A Short Sharpe Course

v0.2.999

Steven E. Pav *

Sunday 11th April, 2021

* steven@gilgamath.com Gilgamath Consulting.

This is version v0.2.999, revision `f342e6efb6e565e824c55d416cae481051812b7e` of this text. The latest version of this text is available at http://www.gilgamath.com/pages/ssc.html.

For Oskar and Laszlo.

Contents

List of Figures

List of Tables

Preface

> I am at the moment writing a lengthy indictment against our century. When my brain begins to reel from my literary labors, I make an occasional cheese dip.
>
> *(John Kennedy Toole,* A Confederacy of Dunces*)*

I wrote this book for myself. The material here was born from notes I kept through a eight year struggle to find profitable quantitative trading strategies. My notes contained known results and simple exercises. Over time, I made what I thought were novel discoveries, resulting in a few papers. [121, 123, 125, 129, 130] This is an odd book–probably too theoretical for most practicing 'quants', too basic for most statisticians, and not rigorous enough for most mathematicians; as a textbook it lacks enough motivation, as a reference, too few citations. I hope you, the reader, will find this book useful.

It should be no surprise to the reader that I have *no* formal training in statistics; after all, no classically trained statistician would devote so much attention to as mundane a topic as the t-statistic. My initial interest in statistics was sparked by an allergic reaction to the over-reliance on heuristics during my tenure as a quant. Perhaps only because I was allowed to discover the topic via my own random walk through papers, books, and blog posts was it that I came to appreciate and (mis)understand statistics as I do. The gaps in my education are undeniable, and may appear as the occasional heterodoxy. I apologize to the reader in advance.

About this edition

My original intent was to write a book with two major parts: one on the Sharpe ratio, the other on the Markowitz portfolio. This text is a rough draft of the first part, focused entirely on the Sharpe ratio. The remainder, which explores the statistical properties of the portfolio that maximizes the Sharpe ratio, overfitting, backtesting, and market timing will only be available in the book, to be published in August, 2021. [131]

This document was assembled using the **knitr** package; the analysis and presentation relied heavily on the **ggplot2**, **dplyr**, **xts**, and **Rcpp** packages. [176, 169, 170, 145, 40] Much of the bespoke analysis of the Sharpe ratio and the Markowitz portfolio is (or ought to be) via the author's own **SharpeR** and **MarkowitzR** packages. [122, 124]

Acknowledgements

Thanks to those who have taught me this. Thanks to the days which have taught me this.

(Dag Hammarsköld, Markings*)*

For good cheer over the years, I thank Daniel Seifert, Stephen Markacs, Marius Fromm, Paul Ford, Joy Patterson, Heather Rowe, Ian Rothwell, Mark Bishop, Abie Flaxmann, Jessi Berklehammer, Devon Yates, Annie Owens, Markus Fuchs. My work days have benefited from the camraderie of many, notably George Papcun, Jim Bunch, Mike O'Neil, Radu Mihaescu, Jesse Davis, Vincent Pribish, Jake Freifeld, Geoff Anders, Max Dama, Nate Eiseman, and Eric Wiewiora. I owe a great debt those teachers who taught me math, and who taught me to love math, including Daniel Hall, Arlen Jividen, David Kelley, Thomas Hull, Addison Frey, Larry Moss, Bill Hrusa, Noel Walkington, my first math teacher, Nancy Pav, and many others.

This work would have been considerably different without the **R** language and the packages freely available from CRAN; I give thanks to those who maintain the language and the repository, and who author and maintain packages. A shoutout is owed to all scholars who make their work freely accessible via the internet–in so doing, they have made my study possible without access to a good physical library. As always, acknowledgments to Dickie Birkenbush.

For firm support and love over the years, much thanks to Richard, Nancy, Bill, and Dan Pav, and to Linda Shipley, Oskar and Laszlo Pav.

Many years ago I found my way through a crossroads through the effort of William Zeimer, to whom I am ever grateful.

Lastly, I must thank Linda, for teaching me the value of everything.

San Francisco,
April, 2021

1. Preliminaries

Every journey of a thousand miles begins with a single generous donation.

(Matt Groening, Life in Hell*)*

The trouble with being educated is that it takes a long time; it uses up the better part of your life and when you are finished what you know is that you would have benefited more by going into banking.

(Phillip K. Dick, The Transmigration of Timothy Archer*)*

This chapter contains some preliminary material needed in the rest of the text: a hodgepodge of definitions, notational declarations, data descriptions, *etc.*

1.1. Linear Algebra

As much as possible, we will denote vectors by bold lower case Roman letters: $\boldsymbol{v}$, $\boldsymbol{x}$, $\boldsymbol{y}$, *etc.* Matrices will typically be denoted in Roman or Greek bold upper case letters: A, X, $\mathsf{\Sigma}$. A vector is typically considered a column vector so that $\mathsf{A}\boldsymbol{x}$ and $\boldsymbol{x}^{\top}\mathsf{Q}\boldsymbol{x}$ are well-formed. We will denote a vector of all ones by $\mathbf{1}$; all zeros by $\mathbf{0}$. The identity matrix is denoted I; the zero matrix $\mathsf{0}$. Let $\boldsymbol{e}_i$ be the i^{th} column of the identity matrix, where the size of the vector should be implicit from context. Similarly, define the 'single entry' matrix, J^{ij} as the matrix of all zeros but for a single 1 in the i, j location. Equivalently, we can write $\mathsf{J}^{ij} = \boldsymbol{e}_i{\boldsymbol{e}_j}^{\top}$.

Definition 1.1.1 (Matrix operations)**.** We will write A^{-1} for the inverse of non-singular matrix A. Rarely we will write A^{+} for the *Moore-Penrose pseudoinverse.* For square symmetric non-singular matrix A, let $\mathsf{A}^{1/2}$ be the lower triangular *Cholesky factor* of A, defined as the lower triangular matrix such that $\mathsf{A}^{1/2}\mathsf{A}^{\top/2} = \mathsf{A}$. Let the inverse of $\mathsf{A}^{1/2}$ be denoted by $\mathsf{A}^{-1/2}$; the inverse of $\mathsf{A}^{\top/2}$ is $\mathsf{A}^{-\top/2}$.

Definition 1.1.2 (Matrix functionals)**.** For square matrix A, the *trace* of A, denoted

$\operatorname{tr}(\mathsf{A})$ is the sum of the diagonal elements:

$$\operatorname{tr}(\mathsf{A}) = \sum_i A_{ii}.$$

Let $|\mathsf{A}|$ be the determinant of A.

Definition 1.1.3 (Converting between matrices and vectors)**.** For vector $\boldsymbol{x}$, define $\operatorname{Diag}(\boldsymbol{x})$ as the diagonal matrix whose diagonal equals $\boldsymbol{x}$. For matrix A, define $\operatorname{diag}(\mathsf{A})$ as the vector of the diagonal of A. For matrix A, let $\operatorname{vec}(\mathsf{A})$, and $\operatorname{vech}(\mathsf{A})$ 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. Let L be the 'Elimination matrix,' a matrix of zeros and ones with the property that $\operatorname{vech}(\mathsf{A}) = \mathsf{L}\operatorname{vec}(\mathsf{A})$. The 'Duplication matrix,' D, is the matrix of zeros and ones that reverses this operation: $\mathsf{D}\operatorname{vech}(\mathsf{A}) = \operatorname{vec}(\mathsf{A})$. [104] Note that this implies that

$$\mathsf{LD} = \mathsf{I}\left(\neq \mathsf{DL}\right).$$

Let K be the 'commutation matrix', the matrix whose rows are a permutation of the rows of the identity matrix such that $\mathsf{K}\operatorname{vec}(\mathsf{A}) = \operatorname{vec}\left(\mathsf{A}^{\top}\right)$ for square matrix A. [106] Let N be the 'symmetric idempotent matrix,' defined as $\mathsf{N} =_{\mathrm{df}} \frac{1}{2}\left(\mathsf{I} + \mathsf{K}\right)$. (Here we use '$=_{\mathrm{df}}$' to mean "defined as." The equation above should be read as "N is defined as $\frac{1}{2}\left(\mathsf{I} + \mathsf{K}\right)$.") This matrix has many interesting properties, for which we refer the reader to Magnus and Neudecker. [105]

Definition 1.1.4 (Matrix products)**.** For conformable matrices A, B, the matrix product is denoted by string concatenation: AB is the product. The *Kronecker product* will be denoted by $\mathsf{A} \otimes \mathsf{B}$. This is defined blockwise as

$$\mathsf{A} \otimes \mathsf{B} =_{\mathrm{df}} \begin{bmatrix} A_{11}\mathsf{B} & A_{12}\mathsf{B} & \ldots & A_{1n}\mathsf{B} \\ A_{21}\mathsf{B} & A_{22}\mathsf{B} & \ldots & A_{2n}\mathsf{B} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1}\mathsf{B} & A_{m2}\mathsf{B} & \ldots & A_{mn}\mathsf{B} \end{bmatrix}. \tag{1.1}$$

The *Hadamard product*, or elementwise product, is denoted by $\mathsf{A} \odot \mathsf{B}$. This is defined only for matrices of the same size, and is defined elementwise: $\left(\mathsf{A} \odot \mathsf{B}\right)_{ij} = A_{ij}B_{ij}$. Similarly we may define Hadamard ratios: $\mathsf{A} \oslash \mathsf{B}$ is the elementwise ratio $\left(\mathsf{A} \oslash \mathsf{B}\right)_{ij} = A_{ij}/B_{ij}$. At times we may write the Hadamard power (*i.e.*, elementwise) of a vector or matrix as $\boldsymbol{v}^k$ or A^k.

1.2. The Data

Throughout this text, a number of datasets of real historical returns will be considered.

Example 1.2.1 (Fama-French Four Factor monthly returns). The monthly returns of 'the Market' portfolio, the small cap portfolio (known as SMB, for 'small minus big'), and the value portfolio (known as HML, for 'high minus low'), are published periodically by Kenneth French. [50] These are the celebrated three factor portfolios of Fama and French. [43] Together with these three, we also consider the returns of the momentum portfolio (know as UMD, for 'up minus down'), making a total of four factors. [26] Throughout this text, we refer to these as "the" four factors.

The monthly returns data were downloaded directly from French's site [50] and stored in the data package **aqfb.data**, which is a companion package to this text[1][2]. [128] One can access this data in **R** as follows, which puts an `xts` object called `mff4` into the environment:

1. The data are also attached to this PDF.

```
require(aqfb.data)
data(mff4)
```

The data are distributed as monthly relative returns, quoted in percents. The Market return is quoted as an excess return, with the risk free rate subtracted out. The risk free rate is also tabulated. For our purposes, the raw market returns are needed, so the risk free rate is added back to the market returns. The risk free rate, denoted RF, is kept as a kind of inflationary benchmark.

The set consists of 1104mo. of data, from Jan 1927 through Dec 2018. The cumulative value of the factor returns (and of the risk-free rate) are shown in Figure 1.1. By 'cumulative value', we mean the value of a one dollar initial investment in the portfolio replicating the factor returns, ignoring all costs and trade frictions. These are shown in nominal dollars, ignoring the changes in the real buying power of a dollar due to inflation, which is presumably related to the cumulative value of RF. Correlations of the returns, for a subset of the time history, are given in Example 1.2.2.

Typically in this text, the monthly returns of the factors will be used. However, at times the daily returns may be needed. These are available in **aqfb.data** as follows:

```
library(aqfb.data)
data(dff4)
```

⊣

Caution (Factors are not assets). We will use the Fama French factor returns extensively in this text. One should not assume, however, that these returns can easily be captured. While there are many investment vehicles that track the Market returns very well, the other factors' returns are harder to realize, especially under, say, a long-only investment constraint.

Despite that deficiency, we consider the Fama French factor returns here for a number of reasons: 1. they exhibit the oddities of investment returns, like correlation, heteroskedasticity, autocorrelation, *etc.* 2. they are well defined and freely available.

[2] Ken French does a great service to the community by computing and publishing his data. However, there is a 'last mile' delivery problem, as the data are published in zipped, oddly named, fixed-width files with header and footer junk. To simplify access, the data have been repackaged in **aqfb.data**.

