolio weights. *International Journal of Theoretical and Applied Finance*, 11 (03):249–276, 2008. URL http://www.worldscientific.com/doi/abs/10.1142/S0219024908004798.

[50] 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.

[51] Steven E. Pav. Asymptotic distribution of the markowitz portfolio. Privately Published, 2013. URL http://arxiv.org/abs/1312.0557.

[52] 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.

[53] Kaare Brandt Petersen and Michael Syskind Pedersen. The matrix cookbook. Privately Published, 2006. URL https://ece.uwaterloo.ca/~ece603/Matrix-Cookbook.pdf.

[54] S. J. Press. *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*. Dover Publications, Incorporated, 2012. ISBN 9780486139388. URL http://books.google.com/books?id=WneJJEHYHLYC.

[55] C. Radhakrishna Rao. *Advanced Statistical Methods in Biometric Research*. John Wiley and Sons, 1952. URL http://books.google.com/books?id=HvFLAAAAMAAJ.

[56] C. Radhakrishna Rao, Helge Toutenburg, Shalabh, Christian Heumann, and Michael Schomaker. *Linear Models and Generalizations: Least Squares and Alternatives*. Springer Series in Statistics Series. Springer, 3rd edition, 2010. ISBN 9783642093531. URL http://books.google.com/books?id=seedcQAACAAJ.

[57] Wolfgang Schmid and Taras Zabolotskyy. On the existence of unbiased estimators for the portfolio weights obtained by maximizing the sharpe ratio. *AStA Advances in Statistical Analysis*, 92(1):29–34, 2008. ISSN 1863-8171. doi: 10.1007/s10182-008-0054-5. URL http://dx.doi.org/10.1007/s10182-008-0054-5.

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

[59] 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.

[60] 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.

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

[62] 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.

[63] Mervyn J. Silvapulle and Pranab Kumar Sen. *Constrained statistical inference : inequality, order, and shape restrictions*. Wiley-Interscience, Hoboken, N.J., 2005. ISBN 0471208272. URL books.google.com/books?isbn=0471208272.

[64] 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.

[65] Nassim Nicholas Taleb. *The Black Swan. The Impact of the Highly Improbable*. Random House Inc., 2008. ISBN 0812979184. URL http://www.fooledbyrandomness.com/.

[66] Gerald van Belle. *Statistical Rules of Thumb (Wiley Series in Probability and Statistics)*. Wiley-Interscience, 1 edition, March 2002. ISBN 0471402273. URL http://www.vanbelle.org/.

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

[68] Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer, 2004. ISBN 9780387402727. URL http://books.google.com/books?id=th3fbFI1DaMC.

[69] 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].

[70] S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):pp. 60–62, 1938. ISSN 00034851. URL http://www.jstor.org/stable/2957648.

[71] John Alexander Wright, Sheung Chi Phillip Yam, and Siu Pang Yung. A note on a test for the equality of multiple Sharpe ratios and its application on the evaluation of iShares. Technical report, 2012. URL http://www.sta.cuhk.edu.hk/scpy/Preprints/John%20Wright/A%20test%20for%20the%20equality%20of%20multiple%20Sharpe%20ratios.pdf. to appear.

A Glossary

μ The true, or population, mean return of a single asset.

σ The population standard deviation of a single asset.

ζ The population signal-to-noise ratio (SNR), defined as $\zeta =_{\mathrm{df}} \mu/\sigma$.

$\hat{\mu}$ The unbiased sample mean return of a single asset.

$\hat{\sigma}$ The sample standard deviation of returns of a single asset.

$\hat{\zeta}$ The sample Sharpe ratio, defined as $\hat{\zeta} =_{\mathrm{df}} \hat{\mu}/\hat{\sigma}$.

n Typically the sample size, the number of observations of the return of an asset or collection of assets.

r_0 The risk-free, or disastrous rate of return.

p Typically the number of assets in the multiple asset case.

$\boldsymbol{\mu}$ The population mean return vector of p assets.

$\boldsymbol{\Sigma}$ The population covariance matrix of p assets.

$\boldsymbol{\nu}_*$ The maximal SNR portfolio, constructed using population data: $\boldsymbol{\nu}_* =_{\mathrm{df}} \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$.

ζ_* The SNR of $\boldsymbol{\nu}_*$.

$\hat{\boldsymbol{\mu}}$ The Sample mean return vector of p assets.

$\hat{\boldsymbol{\Sigma}}$ The sample covariance matrix of p assets.

$\hat{\boldsymbol{\nu}}$ A portfolio, built on sample data.