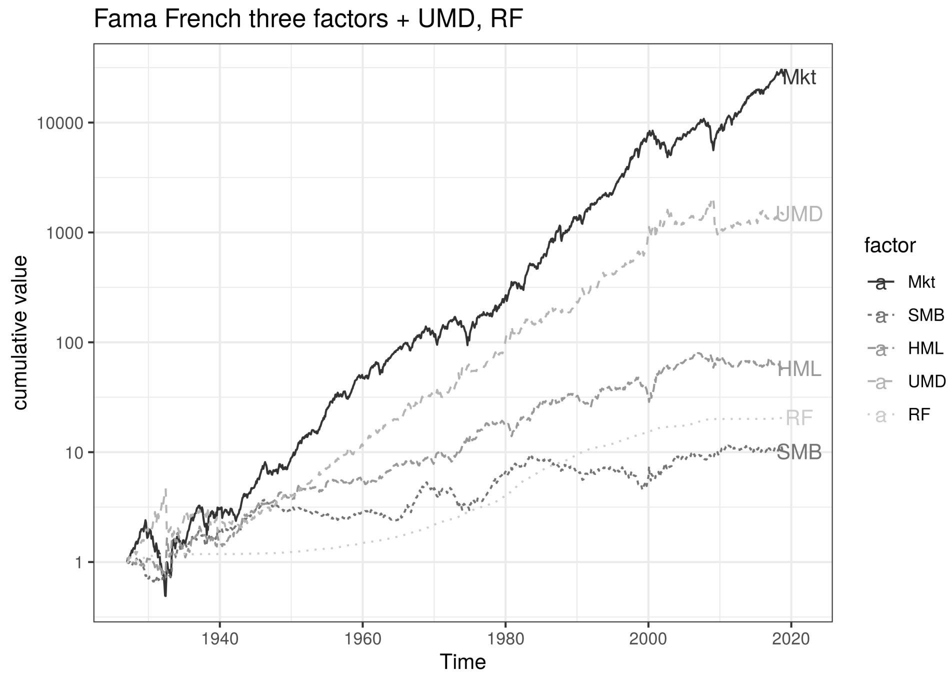

Figure 1.1.: The cumulative value of the four Fama French factors, plus the risk-free rate, are shown, from Jan 1927 through Dec 2018. Cumulative value is normalized to the initial investment, ignores all trading costs, and is not adjusted for inflation.

Example 1.2.2 (Fama-French Six Factor monthly returns)*.* Fama and French also consider a five-factor model of returns, adding profitability (yclept RMW, for 'robust minus weak') and investment (CMA, for 'conservative minus aggressive') factors to the classic three factor model. [44, 8] These are computed and distributed by Ken French, and available via the **aqfb.data** package[3]. [50, 128]

The monthly returns can be loaded via:

```
# a0c1a9e8-aeb3-41b5-b1da-522d8e5a2c04
require(xts)
library(aqfb.data)
data(mff6)
```

Again, the data are distributed as monthly relative returns, quoted in percents. The risk free rate is added back to excess market returns to get raw market returns. Since the upstream data used to determine the RMW and CMA factors 'only' goes back to the 1960's, the set consists of 'only' 666mo. of data, from Jul 1963 through Dec 2018. The cumulative value of the six factors, and RF, are shown in Figure 1.2. The

3. Data are attached to this PDF.

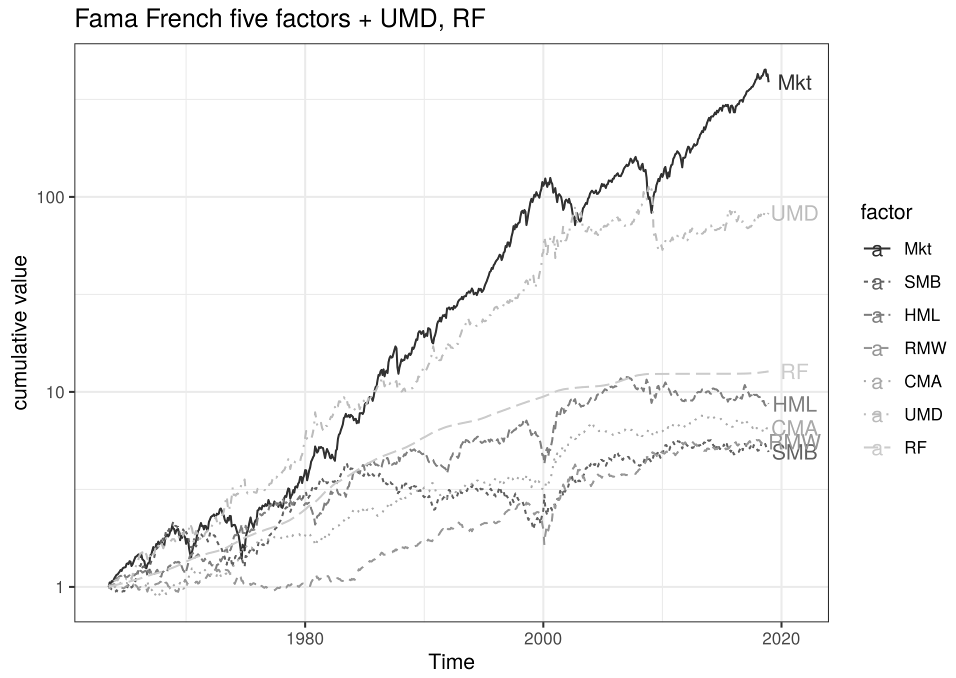

Figure 1.2.: The cumulative value of the six Fama French factors, plus the risk-free rate, are shown, from Jul 1963 through Dec 2018. Cumulative value is normalized to the initial investment, ignores all trading costs, and is not adjusted for inflation.

monthly returns of the six factors and RF are shown in Figure 1.3. The monthly returns are scattered against each other in Figure 1.3, which also shows the Pearson correlations on the upper triangle. The maximum correlation is approximately 0.7, achieved between CMA and HML.

⊣

Example 1.2.3 (French 5 Industry monthly returns). Kenneth French tabulates the returns of five portfolios constructed by industry classification. The five portfolios correspond to 'Consumer goods,' 'Manufacturing,' 'High-Tech,' 'Healthcare,' and 'Other'. [51, 50] The monthly returns of value-weighted portfolios were downloaded from French's library and available in the **aqfb.data** package[4]: [50, 128]

4. Data are attached to this PDF.

```
library(aqfb.data)
data(mind5)
```

The data are distributed as monthly relative returns, quoted in percents. The set consists of 1104mo. of data, from Jan 1927 through Dec 2018. The cumulative values of the industry portfolios are shown in Figure 1.4. The portfolios are clearly highly

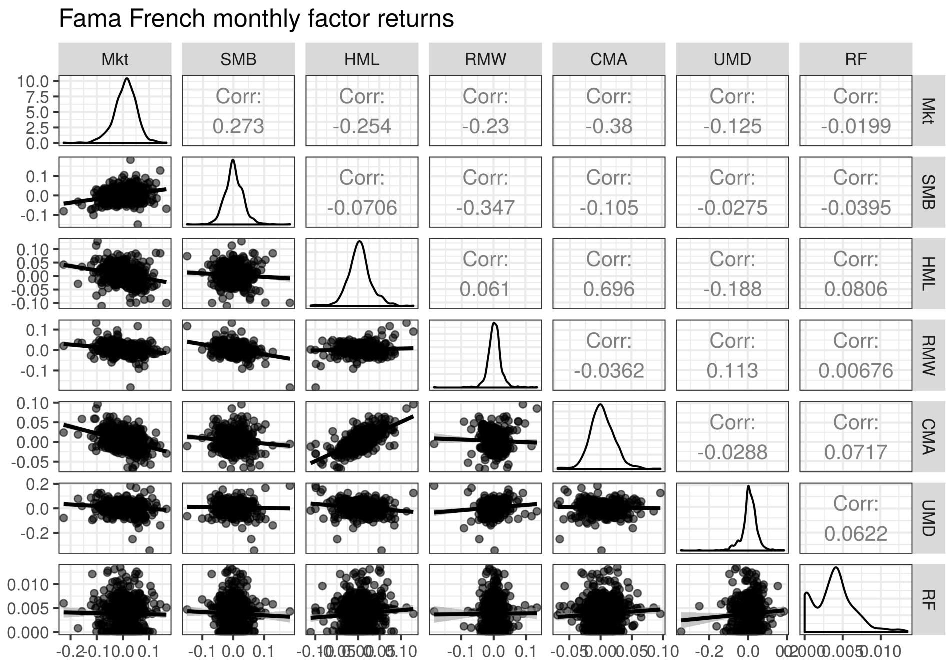

Figure 1.3.: The monthly returns of the six Fama French factors, plus the risk-free rate, from Jul 1963 through Dec 2018 are scattered against each other. In the lower triangle, returns are scattered against each other. Pearson correlations are given in the upper triangle. On the diagonal, the empirical distribution of returns are plotted.

correlated. Monthly returns are scattered against each other in Figure 1.5; the smallest correlation between two portfolios is 0.71.

⊣

Example 1.2.4 (The VIX). The *VIX index* is a 'model-free' estimate of the volatility of the market over the next thirty calendar days, constructed from the bid and ask prices of options on the S&P 500 index, expressed in units of annualized percent. [27] The daily value of the VIX index is computed by the CBOE, and available from the **aqfb.data** package[5]. [29, 128] This dataset includes the 'back-computation' of the VIX using the post-2004 methodology on data back to 1999. So the following **R** code gives access to 7304 day of data, from 1990-01-02 through 2018-12-31:

5. Data are attached to this PDF.

```
library(aqfb.data)
data(dvix)
```

We plot the VIX index in Figure 1.6. The mean value of the VIX index over this period is approximately 19.3, and the standard deviation is 7.81. The coefficient of variation of the VIX is thus around 0.41. The median level is around 17.4, indicating

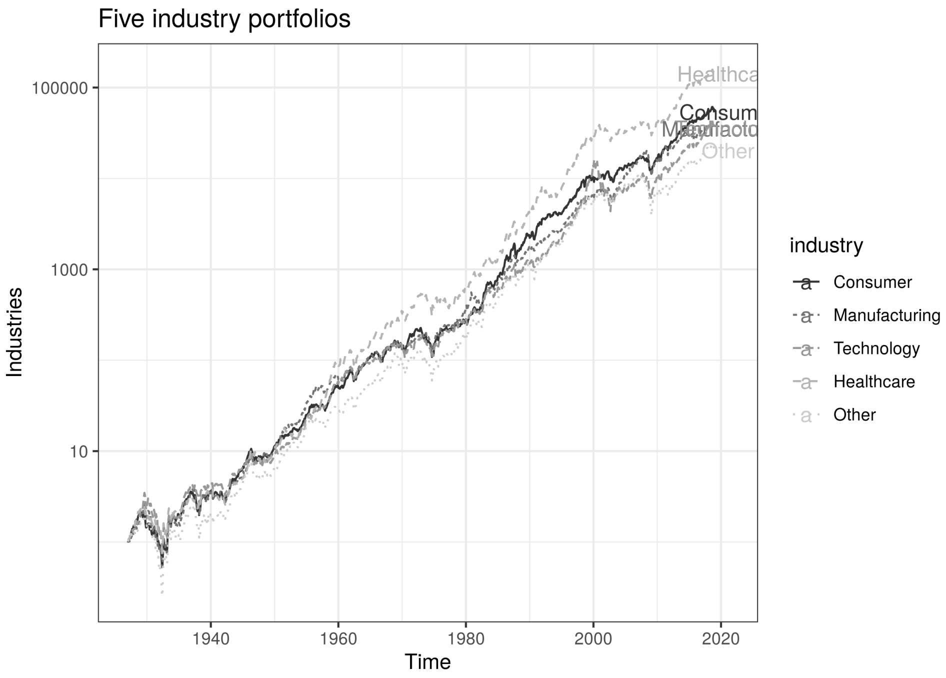

Figure 1.4.: The cumulative value of the five industry portfolios are shown, from Jan 1927 through Dec 2018. Cumulative value is normalized to the initial investment, ignores all trading costs, and is not adjusted for inflation.

some amount of positive skew. The empirical skewness is computed as 2.09. We will later need to consider a normalized version of the VIX, with mean value one; we estimate the third centered moment of this normalized VIX to be around 0.139. That is

$$\frac{1}{n}\sum_{1\le i\le n}\left(\frac{v_i}{\frac{1}{n}\sum_{1\le j\le n} v_j} - 1\right)^3 \approx 0.139.$$

The VIX level is clearly autocorrelated. The first differences in the VIX level are mean reverting, as shown in Figure 1.7. The VIX level is often modeled as an *Ornstein-Uhlenbeck process.* [101, 181]

⊣

1.3. Probability Distributions

A comprehensive treatment of probability distributions is beyond the scope of this book, or any one single book. The reader is directed to the texts of: Walck a good free

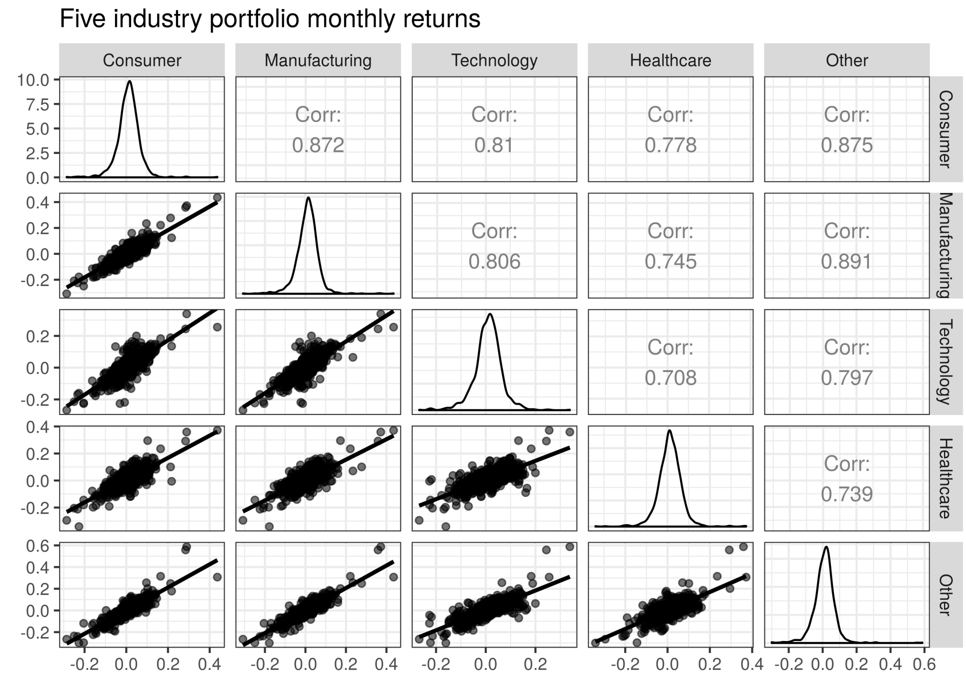

Figure 1.5.: The monthly returns of the five industry portfolios, from Jan 1927 through Dec 2018 are scattered against each other. In the lower triangle, returns are scattered against each other. Pearson correlations are given in the upper triangle. On the diagonal, the empirical distribution of returns are plotted. Industry portfolio returns are clearly highly correlated.

overview of univariate distributions [165]; Krishnamoorthy, for a more theoretical, but similar overview [83]; Press, for multivariate distributions. [138]

Vaguely speaking, the various distributions can be described as follows:

Normal distribution The granddaddy of continuous distributions tends to arise whenever you sum independent random variables, a consequence of the Central limit theorem.

Chi square This is the sum of the squares of zero mean normal random variables. These can also arise as the sums of squares of independent random variables, due to a kind of central limit theorem for squared variables. [63] The more general form is the Gamma distribution.

non-central Chi square This is a Chi square where the summed normal variates do not have zero mean.

Chi This is the positive square root of a Chi square.

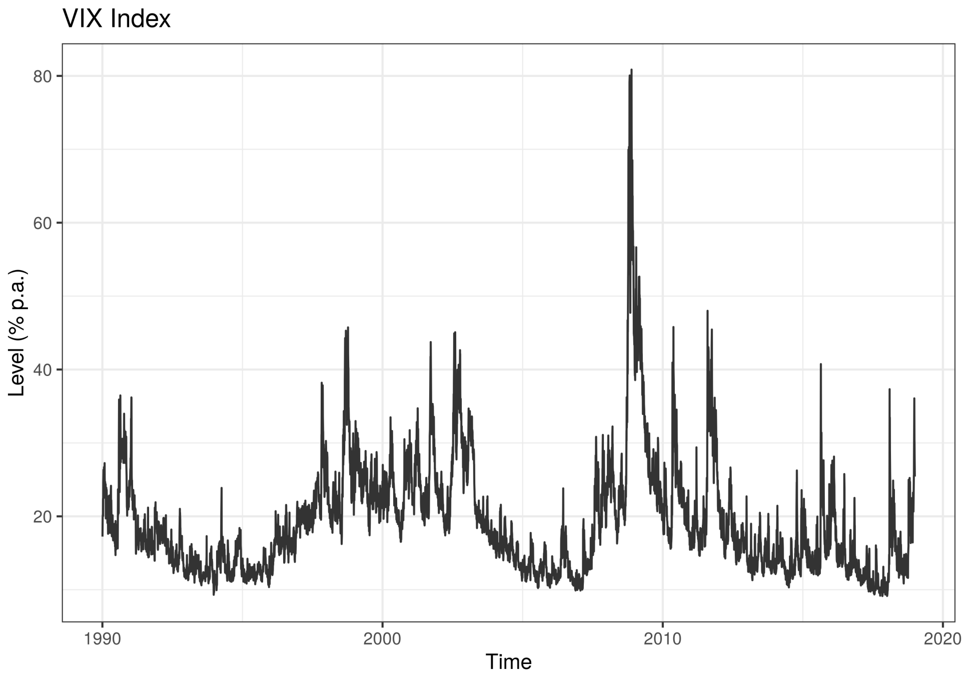

Figure 1.6.: The level of the VIX index is shown, from 1990-01-02 through 2018-12-31

t distribution This arises as the ratio of a normal to a (independent) Chi variate. Typically this occurs when dividing a sample average by some estimate of the standard deviation of the process. For large samples it is basically normal.

non-central t distribution The ratio of a normal with non-zero mean to an independent Chi.

F This is the (rescaled) ratio of Chi square variates. The square of a t is an F, up to scaling. It typically arises as the ratio of the square of some averages to the estimated variance of the process. For large sample sizes, the variation in the denominator is often ignored, and a Chi squared approximation is made.

beta This is the ratio of a Chi square to itself plus another independent Chi square. A trigonometric transform (Equation 2.24) relates the beta to the (rescaled) F distribution.

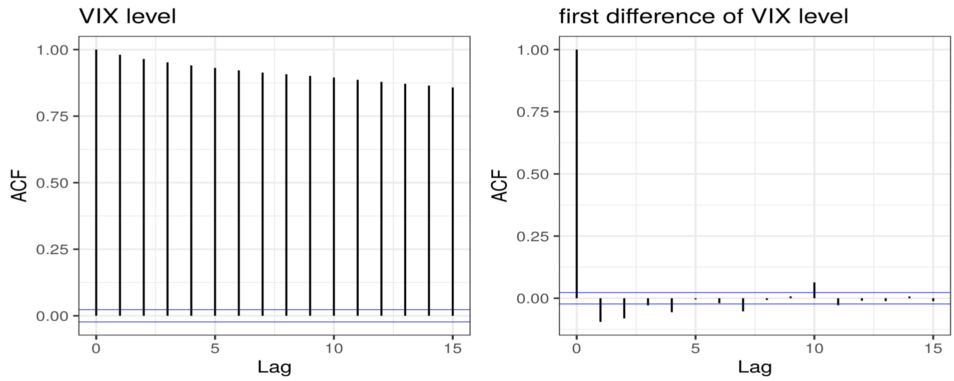

Figure 1.7.: Autocorrelation plots of the VIX level, and the first difference of the VIX level are shown, along with the standard 95% confidence intervals.

1.3.1. The Multivariate Normal Distribution

We say that k-vector $\boldsymbol{x}$ follows a *multivariate normal* (or *Gaussian*) distribution with mean $\boldsymbol{\mu}$ and covariance Σ, if the density of $\boldsymbol{x}$ is

$$\phi\left(\boldsymbol{x}\right) = \frac{1}{\sqrt{\left(2\pi\right)^{k}\left|\Sigma\right|}} \exp\left(-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^{\top}\Sigma^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right). \tag{1.2}$$

We write this as $\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \Sigma\right)$.

The family of multivariate normal random variables is 'closed' under affine shifts: let B be a $p \times k$ matrix whose row space has rank p, for $p \leq k$, and let $\boldsymbol{a}$ be a p-vector. Then $\mathsf{B}\boldsymbol{x} \sim \mathcal{N}\left(\mathsf{B}\boldsymbol{\mu} + \boldsymbol{a}, \mathsf{B}\Sigma\mathsf{B}^{\top}\right)$.

The density of the normal distribution can be expressed on a matrix parameter that combines the first two moments. [121] Suppose $\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \Sigma\right)$. Define the matrix Θ as

$$\Theta =_{\mathrm{df}} \left[\begin{array}{cc} 1 & \boldsymbol{\mu}^{\top} \\ \boldsymbol{\mu} & \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^{\top} \end{array}\right]. \tag{1.3}$$

Then the density of $\boldsymbol{x}$ is

$$\phi\left(\boldsymbol{x}\right) = \frac{e^{\frac{1}{2}}}{\sqrt{\left(2\pi\right)^{k}\left|\Sigma\right|}} \exp\left(-\frac{1}{2}\operatorname{tr}\left(\Theta^{-1}\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\top}\right)\right). \tag{1.4}$$

1.3.2. Elliptical Distributions

We can generalize the multivariate normal distribution to a fatter tailed distribution by considering elliptical distributions. The idea is that the norm of the vector-valued random variable is determined separately from its direction, and the direction follows a simple covariance-like structure. One simple way to define an elliptical distribution is to let $\boldsymbol{z}$ be a n-dimensional normal distribution with mean $\mathbf{0}$ and covariance I. Then let

$$\boldsymbol{x} = \boldsymbol{\mu} + a\Lambda^{1/2}\frac{\boldsymbol{z}}{\left\|z\right\|_2},$$

where a is some random variable, $\boldsymbol{\mu}$ is the mean of $\boldsymbol{x}$, and Λ is related to the covariance of $\boldsymbol{x}$. When a is takes a Chi distribution (not Chi-squared) with n degrees of freedom, we recover the multivariate normal.

Moments of products of elements of $\boldsymbol{x}$ are given, for the normal case, by *Isserlis' Theorem*; an extension of this theorem to elliptical distributions gives the following moment relations. [74, 163] By this theorem, we have

$$\mathrm{E}\left[x_i x_j\right] = \mu_i \mu_j + \frac{\mathrm{E}\left[a^2\right]}{n}\lambda_{i,j}, \tag{1.5}$$

and thus the covariance of $\boldsymbol{x}$ is

$$\Sigma = \frac{\mathrm{E}\left[a^2\right]}{n}\Lambda.$$

The third moment is

$$\mathrm{E}\left[x_i x_j x_k\right] = \mu_i \mu_j \mu_k + \mu_i \Sigma_{j,k} + \mu_j \Sigma_{i,k} + \mu_k \Sigma_{i,l}. \tag{1.6}$$

When considering a centered version of $\boldsymbol{x}$, we can treat $\boldsymbol{\mu}$ as the zero vector, in which case the third moment is zero. Elliptically distributed variables have no skew, which may make them a poor choice for modeling asset returns[6].

The fourth moment is

$$\begin{aligned}\mathrm{E}\left[x_i x_j x_k x_l\right] = \mu_i \mu_j \mu_k \mu_l &+ \frac{n}{n+2}\frac{\mathrm{E}\left[a^4\right]}{\mathrm{E}\left[a^2\right]^2}\left(\Sigma_{i,j}\Sigma_{k,l} + \Sigma_{i,k}\Sigma_{j,l} + \Sigma_{i,l}\Sigma_{j,k}\right)\\ &+ \mu_i \mu_j \Sigma_{k,l} + \mu_i \mu_k \Sigma_{j,l} + \mu_i \mu_l \Sigma_{j,k}\\ &+ \mu_j \mu_k \Sigma_{i,l} + \mu_j \mu_l \Sigma_{i,k} + \mu_k \mu_l \Sigma_{i,j}.\end{aligned} \tag{1.7}$$

The kurtosis (not the *excess* kurtosis) of the i^{th} element of $\boldsymbol{x}$ is then

$$\frac{3n}{n+2}\frac{\mathrm{E}\left[a^4\right]}{\mathrm{E}\left[a^2\right]^2}.$$

Often we need only consider the first four moments of an elliptical distribution. In this case, we can think of the elliptical distribution as parametrized by $\boldsymbol{\mu}$, Σ, and the *kurtosis factor* defined as

$$\kappa =_{\mathrm{df}} \frac{n}{n+2}\frac{\mathrm{E}\left[a^4\right]}{\mathrm{E}\left[a^2\right]^2}. \tag{1.8}$$

The multivariate normal is an elliptical distribution with $\kappa = 1$. Distributions with larger kurtosis factor are considered 'fat tailed.'