$\hat{\boldsymbol{\nu}}_*$ The maximal Sharpe ratio portfolio, constructed using sample data: $\hat{\boldsymbol{\nu}}_* =_{\mathrm{df}} \hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$.

$\hat{\zeta}_*$ The Sharpe ratio of $\hat{\boldsymbol{\nu}}_*$.

$F_t\left(x; v_1, \delta\right)$ the CDF of the non-central t distribution, with v_1 degrees of freedom and non-centrality parameter δ, evaluated at x.

$t_q\left(v_1, \delta\right)$ the inverse CDF, or q-quantile of the non-central t distribution, with v_1 degrees of freedom and non-centrality parameter δ.

$F_f\left(x; v_1, v_2\right)$ the CDF of the F distribution, with degrees of freedom v_1 and v_2, evaluated at x.

$F_f\left(x; v_1, v_2, \delta\right)$ the CDF of the non-central F distribution, with degrees of freedom v_1 and v_2 and non-centrality parameter δ, evaluated at x.

$f_q\left(v_1, v_2, \delta\right)$ the inverse CDF, or q-quantile of the non-central F distribution, with degrees of freedom v_1 and v_2 and non-centrality parameter δ.

γ_3 the skew of a random variable.

γ_4 the excess kurtosis of a random variable.

κ_i the i^{th} uncentered moment of a random variable.

$\hat{\kappa}_i$ the i^{th} uncentered sample moment of a sample.

B 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:

$$\mathcal{I}\left(\zeta, \sigma\right) = n \left(\begin{array}{cc} 1 & \frac{\zeta}{2\sigma^2} \\ \frac{\zeta}{2\sigma^2} & \frac{2+\zeta^2}{4\sigma^4} \end{array} \right) \tag{68}$$

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

$$\mathcal{I}^{-1}\left(\zeta, \sigma\right) = \frac{1}{n} \left(\begin{array}{cc} 1 + \zeta^2/2 & -\zeta\sigma^2 \\ -\zeta\sigma^2 & 2\sigma^4 \end{array} \right) \tag{69}$$

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

$$\operatorname{Var}\left(\left[\tilde{\zeta}, \hat{\sigma}^2\right]^{\top}\right) = \left(\begin{array}{cc} \frac{(1+n\zeta^2)(n-1)}{{d_n}^2 n(n-3)} - \zeta^2 & \zeta\sigma^2\left(\frac{1}{d_n} - 1\right) \\ \zeta\sigma^2\left(\frac{1}{d_n} - 1\right) & \frac{2\sigma^4}{n-1} \end{array} \right). \tag{70}$$

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

C 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 [67]:

$$\begin{aligned} \operatorname{E}\left[F\right] &= \frac{v_2}{v_1}\frac{v_1 + \delta}{v_2 - 2}, \\ \operatorname{Var}\left(F\right) &= \left(\frac{v_2}{v_1}\right)^2 \frac{2}{(v_2-2)(v_2-4)}\left(\frac{(\delta+v_1)^2}{v_2-2} + 2\delta + v_1\right). \end{aligned} \tag{71}$$

Suppose T^2 is distributed as a (non-central) Hotelling's statistic for n observations on p assets, with non-centrality parameter δ. Then [4]

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

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

$$\begin{aligned}
\mathrm{E}\left[T^2\right] &= \frac{(n-1)\,(p+\delta)}{n-p-2},\\
\mathrm{Var}\left(T^2\right) &= \frac{2\,(n-1)^2}{(n-p-2)(n-p-4)}\left(\frac{(\delta+p)^2}{n-p-2}+2\delta+p\right).
\end{aligned} \tag{72}$$

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:

$$\begin{aligned}
\mathrm{E}\left[\hat{\zeta}_*^2\right] &= \frac{n-1}{n}\frac{\left(p+n\zeta_*^2\right)}{n-p-2} = \left(1-\frac{1}{n}\right)\frac{\left(c_a+\zeta_*^2\right)}{1-c_a-\frac{2}{n}},\\
\mathrm{Var}\left(\hat{\zeta}_*^2\right) &= \left(\frac{n-1}{n}\right)^2\frac{2}{(n-p-2)(n-p-4)}\left(\frac{(n\zeta_*^2+p)^2}{n-p-2}+2n\zeta_*^2+p\right),\\
&= \left(1-\frac{1}{n}\right)^2\frac{1}{n}\frac{2}{(1-c_a-\frac{2}{n})(1-c_a-\frac{4}{n})}\left(\frac{(\zeta_*^2+c_a)^2}{1-c_a-\frac{2}{n}}+2\zeta_*^2+c_a\right),
\end{aligned} \tag{73}$$

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

The distribution of Hotelling's statistic is known [4] for general $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and can be expressed in terms of a noncentral F-distribution:

$$\frac{n-p}{p(n-1)}T^2 = \frac{n(n-p)}{p(n-1)}\hat{\zeta}_*^2 \sim F\left(\delta, p, n-p\right),$$

where the distribution has p and $n-p$ degrees of freedom, and

$$\delta = n\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = n\zeta_*^2$$

is the non-centrality parameter, and ζ_* is the population optimal SNR.

C.1 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 [67]:

$$\begin{aligned}
\mathrm{E}\left[F\right] &= \frac{v_2}{v_1}\frac{v_1+\delta}{v_2-2},\\
\mathrm{Var}\left(F\right) &= \left(\frac{v_2}{v_1}\right)^2\frac{2}{(v_2-2)(v_2-4)}\left(\frac{(\delta+v_1)^2}{v_2-2}+2\delta+v_1\right).
\end{aligned} \tag{74}$$

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

$$\mathrm{E}\left[\sqrt{F}\right] \approx \sqrt{\mathrm{E}\left[F\right]} - \frac{\frac{{v_2}^2\left(\delta^2+(v_1+2)\,(2\,\delta+v_1)\right)}{{v_1}^2\,(v_2-4)\,(v_2-2)} - \left(\mathrm{E}\left[F\right]\right)^2}{8\,\left(\mathrm{E}\left[F\right]\right)^{\frac{3}{2}}}. \tag{75}$$

D Untangling Giri

Here I translate Giri's work on Rao's LRT into the terminology used in the rest of this note. [18] In equation (1.9), Giri defines the LRT statistic Z by

$$Z =_{\mathrm{df}} \frac{1 - N\bar{X}_{[2]}^{\top}\left(S_{22} + N\bar{X}_{[2]}\bar{X}_{[2]}^{\top}\right)^{-1}\bar{X}_{[2]}}{1 - N\bar{X}_{[1]}^{\top}\left(S_{11} + N\bar{X}_{[1]}\bar{X}_{[1]}^{\top}\right)^{-1}\bar{X}_{[1]}}. \tag{76}$$

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

$$Z = \frac{1 + N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}}{1 + N\bar{X}_{[2]}^{\top}{S_{22}}^{-1}\bar{X}_{[2]}}. \tag{77}$$

In section 3, Giri notes that, conditional on observing R_1, Z takes a (non-central) 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' non-central beta, which can be transformed into a noncentral F:

$$\frac{N-p}{p-q}\frac{1-Z}{Z} = \frac{N-p}{p-q}\frac{N\bar{X}_{[2]}^{\top}{S_{22}}^{-1}\bar{X}_{[2]} - N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}}{1 + N\bar{X}_{[1]}^{\top}{S_{11}}^{-1}\bar{X}_{[1]}}. \tag{78}$$

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 $\bar{X}_{[1]}^{\top} {S_{11}}^{-1} \bar{X}_{[1]}$,

$$\frac{N-p}{p-q} \frac{N \bar{X}_{[2]}^{\top} {S_{22}}^{-1} \bar{X}_{[2]} - N \bar{X}_{[1]}^{\top} {S_{11}}^{-1} \bar{X}_{[1]}}{1 + N \bar{X}_{[1]}^{\top} {S_{11}}^{-1} \bar{X}_{[1]}} \sim F\left(p-q, N-p, \frac{N\left(\xi^{\top} \Sigma^{-1} \xi - {\xi_{[1]}}^{\top} {\Sigma_{11}}^{-1} \xi_{[1]}\right)}{1 + N \bar{X}_{[1]}^{\top} {S_{11}}^{-1} \bar{X}_{[1]}}\right). \tag{79}$$

Now note that S_{11} refers to the sample Gram matrix, and thus S_{11}/N is the biased covariance estimate, $\tilde{\mathsf{\Sigma}}$ 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 3.2 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

$$\frac{n-k}{k-k_g} \frac{\hat{\zeta}_{*,\mathsf{I}}^{2} - \hat{\zeta}_{*,\mathsf{G}}^{2}}{(n-1)/n + \hat{\zeta}_{*,\mathsf{G}}^{2}} \sim F\left(k-k_g, n-k, \frac{n}{1 + \frac{n}{n-1} \hat{\zeta}_{*,\mathsf{G}}^{2}} \left(\zeta_{*,\mathsf{I}}^{2} - \zeta_{*,\mathsf{G}}^{2}\right)\right). \tag{80}$$