An alternative characterization is as follows: random vector $\boldsymbol{x}$ has an elliptical distribution if and only if the density of $\boldsymbol{x}$ is some function of the quadratic form $\left(\boldsymbol{x} - \boldsymbol{\mu}\right)^\top \Lambda^{-1}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)$.

[6] Although slightly better than Gaussian, which have too modest kurtosis.

1.3.3. The Wishart Distribution

Let the rows of the $n \times p$ matrix X be *i.i.d.* multivariate normals with zero mean and covariance $\mathsf{\Sigma}$. Then $\mathsf{W} = \mathsf{X}^{\top}\mathsf{X}$ follows a Wishart distribution with parameter $\mathsf{\Sigma}$ and n degrees of freedom, written as $\mathsf{W} \sim \mathcal{W}\left(\mathsf{\Sigma}, n\right)$, or sometimes as $\mathsf{W} \sim \mathcal{W}_p\left(\mathsf{\Sigma}, n\right)$ to emphasize that W is $p \times p$. [138, 5, 159] The Wishart distribution generalizes the chi-square distribution, or rather the gamma distribution, of which the chi-square is an example.

The density of the Wishart is

$$f_{\mathcal{W}}\left(\mathsf{W}; \mathsf{\Sigma}, n\right) = \frac{\left|\mathsf{W}\right|^{n-p-1/2}}{2^{np/2}\left|\mathsf{\Sigma}\right|^{n/2}\Gamma\left(n/2\right)} e^{-\frac{1}{2}\operatorname{tr}\left(\mathsf{\Sigma}^{-1}\mathsf{W}\right)}. \tag{1.9}$$

This density is defined for non-integral n, a case not covered by the stochastic representation above.

The Wishart enjoys a 'closure property', that is, informally, a projection of a Wishart is also a Wishart. If A is a $p \times k$ matrix of rank k where $k \leq p$, then

$$\mathsf{W} \sim \mathcal{W}_p\left(\mathsf{\Sigma}, n\right) \Rightarrow \mathsf{A}^{\top}\mathsf{W}\mathsf{A} \sim \mathcal{W}_k\left(\mathsf{A}^{\top}\mathsf{\Sigma}\mathsf{A}, n\right). \tag{1.10}$$

The Non-Central Wishart Distribution

As with typical non-central distributions, the non-central Wishart distribution arises when non-centered variates are used in place of centered variates. So imagine that the rows of the $n \times p$ matrix X are *i.i.d.* multivariate normals with mean $\boldsymbol{\mu}$ and covariance $\mathsf{\Sigma}$. Then $\mathsf{W} = \mathsf{X}^{\top}\mathsf{X}$ follows a non-central Wishart distribution with parameter $\mathsf{\Sigma}$, n degrees of freedom, and non-centrality paramater $\boldsymbol{\mu}$, written as $\mathsf{W} \sim \mathcal{W}_p\left(\mathsf{\Sigma}, \boldsymbol{\mu}, n\right)$. [95] The non-central Wishart also satisfies a closure property.

1.3.4. The Inverse Wishart Distribution

Let $\mathsf{Y}^{-1} \sim \mathcal{W}\left(\mathsf{\Psi}^{-1}, n\right)$. Then Y follows an Inverse Wishart Distribution with parameter $\mathsf{\Psi}$ and n degrees of freedom, written as $\mathsf{Y} \sim \mathcal{IW}\left(\mathsf{\Psi}, n\right)$.

The density of the inverse Wishart is

$$f_{\mathcal{IW}}\left(\mathsf{Y}; \mathsf{\Psi}, n\right) = \frac{\left|\mathsf{Y}\right|^{n+p+1/2}\left|\mathsf{\Psi}\right|^{n/2}}{2^{np/2}\Gamma\left(n/2\right)} e^{-\frac{1}{2}\operatorname{tr}\left(\mathsf{\Psi}\mathsf{Y}\right)}. \tag{1.11}$$

1.3.5. The Multivariate t Distribution

The *Multivariate t distribution* generalizes the t-distribution to the multivariate setting. [81] There are two equivalent stochastic representations of this distribution. Suppose that

$$\begin{aligned} Y &\sim \mathcal{N}\left(\mathbf{0}_p, \mathsf{\Sigma}\right), \\ S^2 &\sim \chi^2\left(n\right), \end{aligned}$$

with Y and S^2 independent. Then $X = \frac{Y}{\sqrt{S^2/n}} + \boldsymbol{\mu}$ follows a multivariate t distribution with n degrees of freedom, matrix Σ, and location parameter $\boldsymbol{\mu}$, written as

$$X \sim \mathcal{T}\left(n, \Sigma, \boldsymbol{\mu}\right). \tag{1.12}$$

X has the density [37]

$$f_{\mathcal{T}}\left(X; n, \Sigma, \boldsymbol{\mu}\right) = \frac{\Gamma\left(\frac{1}{2}\left(n+p\right)\right) n^{n/2}}{\Gamma\left(\frac{1}{2}n\right) \pi^{p/2} \left|\Sigma\right|^{\frac{1}{2}}} \left(n + \left(X - \boldsymbol{\mu}\right)^{\top} \Sigma^{-1} \left(X - \boldsymbol{\mu}\right)\right)^{-n+p/2}. \tag{1.13}$$

The multivariate t distribution admits an alternative stochastic representation. [37, 6, 81] Let

$$\begin{aligned} Y &\sim \mathcal{N}\left(\mathbf{0}_p, n\mathsf{I}_p\right), \\ U &\sim \mathcal{W}\left(\Sigma^{-1}, n+p-1\right), \end{aligned}$$

with Y and U independent. Then

$$X = \left(U^{1/2}\right)^{-1} Y + \boldsymbol{\mu} \sim \mathcal{T}\left(n, \Sigma, \boldsymbol{\mu}\right),$$

where here $\mathsf{A}^{1/2}$ represents the symmetric square root of A. Another way of stating this is in terms of an inverse Wishart, as follows:

$$\begin{aligned} W &\sim \mathcal{IW}\left(\Sigma, n+p-1\right), \\ X\left|W\right. &\sim \mathcal{N}\left(\boldsymbol{\mu}, nW\right). \end{aligned} \tag{1.14}$$

This form will appear in the context of posterior marginal distributions in Bayesian analysis.

Kotz and Nadarajah claim that the distribution is “said to be central if $\boldsymbol{\mu} = \mathbf{0}$; otherwise, it is said to be noncentral.” [81, page 1] It is not clear to whom they refer in the third person invisible, but the $\boldsymbol{\mu} \neq \mathbf{0}_p$ case should emphatically *not* be called a ‘non-central’ multivariate t, since it does *not* reduce to the scalar non-central t in the $p = 1$ case. Instead one should consider the $\boldsymbol{\mu}$ as a locational shift. The distribution of X is spherically symmetric around $\boldsymbol{\mu}$, a property not shared by the scalar non-central t.

The Multivariate t satisfies a multiplicative closure property. If A is a $p \times k$ matrix of rank k where $k \leq p$, then

$$X \sim \mathcal{T}\left(n, \Sigma, \boldsymbol{\mu}\right) \Rightarrow \mathsf{A}^{\top} X \sim \mathcal{T}\left(n, \mathsf{A}^{\top}\Sigma\mathsf{A}, \mathsf{A}^{\top}\boldsymbol{\mu}\right). \tag{1.15}$$

As a consequence, marginals of the Multivariate t distribution follow the (scaled, shifted) scalar t distribution.

The Multivariate t distribution is an elliptical distribution, with mean $\boldsymbol{\mu}$, covariance $\frac{n}{n-2}\Sigma$ and kurtosis factor $\kappa = \frac{n-2}{n-4}$.

1.3.6. The Non-central Multivariate t Distribution

Kshirsagar described a truly non-central multivariate t distribution that generalizes the scalar non-central t distribution in the $p = 1$ case. Let

$$\begin{aligned} Y &\sim \mathcal{N}\left(\boldsymbol{\mu}, \sigma^2 R\right), \\ S^2 &\sim \chi^2\left(n\right), \end{aligned}$$

where R is a correlation matrix, *i.e.*, symmetric positive definite with ones along the diagonal. Then

$$X = \frac{Y}{\sqrt{\sigma^2 S^2/n}}$$

follows the non-central multivariate t distribution with covariance $\Sigma = \sigma^2 R$, degrees of freedom n and non-centrality parameter $\boldsymbol{\mu}$. Kshirsagar gives the density of this distribution and notes that its marginals follow the scalar non-central t distribution. [81, 85] This distribution is *not* an elliptical distribution.

1.4. † Matrix Derivatives

In the more advanced parts of this text, we will need to compute the derivatives of matrices and the derivatives of quantities with respect to matrices. This is somewhat complicated by considerations of notation (the derivative of an $m \times n$ matrix with respect to a $p \times q$ matrix consists of $mnpq$ distinct elements) as well as symmetry (*e.g.*, we typically want to consider the derivative of a quantity with respect to a symmetric matrix, only considering changes which respect symmetry). For more details Magnus and Neudecker give a very good introduction. [105, 106] The Matrix Cookbook provides a good cheat-sheet of results. [134]

Definition 1.4.1 (Derivatives)**.** For m-vector $\boldsymbol{x}$, and n-vector $\boldsymbol{y}$, let the derivative $\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}\boldsymbol{x}}$ be the $n \times m$ matrix whose first column is the partial derivative of $\boldsymbol{y}$ with respect to x_1. This follows the so-called 'numerator layout' convention. For matrices Y and X, define

$$\frac{\mathrm{d}\mathsf{Y}}{\mathrm{d}\mathsf{X}} =_{\mathrm{df}} \frac{\mathrm{d}\operatorname{vec}\left(\mathsf{Y}\right)}{\mathrm{d}\operatorname{vec}\left(\mathsf{X}\right)}.$$

Lemma 1.4.2 (Miscellaneous Derivatives)**.** *For symmetric matrices* Y *and* X*,*

$$\frac{\mathrm{d}\operatorname{vech}\left(\mathsf{Y}\right)}{\mathrm{d}\operatorname{vec}\left(\mathsf{X}\right)} = \mathsf{L}\frac{\mathrm{d}\mathsf{Y}}{\mathrm{d}\mathsf{X}}, \quad \frac{\mathrm{d}\operatorname{vec}\left(\mathsf{Y}\right)}{\mathrm{d}\operatorname{vech}\left(\mathsf{X}\right)} = \frac{\mathrm{d}\mathsf{Y}}{\mathrm{d}\mathsf{X}}\mathsf{D}, \quad \frac{\mathrm{d}\operatorname{vech}\left(\mathsf{Y}\right)}{\mathrm{d}\operatorname{vech}\left(\mathsf{X}\right)} = \mathsf{L}\frac{\mathrm{d}\mathsf{Y}}{\mathrm{d}\mathsf{X}}\mathsf{D}. \tag{1.16}$$

Proof. For the first equation, note that $\operatorname{vech}\left(\mathsf{Y}\right) = \mathsf{L}\operatorname{vec}\left(\mathsf{Y}\right)$, thus by the chain rule:

$$\frac{\mathrm{d}\operatorname{vech}\left(\mathsf{Y}\right)}{\mathrm{d}\operatorname{vec}\left(\mathsf{X}\right)} = \frac{\mathrm{d}\mathsf{L}\operatorname{vec}\left(\mathsf{Y}\right)}{\mathrm{d}\operatorname{vec}\left(\mathsf{Y}\right)} = \mathsf{L}\frac{\mathrm{d}\mathsf{Y}}{\mathrm{d}\mathsf{X}},$$

by linearity of the derivative. The other identities follow similarly. ⊣

Lemma 1.4.3 (Derivative of matrix inverse)**.** *For invertible matrix* A*,*

$$\frac{\mathrm{d}\mathsf{A}^{-1}}{\mathrm{d}\mathsf{A}} = -\left(\mathsf{A}^{-\top} \otimes \mathsf{A}^{-1}\right) = -\left(\mathsf{A}^{\top} \otimes \mathsf{A}\right)^{-1}. \tag{1.17}$$

For symmetric A*, the derivative with respect to the non-redundant part is*

$$\frac{\mathrm{dvech}\left(\mathsf{A}^{-1}\right)}{\mathrm{d}\,\mathrm{vech}\left(\mathsf{A}\right)} = -\mathsf{L}\left(\mathsf{A}^{-1} \otimes \mathsf{A}^{-1}\right)\mathsf{D}. \tag{1.18}$$

Note how this result generalizes the scalar derivative: $\frac{\mathrm{d}x^{-1}}{\mathrm{d}x} = -\left(x^{-1}x^{-1}\right)$.

Proof. Equation 1.17 is a known result. [42, 106] Equation 1.18 then follows using Lemma 1.4.2. ⊣

Lemma 1.4.4 (Miscellaneous Derivatives)**.** *Given conformable, symmetric, matrices* X*,* Y*,* Z*, and constant matrix* J*, define*

$$f_{\mathcal{P}}\left(\mathsf{X};\mathsf{J}\right) =_{\mathrm{df}} -\left(\left(\mathsf{J}^{\top}\mathsf{X}\mathsf{J}\right)^{-1} \otimes \left(\mathsf{J}^{\top}\mathsf{X}\mathsf{J}\right)^{-1}\right)\left(\mathsf{J}^{\top} \otimes \mathsf{J}^{\top}\right).$$

Then

$$\frac{\mathrm{d}\left(\mathsf{J}^{\top}\mathsf{X}\mathsf{J}\right)^{-1}}{\mathrm{d}\mathsf{Z}} = f_{\mathcal{P}}\left(\mathsf{X};\mathsf{J}\right)\frac{\mathrm{d}\mathsf{X}}{\mathrm{d}\mathsf{Z}}. \tag{1.19}$$

$$\frac{\mathrm{d}\mathsf{X}\mathsf{Y}}{\mathrm{d}\mathsf{Z}} = \left(\mathsf{I} \otimes \mathsf{X}\right)\frac{\mathrm{d}\mathsf{Y}}{\mathrm{d}\mathsf{Z}} + \left(\mathsf{Y}^{\top} \otimes \mathsf{I}\right)\frac{\mathrm{d}\mathsf{X}}{\mathrm{d}\mathsf{Z}}. \tag{1.20}$$

$$\frac{\mathrm{d}\mathsf{X}\mathsf{X}^{\top}}{\mathrm{d}\mathsf{Z}} = \left(\mathsf{I} + \mathsf{K}\right)\left(\mathsf{X} \otimes \mathsf{I}\right)\frac{\mathrm{d}\mathsf{X}}{\mathrm{d}\mathsf{Z}}. \tag{1.21}$$

$$\frac{\mathrm{dtr}\left(\mathsf{X}\mathsf{Y}\right)}{\mathrm{d}\mathsf{Z}} = \mathrm{vec}\left(\mathsf{X}^{\top}\right)^{\top}\frac{\mathrm{d}\mathsf{Y}}{\mathrm{d}\mathsf{Z}} + \mathrm{vec}\left(\mathsf{Y}\right)^{\top}\frac{\mathrm{d}\mathsf{X}^{\top}}{\mathrm{d}\mathsf{Z}}. \tag{1.22}$$

$$\frac{\mathrm{d}\left|\mathsf{X}\right|}{\mathrm{d}\mathsf{Z}} = \left|\mathsf{X}\right|\mathrm{vec}\left(\mathsf{X}^{-\top}\right)^{\top}\frac{\mathrm{d}\mathsf{X}}{\mathrm{d}\mathsf{Z}}. \tag{1.23}$$

$$\frac{\mathrm{d}\left|\mathsf{X}\mathsf{Y}\right|}{\mathrm{d}\mathsf{Z}} = \left|\mathsf{X}\mathsf{Y}\right|\left(\mathrm{vec}\left(\mathsf{X}^{-\top}\right)^{\top}\frac{\mathrm{d}\mathsf{X}}{\mathrm{d}\mathsf{Z}} + \mathrm{vec}\left(\mathsf{Y}^{-\top}\right)^{\top}\frac{\mathrm{d}\mathsf{Y}}{\mathrm{d}\mathsf{Z}}\right). \tag{1.24}$$

$$\frac{\mathrm{d}\left|\left(\mathsf{X}\mathsf{Y}\right)^{-1}\right|}{\mathrm{d}\mathsf{Z}} = \left|\mathsf{X}\mathsf{Y}\right|^{-1}\left(\mathrm{vec}\left(\mathsf{X}^{-\top}\right)^{\top}\frac{\mathrm{d}\mathsf{X}}{\mathrm{d}\mathsf{Z}} + \mathrm{vec}\left(\mathsf{Y}^{-\top}\right)^{\top}\frac{\mathrm{d}\mathsf{Y}}{\mathrm{d}\mathsf{Z}}\right). \tag{1.25}$$

Here K *is the 'commutation matrix.'*

Let λ_j *be the* j^{th} *eigenvalue of* X*, with corresponding eigenvector* $\boldsymbol{\nu}_j$*, normalized so that* ${\boldsymbol{\nu}_j}^{\top}\boldsymbol{\nu}_j = 1$*. Then*

$$\frac{\mathrm{d}\lambda_j}{\mathrm{d}\mathsf{Z}} = \left({\boldsymbol{\nu}_j}^{\top} \otimes {\boldsymbol{\nu}_j}^{\top}\right)\frac{\mathrm{d}\mathsf{X}}{\mathrm{d}\mathsf{Z}}. \tag{1.26}$$

Proof. For Equation 1.19, write

$$\frac{\mathrm{d}\left(\mathsf{J}^\top\mathsf{X}\mathsf{J}\right)^{-1}}{\mathrm{d}\mathsf{Z}} = \frac{\mathrm{d}\left(\mathsf{J}^\top\mathsf{X}\mathsf{J}\right)^{-1}}{\mathrm{d}\left(\mathsf{J}^\top\mathsf{X}\mathsf{J}\right)}\frac{\mathrm{d}\left(\mathsf{J}^\top\mathsf{X}\mathsf{J}\right)}{\mathrm{d}\mathsf{Z}}.$$

Lemma 1.4.3 gives the derivative on the left; to get the derivative on the right, note that $\mathrm{vec}\left(\mathsf{J}^\top\mathsf{X}\mathsf{J}\right) = \left(\mathsf{J}^\top \otimes \mathsf{J}^\top\right)\mathrm{vec}\left(\mathsf{X}\right)$, then use linearity of the derivative.

For Equation 1.20, write $\mathrm{vec}\left(\mathsf{X}\mathsf{Y}\right) = \left(\mathsf{Y}^\top \otimes \mathsf{X}\right)\mathrm{vec}\left(\mathsf{I}\right)$. Then consider the derivative of $\mathrm{vec}\left(\mathsf{X}\mathsf{Y}\right)$ with respect to any *scalar* z :

$$\frac{\mathrm{dvec}\left(\mathsf{X}\mathsf{Y}\right)}{\mathrm{d}z} = \frac{\mathrm{d}\left(\mathsf{Y}^\top \otimes \mathsf{X}\right)\mathrm{vec}\left(\mathsf{I}\right)}{\mathrm{d}z} = \mathrm{vec}^{-1}\left(\frac{\mathrm{d}\left(\mathsf{Y}^\top \otimes \mathsf{X}\right)}{\mathrm{d}z}\right)\mathrm{vec}\left(\mathsf{I}\right),$$

where $\mathrm{vec}^{-1}\left(\cdot\right)$ is the inverse of $\mathrm{vec}\left(\cdot\right)$. That is, $\mathrm{vec}^{-1}\left(\mathrm{vec}\left(\cdot\right)\right)$ is the identity over square matrices. (This wrinkle is needed because we have defined derivatives of matrices to be the derivative of their vectorization.)

For Equation 1.21, by Equation 1.20,

$$\begin{aligned}\frac{\mathrm{dvec}\left(\mathsf{X}\mathsf{X}^\top\right)}{\mathrm{d}\,\mathrm{vec}\left(\mathsf{Z}\right)} &= \left(\mathsf{I}\otimes\mathsf{X}\right)\frac{\mathrm{dvec}\left(\mathsf{X}^\top\right)}{\mathrm{d}\,\mathrm{vec}\left(\mathsf{Z}\right)} + \left(\mathsf{X}\otimes\mathsf{I}\right)\frac{\mathrm{dvec}\left(\mathsf{X}\right)}{\mathrm{d}\,\mathrm{vec}\left(\mathsf{Z}\right)},\\ &= \left(\mathsf{I}\otimes\mathsf{X}\right)\mathsf{K}\frac{\mathrm{dvec}\left(\mathsf{X}\right)}{\mathrm{d}\,\mathrm{vec}\left(\mathsf{Z}\right)} + \left(\mathsf{X}\otimes\mathsf{I}\right)\frac{\mathrm{dvec}\left(\mathsf{X}\right)}{\mathrm{d}\,\mathrm{vec}\left(\mathsf{Z}\right)}.\end{aligned}$$

Now let A be any conformable square matrix. We have:

$$\begin{aligned}&\left(\mathsf{I}\otimes\mathsf{X}\right)\mathsf{K}\,\mathrm{vec}\left(\mathsf{A}\right) = \left(\mathsf{I}\otimes\mathsf{X}\right)\mathrm{vec}\left(\mathsf{A}^\top\right) = \mathrm{vec}\left(\mathsf{X}\mathsf{A}^\top\right) = \\ &\qquad\qquad \mathsf{K}\,\mathrm{vec}\left(\mathsf{A}\mathsf{X}^\top\right) = \mathsf{K}\left(\mathsf{X}\otimes\mathsf{I}\right)\mathrm{vec}\left(\mathsf{A}\right).\end{aligned}$$

Because A was arbitrary, we have $\left(\mathsf{I}\otimes\mathsf{X}\right)\mathsf{K} = \mathsf{K}\left(\mathsf{X}\otimes\mathsf{I}\right)$, and the result follows.

Using the product rule for Kronecker products [134], then using the vector identity again we have

$$\begin{aligned}\frac{\mathrm{dvec}\left(\mathsf{X}\mathsf{Y}\right)}{\mathrm{d}z} &= \left(\mathrm{vec}^{-1}\left(\frac{\mathrm{d}\mathsf{Y}^\top}{\mathrm{d}z}\right)\otimes\mathsf{X} + \mathsf{Y}^\top\otimes\mathrm{vec}^{-1}\left(\frac{\mathrm{d}\mathsf{X}}{\mathrm{d}z}\right)\right)\mathrm{vec}\left(\mathsf{I}\right),\\ &= \left(\mathsf{I}\otimes\mathsf{X}\right)\frac{\mathrm{d}\mathsf{Y}}{\mathrm{d}z} + \left(\mathsf{Y}^\top\otimes\mathsf{I}\right)\frac{\mathrm{d}\mathsf{X}}{\mathrm{d}z}.\end{aligned}$$

Then apply this result to every element of $\mathrm{vec}\left(\mathsf{Z}\right)$ to get the result.

For Equation 1.22, write

$$\mathrm{tr}\left(\mathsf{X}\mathsf{Y}\right) = \mathrm{vec}\left(\mathsf{X}^\top\right)^\top\mathrm{vec}\left(\mathsf{Y}\right),$$

then use the product rule.

For Equation 1.23, first consider the derivative of $\left|\mathsf{X}\right|$ with respect to a scalar z. This is known to take form: [134]

$$\frac{\mathrm{d}\left|\mathsf{X}\right|}{\mathrm{d}z} = \left|X\right|\mathrm{tr}\left(\mathsf{X}^{-1}\mathrm{vec}^{-1}\left(\frac{\mathrm{d}\mathsf{X}}{\mathrm{d}z}\right)\right),$$

where the $\mathrm{vec}^{-1}(\cdot)$ is here because of how we have defined derivatives of matrices. Rewrite the trace as the dot product of two vectors:

$$\frac{\mathrm{d}|\mathsf{X}|}{\mathrm{d}z} = |X| \operatorname{vec}\left(\mathsf{X}^{-\top}\right)^\top \frac{\mathrm{d}\mathsf{X}}{\mathrm{d}z}.$$

Using this to compute the derivative with respect to each element of $\operatorname{vec}(\mathsf{Z})$ gives the result. Equation 1.24 follows from the scalar product rule since $|\mathsf{XY}| = |\mathsf{X}|\,|\mathsf{Y}|$. Equation 1.25 then follows, using the scalar chain rule.

For Equation 1.26, the derivative of the j^{th} eigenvalue of matrix X with respect to a scalar z is known to be: [134, equation (67)]

$$\frac{\mathrm{d}\lambda_j}{\mathrm{d}z} = \boldsymbol{\nu}_j{}^\top \mathrm{vec}^{-1}\left(\frac{\mathrm{d}\mathsf{X}}{\mathrm{d}z}\right)\boldsymbol{\nu}_j.$$

Take the vectorization of this scalar, and rewrite it in Kronecker form:

$$\frac{\mathrm{d}\lambda_j}{\mathrm{d}z} = \left(\boldsymbol{\nu}_j{}^\top \otimes \boldsymbol{\nu}_j{}^\top\right)\frac{\mathrm{d}\mathsf{X}}{\mathrm{d}z}.$$

Use this to compute the derivative of λ_j with respect to elements of $\operatorname{vec}(\mathsf{Z})$. ⊣

Lemma 1.4.5 (Cholesky Derivatives)**.** *Let X be a symmetric positive definite matrix. Let Y be its lower triangular Cholesky factor. That is, Y is the lower triangular matrix such that $\mathsf{Y}\mathsf{Y}^\top = \mathsf{X}$. Then*

$$\frac{\mathrm{d}\operatorname{vech}(\mathsf{Y})}{\mathrm{d}\operatorname{vech}(\mathsf{X})} = \left(\mathsf{L}\left(\mathsf{I}+\mathsf{K}\right)\left(\mathsf{Y}\otimes\mathsf{I}\right)\mathsf{L}^\top\right)^{-1}, \tag{1.27}$$

where K is the 'commutation matrix.'

Proof. By Equation 1.21 of Lemma 1.4.4,

$$\frac{\mathrm{d}\operatorname{vec}\left(\mathsf{Y}\mathsf{Y}^\top\right)}{\mathrm{d}\operatorname{vec}(\mathsf{Y})} = \left(\mathsf{I}+\mathsf{K}\right)\left(\mathsf{Y}\otimes\mathsf{I}\right).$$

By the chain rule, for lower triangular matrix Y, we have

$$\begin{aligned}\frac{\mathrm{d}\operatorname{vech}\left(\mathsf{Y}\mathsf{Y}^\top\right)}{\mathrm{d}\operatorname{vech}(\mathsf{Y})} &= \frac{\mathrm{d}\operatorname{vech}\left(\mathsf{Y}\mathsf{Y}^\top\right)}{\mathrm{d}\operatorname{vec}\left(\mathsf{Y}\mathsf{Y}^\top\right)}\frac{\mathrm{d}\operatorname{vec}\left(\mathsf{Y}\mathsf{Y}^\top\right)}{\mathrm{d}\operatorname{vec}(\mathsf{Y})}\frac{\mathrm{d}\operatorname{vec}(\mathsf{Y})}{\mathrm{d}\operatorname{vech}(\mathsf{Y})},\\ &= \mathsf{L}\left(\mathsf{I}+\mathsf{K}\right)\left(\mathsf{Y}\otimes\mathsf{I}\right)\mathsf{L}^\top.\end{aligned}$$

The result now follows since

$$\frac{\mathrm{d}\operatorname{vech}(\mathsf{Y})}{\mathrm{d}\operatorname{vech}(\mathsf{X})} = \frac{\mathrm{d}\operatorname{vech}(\mathsf{Y})}{\mathrm{d}\operatorname{vech}\left(\mathsf{Y}\mathsf{Y}^\top\right)} = \left(\frac{\mathrm{d}\operatorname{vech}\left(\mathsf{Y}\mathsf{Y}^\top\right)}{\mathrm{d}\operatorname{vech}(\mathsf{Y})}\right)^{-1}.$$

⊣

Lemma 1.4.6 (Inverse Cholesky Derivatives)**.** *Let* X *be a symmetric positive definite matrix. Let* Y *be the lower triangular Cholesky factor of the* inverse *of* X*. That is,* Y *is the lower triangular matrix such that* $\mathsf{Y}\mathsf{Y}^\top = \mathsf{X}^{-1}$*, sometimes written as* $\mathsf{Y} = \mathsf{X}^{-1/2}$*. Then*

$$\frac{\mathrm{dvech}\left(\mathsf{Y}\right)}{\mathrm{d}\,\mathrm{vech}\left(\mathsf{X}\right)} = -\left(\mathsf{L}\left(\mathsf{I}+\mathsf{K}\right)\left(\left(\mathsf{Y}^\top\right)^{-1} \otimes \left(\mathsf{Y}\mathsf{Y}^\top\right)^{-1}\right)\mathsf{L}^\top\right)^{-1}. \tag{1.28}$$

Proof. By the chain rule,

$$\frac{\mathrm{dvech}\left(\left(\mathsf{Y}\mathsf{Y}^\top\right)^{-1}\right)}{\mathrm{d}\,\mathrm{vech}\left(\mathsf{Y}\right)} = \frac{\mathrm{dvech}\left(\left(\mathsf{Y}\mathsf{Y}^\top\right)^{-1}\right)}{\mathrm{d}\,\mathrm{vech}\left(\mathsf{Y}\mathsf{Y}^\top\right)}\frac{\mathrm{dvech}\left(\mathsf{Y}\mathsf{Y}^\top\right)}{\mathrm{d}\,\mathrm{vech}\left(\mathsf{Y}\right)}.$$

The term on the left is given by Lemma 1.4.3. The one on the right was derived in the proof of Lemma 1.4.5. Together they give us

$$\frac{\mathrm{dvech}\left(\left(\mathsf{Y}\mathsf{Y}^\top\right)^{-1}\right)}{\mathrm{d}\,\mathrm{vech}\left(\mathsf{Y}\right)} = \left(-\mathsf{L}\left(\left(\mathsf{Y}\mathsf{Y}^\top\right)^{-1} \otimes \left(\mathsf{Y}\mathsf{Y}^\top\right)^{-1}\right)\mathsf{D}\right)\left(\mathsf{L}\left(\mathsf{I}+\mathsf{K}\right)\left(\mathsf{Y}\otimes\mathsf{I}\right)\mathsf{L}^\top\right).$$

Now note that $\mathsf{I}+\mathsf{K}$ is a special matrix which symmetrizes quantities to the right of it. (One half of this quantity is what Magnus calls the 'symmetric idempotent matrix.' [105]) Then note that $\mathsf{D}\mathsf{L}$ is idempotent on symmetric matrices. Thus $\mathsf{D}\mathsf{L}\left(\mathsf{I}+\mathsf{K}\right) = \mathsf{I}+\mathsf{K}$, and we can eliminate the $\mathsf{D}\mathsf{L}$ from our expression. From this it follows, via implicit differentiation, that

$$\frac{\mathrm{dvech}\left(\mathsf{Y}\right)}{\mathrm{d}\,\mathrm{vech}\left(\mathsf{X}\right)} = -\left(\mathsf{L}\left(\left(\mathsf{Y}\mathsf{Y}^\top\right)^{-1} \otimes \left(\mathsf{Y}\mathsf{Y}^\top\right)^{-1}\right)\left(\left(\mathsf{I}+\mathsf{K}\right)\left(\mathsf{Y}\otimes\mathsf{I}\right)\mathsf{L}^\top\right)\right)^{-1}.$$

Now note that the matrix $\mathsf{I}+\mathsf{K}$ (equal to twice the symmetric idempotent matrix N_n) can be placed on either side of expressions of the form $\mathsf{A}\otimes\mathsf{A}$. [105, eqn. (40)] Thus it can be shifted to the left of the equation. We then use the fact that matrix product commutes with Kronecker product to arrive at

$$\frac{\mathrm{dvech}\left(\mathsf{Y}\right)}{\mathrm{d}\,\mathrm{vech}\left(\mathsf{X}\right)} = -\left(\mathsf{L}\left(\mathsf{I}+\mathsf{K}\right)\left(\left(\left(\mathsf{Y}\mathsf{Y}^\top\right)^{-1}\mathsf{Y}\right) \otimes \left(\mathsf{Y}\mathsf{Y}^\top\right)^{-1}\right)\mathsf{L}^\top\right)^{-1}. \tag{1.29}$$

From this the result follows. ⊣

Note how this result generalizes the scalar derivative: $\frac{\mathrm{d}x^{-1/2}}{\mathrm{d}x} = -\left(2x^{1/2}x\right)^{-1}$.

Exercises

* **Ex. 1.1 Note on exercises** Exercises which are more difficult than others will typically be marked with a star, as above. Some questions may be marked with they symbol "§" to denote that the author is not sure that there is a nice answer.

1. Some exercises might be labelled as "boring". The point of these is not to bore you, rather to ensure that you have understood some computational recipe well enough to reproduce on your own.

Ex. 1.2 Trace Let A be a square matrix.

1. Show that if A is diagonal, then $\operatorname{tr}(\mathsf{A}) = \mathbf{1}^\top \mathsf{A}\mathbf{1}$.
2. Show that $\operatorname{tr}\left(\mathsf{A}^2\right) = \operatorname{vec}(\mathsf{A})^\top \operatorname{vec}(\mathsf{A})$ for symmetric matrix A.

Ex. 1.3 Elimination and Duplication Write out the 3×4 Elimination matrix. Write out the 4×3 Duplication matrix.

Ex. 1.4 Matrix Calculus Exercises Assume X is symmetric.

1. What is $\frac{\mathrm{d}\operatorname{tr}(\mathsf{X})}{\mathrm{d}\mathsf{X}}$?
2. Compute $\frac{\mathrm{d}\operatorname{tr}\left(\mathsf{X}^2\right)}{\mathrm{d}\mathsf{X}}$.

Ex. 1.5 Inverse Kronecker Suppose A and B are square non-singular matrices. Prove that

$$(\mathsf{A} \otimes \mathsf{B})^{-1} = \mathsf{A}^{-1} \otimes \mathsf{B}^{-1}.$$

Ex. 1.6 Numerical Derivatives Use a differencing scheme to approximate the derivative of matrix-valued operations of matrix input. That is, letting ϵ be a small quantity, compute, for some input matrix X and function f

$$\frac{f\left(\mathsf{X} + \epsilon \mathsf{J}^{ij}\right) - f\left(\mathsf{X}\right)}{\epsilon},$$

vectorize it, and set this equal to the i, j^{th} column of the approximate derivative.

1. Using the numerical differencing scheme, confirm the result of Lemma 1.4.5. Note that some implementations of the Cholesky operation will assume the input is symmetric and thus ignore the lower or upper triangular part of the input. Take care to deal with this in your numerical approximation. (Hint: apply the symmetric idempotent matrix to your input.)
2. Using the numerical differencing scheme, confirm the result of Lemma 1.4.6.

Ex. 1.7 Wishart Facts 1. Prove the closure property of the Wishart distribution, Equation 1.10.

Part I.

The Sharpe ratio

2. The Sharpe ratio and the signal-noise ratio

> A man who seeks advice about his actions will not be grateful for the suggestion that he maximise expected utility.
>
> *(A. D. Roy,* Safety First and the Holding of Assets*)*

> Hey, Paul! Let's get rid of Clarence and steal all his good ideas!
>
> *(John Lennon,* attributed*)*

2.1. Introduction

The Sharpe ratio is arguably the most commonly used metric of the *historical* performance of financial assets–mutual funds, hedge funds, stocks, *etc.* It is defined as

$$\hat{\zeta} =_{\mathrm{df}} \frac{\hat{\mu} - r_0}{\hat{\sigma}}, \tag{2.1}$$

where $\hat{\mu}$ is the historical, or sample, mean return of the mutual fund, $\hat{\sigma}$ is the sample standard deviation of returns, and r_0 is some fixed *risk-free* or *disastrous* rate of return.

Under the original definition of Sharpe's "reward-to-variability ratio", r_0 was equal to zero. [149] One typically uses the vanilla sample mean and the Bessel-corrected sample standard deviation in computing the ratio, *viz.*

$$\hat{\mu} =_{\mathrm{df}} \sum_{1 \le i \le n} x_i, \quad \hat{\sigma} =_{\mathrm{df}} \sqrt{\frac{\sum_{1 \le i \le n} \left(x_i - \hat{\mu}\right)^2}{n-1}}. \tag{2.2}$$

In this text, *Sharpe ratio* will refer to this quantity, computed from sample statistics, whereas *signal-noise ratio* will refer to the analagously defined population parameter,

$$\zeta =_{\mathrm{df}} \frac{\mu - r_0}{\sigma}. \tag{2.3}$$

In general, hats will be placed over quantities to denote population estimates. Sharpe himself notes that ideally population values would be used in the computation of his reward-to-variability ratio, but "Since the predictions cannot be obtained in any satisfactory manner, ... ex post values must be used–the average rate of return of a portfolio must be substituted for its expected rate of return, and the actual standard deviation of its rate of return for its predicted risk." [149, p. 122]

In some situations, the risk-free rate, which acts as a benchmark, varies significantly over the measured period, and this time dependence is incorporated into the computation. In this case, the Sharpe ratio is defined as the ratio of the sample mean to the sample standard deviation of $x_t - r_t$, where x_t is the return of the asset in question, and r_t is the varying risk free rate.

Example 2.1.1 (Sharpe ratio of the Market). Here we consider the raw returns of the Market, introduced in Example 1.2.1. The mean monthly return from Jan 1927 to Dec 2018 is 0.92%; the standard deviation is 5.3%. Thus the Sharpe ratio is $0.17\,\mathrm{mo.}^{-1/2}$. (More on the funny units later.) ⊣

Example 2.1.2 (Sharpe ratio of the Market, excess returns). Continuing Example 2.1.1, if we compute the *excess* returns of 'the Market' by subtracting the risk-free rate from the Market returns, the Sharpe ratio is $0.1211\,\mathrm{mo.}^{-1/2}$. ⊣

In practice, the Sharpe ratio is often used to answer the following kinds of questions:

1. "*Should I invest a predetermined amount of money (long) in a given asset?*"
2. "*Should our quant fund replace strategy A with strategy B?*"
3. "*Should we invest in asset B given that we will keep our investment in asset A?*"

Note that each of these questions is binary: we are not asking *how much* should we invest in a given asset, nor whether we should potentially *short* the asset. Rather, the amount and side are predetermined. Later we will consider portfolio problems, which are less constrained in nature.

We stress again the Sharpe ratio is a *historical* measure, but it is used to guide future investments. But as is often hidden in the fine print, historical performance is no guarantee of future results. In these notes we aim to explain just how much this historical measure can be trusted as an estimate of future results, under certain assumptions.

2.1.1. Which Returns?

Let p be the mark price of an asset. The *relative return* (also known as 'arithmetic', 'percent,' or 'simple' return) from t to $t+1$ is defined as

$$\frac{p_{t+1} - p_t}{p_t} = \frac{p_{t+1}}{p_t} - 1. \tag{2.4}$$

The *log return* (also known as 'geometric' return) is defined as

$$\log \frac{p_{t+1}}{p_t}. \tag{2.5}$$

There is no clear standard which form of returns should be used in computation of the Sharpe ratio. Log returns are typically 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 sensible. For this reason, adjusting the time frame (*e.g.*, annualizing) of log returns is trivial.

However, relative returns have the attractive 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 cash[1] to purchase stock. Under this formulation, the returns aggregate over time by summation just like log returns.

Moreover, log returns are unbounded in the case that there is a non-zero possibility of an investment losing 100%. In such a case the log returns do not have finite first (or higher) moments, and the signal-noise ratio is undefined. While this detail is typically swept under the rug, academics often worry about the assumption of finite *second* moments. In their favor, relative returns are for most, and perhaps all, investments bounded: pick some insanely large number larger than 100, say $10^{10^{10}}$; the return of an investment, in percent, will not exceed this large number, in absolute value, and thus *all* moments of relative returns are finite.

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

$$\frac{\hat{\zeta}_r - \hat{\zeta}}{\hat{\zeta}} \approx \frac{1}{2}\frac{\sum_i x^2}{\sum_i x}, \tag{2.6}$$

where $\hat{\zeta}_r$ is the Sharpe ratio 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. See Exercise 2.16.

It goes without saying that it is generally assumed that prices are measured at equal intervals, and so the returns are 'over' comparable time periods. That is, if x_1 is one (market) day's return, then so is x_2, x_3, and so on. (Though see Example 4.1.3 for how to deal with jumbled periodicities.)

Example 2.1.3 (Sharpe ratio of the Market, log returns). Continuing Example 2.1.1, the returns of the Market were expressed as log returns. The mean monthly return is 0.0078; the standard deviation is 0.0533. Thus the Sharpe ratio is $0.15\,\text{mo.}^{-1/2}$. The boost from computing the Sharpe ratio on relative returns instead of log returns is around 19%. ⊣

Caution. It should be noted, however, that the use of simple returns is considered objectionable in some circles. Because simple returns do not sum as geometric returns do, summing them to estimate long horizon returns is inaccurate. [142] Another way to express this is that the change of time units we will introduce in the next section is misleading, and the Sharpe ratio should only be used on the time horizon over which

[1] At no interest.

returns are measured. Nonetheless, there should be little in this text that is specific to simple returns, and one may apply most or all of the results to the analysis of log returns.

2.2. Units of Sharpe ratio

The Sharpe ratio is often quoted without units. This is a recipe for disaster, since it is emphatically not a unit-less quantity, and often quoted in units different from those which it is measured.

To compute the units of Sharpe ratio, consider the units of μ and σ. Given two successive returns, x_1, x_2, which are assumed independent, let $x_{1:2} = x_1 + x_2$. Note that $x_{1:2}$ is the return of the asset over the two time periods only if the returns are log returns. The expected value of $x_{1:2}$ is 2μ; the variance of $x_{1:2}$ is $2\sigma^2$. Thus the units of μ must be 'return per time,' and the units of σ^2 must be 'return squared per time', where the units of return are those in which x are measured. The units of Sharpe ratio are thus 'per square root time'. Thus when presented as an 'annualized' number, the units of Sharpe ratio are 'per square root years', or $\mathrm{yr}^{-1/2}$.

Once the units are explicit, any engineering student armed with sufficient domain knowledge can translate the units, as in the following example[2].

Example 2.2.1 (Annualization). Suppose one computes the sample mean and standard deviation of daily log returns of Asset Corp. to be, respectively, $3\,\mathrm{bps}\,\mathrm{day}^{-1}$ and $120\,\mathrm{bps}\,\mathrm{day}^{-1/2}$. The Sharpe ratio is then estimated as $0.025\,\mathrm{day}^{-1/2}$. To translate into 'annualized' units, note there are around 252 *market* days per year, thus multiply by the unit quantity $\sqrt{252}\,\mathrm{day}^{1/2}\mathrm{yr}^{-1/2}$ to get a Sharpe ratio of around $0.4\mathrm{yr}^{-1/2}$. ⊣

Example 2.2.2 (Annualized Sharpe ratio, the Market). Continuing Example 2.1.1, the monthly Sharpe ratio of the Market is $0.17\,\mathrm{mo.}^{-1/2}$. To get annualized terms, multiply by the unit quantity $\sqrt{12}\,\mathrm{mo.}^{1/2}\mathrm{yr}^{-1/2}$ to get a Sharpe ratio of around $0.6\mathrm{yr}^{-1/2}$. ⊣

This annualization is exact in the case where one computes the Sharpe ratio on log returns, and the returns are independent. That is, if one defines $x_{1:252} = \sum_{1 \le i \le 252} x_i$, where x_i are daily log returns, then $x_{1:252}$ is the actual annual log return, and the signal-noise ratio of $x_{1:252}$ (its expected value divided by its standard deviation) is exactly equal to $\sqrt{252}$ times the signal-noise ratio of x.

Often the Sharpe ratio is annualized in this manner without checking for potential autocorrelation of returns. Strong positive (negative) autocorrelation tends to cause one to underestimate (overestimate) the volatility of an asset, and thus overestimate (underestimate) the Sharpe ratio. For example, consider a lazy fund manager who provides a historical 'daily' mark-to-market constructed by linear interpolation of the quarerly marks of the fund. The standard deviation of daily returns will be an underestimate, since they do not reflect the real volatility of an accurate mark to market. On the other extreme, consider a fund with a strong negative autocorrelation of monthly returns, resulting in a sawtooth mark to market: observing this pattern, one could

[2] Herein you will see bps (pronounced '*bips*,' or, jocularly, '*beeps*') used to refer to *basis points*, or hundredths of a percent. Sometimes it is erroneously used to mean 10^{-4} in log return units.

time when to buy and when to sell the fund, in order to get in at a trough and out at a peak, and thus the actual volatility of returns is effectively overstated. See also Exercise 2.29 and Section 4.1.4.

2.2.1. Probability of a loss

Because the units are per square root time, it may make more sense to consider the parameter ζ^{-2}, which has time units. Indeed, if one measures signal-noise ratio on log returns, and assuming $\zeta > 0$, then $\kappa^2\zeta^{-2}$ is the time at which having accumulated a loss (or a loss against the risk-free rate) is a κ standard-deviation event. For a fund with a signal-noise ratio of $0.7\text{yr}^{-1/2}$, the event of experiencing a loss over 8.1633yr is a '2-sigma event.'

Typically when one states that an outcome is a '2-sigma event', one uses the CDF of the Gaussian to estimate its probability. While the central limit theorem should make the yearly returns of a trading strategy approximately normal, we can apply a very loose upper bound regardless of the form of the returns (assuming independence). By Cantelli's inequality, the probability of a $\kappa\sigma$ deviation below the mean is no greater than $\left(1 + \kappa^2\right)^{-1}$. This is the loosest bound on the probability of a loss (or underperformance relative to the risk-free rate) over a period of length $\left(\kappa/\zeta\right)^2$.

If the returns are sufficiently well behaved that one can accept the normal approximation, the probability of a loss is bounded from above by $e^{-\kappa^2/2}/\sqrt{2\pi\kappa^2}$, which is to say that doubling the Sharpe ratio of an asset cuts the probability of a loss by much more than a half. To be concrete, under the normal approximation, the probability of a 2-sigma event is around 0.0228; the probability of a 4-sigma event is around 3.1671×10^{-5};

Example 2.2.3 (Probability of a loss, Market returns). Using the monthly *log* returns of the Market (Example 1.2.1), the Sharpe ratio is computed as $0.5\,\text{yr}^{-1/2}$. Under Cantelli's inequality, the probability of a down year is no greater than 0.8; under the normal approximation, the probability is around 0.31. For the 92 year sample, the empirical probability of a down year is around 0.25. Thus Cantelli's inequality is far too conservative, and the normal approximation appears reasonable.

Note that the reasoning is circular here: the Sharpe ratio and empirical rate of down years were computed using the same data. ⊣

2.2.2. Interpretation of Sharpe ratio

The range of sane values one might expect to see of the Sharpe ratio vary depending on the type of asset, the market, relevant risk-free rate, and whether the returns are net or gross[3]. Among mutual funds and hedge funds, loosely, one should consider a signal-noise ratio above $0.5\text{yr}^{-1/2}$ to be 'good', while values above $1\text{yr}^{-1/2}$ to be 'very good.' Apparently Warren Buffet's Bershire Hathaway fund achieved a Sharpe ratio of

[3] Of course, the number of samples over which one computes the Sharpe ratio is also relevant; more on this in the sequel.

'only' $0.76\mathrm{yr}^{-1/2}$ over a 30 year period. [49] Jim Simons is reported to have achieved a Sharpe ratio of $1.89\mathrm{yr}^{-1/2}$ over a 10 year period. [102]

Those who engage in low-latency, 'high frequency trading' will often demur when asked what their achieved Sharpe ratio is, but may quote numbers as high as $20\mathrm{yr}^{-1/2}$ or higher[4]. Often these numbers ignore large fixed costs in maintaining their infrastructure, or keeping up in the technological arms race.

Mkt	SMB	HML	UMD
0.60	0.23	0.37	0.49

Table 2.1.: The Sharpe ratio of the four factor portfolios are given, in units of $\mathrm{yr}^{-1/2}$, computed on monthly relative returns from Jan 1927 to Dec 2018.

Example 2.2.4 (Sharpe ratio of the Fama French Four Factors). The Sharpe ratio of the Market, SMB, HML and UMD portfolios introduced in Example 1.2.1 were computed on the monthly relative returns, annualized, then tabulated in Table 2.1. The takeaway is that very simple monthly-rebalancing strategies have achieved Sharpe ratio in the ballpark of $0.5\,\mathrm{yr}^{-1/2}$, subject to the survivorship bias of examining the U. S. stock market. ⊣

Consumer	Manufacturing	Technology	Healthcare	Other
0.65	0.60	0.59	0.67	0.49

Table 2.2.: The Sharpe ratio of the five industry portfolios are given, in units of $\mathrm{yr}^{-1/2}$, computed on monthly relative returns from Jan 1927 to Dec 2018.

Example 2.2.5 (Sharpe ratio of the five industry portfolios). The Sharpe ratio of the five industry portfolios introduced in Example 1.2.3 were computed on the monthly relative returns, annualized, then tabulated in Table 2.2. These achieve about the same Sharpe ratio as the Market factor, all around $0.6\,\mathrm{yr}^{-1/2}$. ⊣

2.3. Historical perspective

That we call this ratio the 'Sharpe ratio' is another example of Stigler's Law of Eponymy. [156] In 1952, Roy introduced the 'Safety First' criterion as an objective to be maximized in investment decisions. [144] Arguing from Chebyshev's inequality, Roy shows that

$$\Pr\left\{x \leq r_0\right\} \leq \left(\frac{\sigma}{\mu - r_0}\right)^2, \tag{2.7}$$

[4] Novice quants will often quote similar numbers based on broken backtests of low frequency strategies.

and thus one should seek to minimize the right hand side quantity, equivalent to maximizing $(\mu - r_0)/\sigma$. Roy then goes on to illustrate what we now call the mean-variance frontier. Harry Marowitz and William Sharpe shared the 1990 Nobel Prize, while Roy died penniless and insane, trying to play a gramophone record with a peanut[5]. [157]

As a sample statistic, the Sharpe ratio is fairly similar to the t-statistic. The t-test for the hypothesis $H_0 : \mu \leq r_0$ uses the quantity

$$\frac{\hat{\mu} - r_0}{\sqrt{\hat{\sigma}^2/n}} = \sqrt{n}\hat{\zeta}.$$

This modern form of the t-test is *not* the form first considered by Gosset (writing as "Student").[60] 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 $n-1$. Gosset's statistic is asymptotically equivalent to the Sharpe ratio, a fact we will abuse in the sequel.

2.4. Linear attribution models

The Sharpe ratio can be viewed as a specific case of an *attribution model*, or *factor model*. In the general case one attributes the returns of the asset in question as the linear combination of l factors, one of which is typically the constant one:

$$x_t = \beta_0 1 + \sum_{i}^{l-1} \beta_i f_{i,t} + \epsilon_t, \tag{2.8}$$

where $f_{i,t}$ is the value of some i^{th} 'factor' at time t, and the innovations, ϵ, are assumed to be zero mean, and have standard deviation σ. Here we have forced the zeroth factor to be the constant one, $f_{0,t} = 1$.

Given n observations, let F be the $n \times l$ matrix whose rows are the observations of the factors (including a column that is the constant 1), and let $\boldsymbol{x}$ be the n length column vector of returns; then the multiple linear regression estimates are

$$\hat{\boldsymbol{\beta}} =_{\text{df}} \left(\mathsf{F}^{\top}\mathsf{F}\right)^{-1}\mathsf{F}^{\top}\boldsymbol{x}, \qquad \hat{\sigma} =_{\text{df}} \sqrt{\frac{\left(\boldsymbol{x} - \mathsf{F}\hat{\boldsymbol{\beta}}\right)^{\top}\left(\boldsymbol{x} - \mathsf{F}\hat{\boldsymbol{\beta}}\right)}{n-l}}. \tag{2.9}$$

We can then define a *ex-factor Sharpe ratio* as follows: let $\boldsymbol{v}$ be some non-zero vector, and let r_0 be some risk-free, or disastrous, rate of return. Then define

$$\hat{\zeta}_g =_{\text{df}} \frac{\hat{\boldsymbol{\beta}}^{\top}\boldsymbol{v} - r_0}{\hat{\sigma}}. \tag{2.10}$$

[5] This is an exageration; however, Roy has received little recognition for his work.

In all of the factor models we will consider here, we choose $\boldsymbol{v} = \boldsymbol{e}_0$, the vector of all zeros except a one corresponding to the intercept term. Let ex-factor signal-noise ratio be the population analogue.

Remark (Nomenclature). The terms "ex-factor Sharpe ratio" and "ex-factor signal-noise ratio" are idiosyncratic to this text. The author is not aware of any commonly used terms for these concepts, though the term "information ratio" is close. The information ratio refers to the Sharpe ratio of returns from which benchmark returns have been subracted, which is close to what we call the ex-factor Sharpe ratio, except no regression to the benchmark is performed.

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, either because those other factors are some kind of benchmark, or because one believes they have zero expectation in the future.

2.4.1. Examples of linear attribution models

- As noted above, the vanilla Sharpe ratio employs a trivial factor model, *viz.* $x_t = \beta_0 1 + \epsilon_t$. 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 simplest refinement to the trivial model would be to include some form of the interest rate, say $x_t = \beta_0 1 + \beta_f f_{f,t} + \epsilon_t$, where $f_{f,t}$ is the properly scaled 'risk free' rate of return. This model can lead to rank-deficiencies unless the interest rate has changed over the observation period. It is actually uncommon to fit the coefficient β_f, however; typically it is assumed to be 1, and the prevailing risk-free rate is subtracted from the return x before any modeling is performed.
- The models above do 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)[6], 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. [16, 35, 82]
  This is clearly a superior model for stocks and portfolios with a long bias (*e.g.*, typical mutual funds), but may seem inappropriate for *e.g.*, a long-short balanced hedge fund. 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

[6] The term 'CAPM' ' usually encompasses a number of assumptions used to justify the validity of this model for stock returns; here the term is abused to refer only to the factor model.

(one finds that $\hat{\beta}_0$ is very small); 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 model abound. For example, the Fama-French 3-factor model (I drop the risk-free rate for simplicity):

$$x_t = \beta_0 1 + \beta_{\mathrm{M}} f_{\mathrm{M},t} + \beta_{\mathrm{SMB}} f_{\mathrm{SMB},t} + \beta_{\mathrm{HML}} f_{\mathrm{HML},t} + \epsilon_t, \tag{2.11}$$

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. [43] Carhart adds a momentum factor:

$$x_t = \beta_0 1 + \beta_{\mathrm{M}} f_{\mathrm{M},t} + \beta_{\mathrm{SMB}} f_{\mathrm{SMB},t} + \beta_{\mathrm{HML}} f_{\mathrm{HML},t} + \beta_{\mathrm{UMD}} f_{\mathrm{UMD},t} + \epsilon_t, \tag{2.12}$$

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. [26] Alternative factor models include Elton *et al.*, Chen *et al.* and Asness *et al.* [41, 32, 9]
These factor models are designed to explain the returns of stocks, and are typically judged by their parsimony, explanatory power (lack of significant $\hat{\beta}_0$ terms across a wide universe of stocks), orthogonality of the factors, and narrative appeal. However, there is no reason they cannot be used to describe (or explain away, in terms of simpler strategies) the returns of an actively managed portfolio. In fact, there is typically little power lost in a 'kitchen sink' approach, if the objective is to eviscerate a proposed trading strategy.

- 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. [68]
Actually, one or several factor timing terms could 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'. [34]

Example 2.4.1 (UMD attribution). Consider the monthly simple returns of the Market, SMB, HML and UMD portfolios, as described in Example 1.2.1. In a somewhat unorthodox analysis, the returns of UMD are attributed to the other three factors and an intercept term. The regression fit for the monthly returns, against intercept, Market, SMB, and HML, is $\hat{\boldsymbol{\beta}} = [1.0395, -0.2127, -0.0337, -0.4752]^{\top}$; the residual volatility is $\hat{\sigma} = 4.1239\,\mathrm{mo.}^{-1/2}$. This yields $\hat{\zeta}_g = 0.8732\,\mathrm{yr}^{-1/2}$. This is somewhat larger than the value of Sharpe ratio computed without the attribution model, namely $\hat{\zeta} = 0.4884\,\mathrm{yr}^{-1/2}$, as found in Example 2.2.4. The factor model here acted to increase the expected value slightly and decrease the idiosyncratic volatility slightly. This is somewhat atypical: adding more factors to a model tends to decrease the Sharpe ratio. ⊣

Example 2.4.2 (UMD, Chow test). A strawman argument against momentum strategies is that 'they worked in the past,' but no longer do. Here we consider the monthly simple returns of the UMD portfolio, as described in Example 1.2.1. An attribution is made against the factor model where f_0 is the constant 1, and where $f_{1,t}$ is an indicator function for the event that t is prior to 1980-01-01. That is, $f_{1,t}$ is one for all time prior to 1980-01-01 and zero thereafter. (This is essentially a *Chow test*, but we are interested in the significance of the intercept term, rather than looking for a structural break. [34, 79])

The regression fit gives $\hat{\boldsymbol{\beta}} = [0.5829, 0.1367]^{\top}$ in units of $\%\,\mathrm{mo.}^{-1}$; the residual volatility is fit as $\hat{\sigma} = 4.6945\,\mathrm{mo.}^{-1/2}$. This yields $\hat{\zeta}_g = 0.4301\,\mathrm{yr}^{-1/2}$, slightly lower than the value of Sharpe ratio computed in the unattributed model, $\hat{\zeta} = 0.4884\,\mathrm{yr}^{-1/2}$, as found in Example 2.2.4. ⊣

2.4.2. † Heteroskedasticity attribution models

We have defined a linear attribution model where some vector of covariates (be they contemporaneous returns of other assets, random predictive 'signal' variables, seasonality variables, *etc.*) accounts for the expected returns of an asset, especially how they deviate from the unconditional expected return when conditioning on those covariates. Can we imagine a similar model where some variable(s) explains differences in heteroskedasticity? After all, while forecasting returns is considered difficult, there is a well known conditional heteroskedasticity effect in many markets [36, 115] that ought to be accounted for.

As it is much more difficult to deal with the case of multiple variables that explain heteroskedasticity, we will stick to the scalar case for now. Consider a strictly positive scalar random variable, s_t, observable at the time the investment decision is required to capture $\boldsymbol{x}_{t+1}$. It is more convenient to think of s_t as a 'quietude' indicator, or a 'weight' for a weighted regression, rather than as a 'volatility' indicator.

We can then generalize Equation 2.8 to

$$x_t = \beta_0 1 + \sum_{i}^{l-1} \beta_i f_{i,t} + s_{t-1}^{-1}\epsilon_t, \tag{2.13}$$

where we allow the $f_{i,t}$ to be correlated to, or even equal to, the quietude indicator, s_t, and where the error term ϵ_t is homoskedastic.

We present some examples of conditional heteroskedasticity models:

- The *market clock* or *volatility clock* model posits that returns should be measured on a volatility or volume basis rather than wallclock time. As such, under this theory, if variance of returns is, say, twice the long term variance, then expected returns ought to be twice the long term expectation. Another way of putting this is the returns over half a day in this regime should have the same distribution as daily returns over the long term[7]. Supposing that volatility can be measured before the investment decision, we have

  $$x_t = {s_{t-1}}^{-1}\beta_0 + s_{t-1}^{-1}\epsilon_t,$$

  where s_{t-1}^{-1} is the volatility factor. When $s_\cdot$ has a long term average of 1, then β_0 is the long term average return.
  Note that for any given time t, the single period return x_t has a signal-noise ratio independent of s_{t-1}, namely β_0 divided by the standard deviation of ϵ_t. When we consider multiple periods, however, the signal-noise ratio will deviate from this value. Note, however, that there is nothing particular to fear in this model about periods of high volatility since the asset holder is perfectly compensated with increased excess returns.
- A more depressing, though perhaps realistic, model of heteroskedasticity is that in periods of higher volatility there is no compensating increase in expected returns. We can express this as

  $$x_t = \beta_0 + s_{t-1}^{-1}\epsilon_t,$$

  where s_{t-1}^{-1} is the volatility factor. In this case the single period return x_t has signal-noise ratio proportional to s_{t-1}. That is to say the risk to reward is higher in periods of lower volatility, or "volatility drinks your milkshake."
- Rather than be forced to choose among these two models of heteroskedasticity, one can consider a mixed model of the form

  $$x_t = \beta_0 + s_{t-1}^{-1}\beta_1 + s_{t-1}^{-1}\epsilon_t,$$

[7] Never mind that this theory is rarely supported by data.

where now the coefficients β_0 and β_1 may be zero or non-zero, but can give us either of the two previously mentioned models, volatility clock and volatility milkshake, with values to be estimated from the data.

The trick for dealing with conditional heteroskedasticity models is to use the quietudes for weighted estimation. That is, rewrite Equation 2.13 as

$$\tilde{x_t} = \beta_0 s_{t-1} + \sum_{i}^{l-1} \beta_i \tilde{f_{i,t}} + \epsilon_t, \tag{2.14}$$

where we define $\tilde{x_t} =_{\mathrm{df}} s_{t-1} x_t$, and $\tilde{f_{i,t}} =_{\mathrm{df}} s_{t-1} f_{i,t}$. We can now fit the data $\tilde{x_t}$ on the new attribution factors $\tilde{f_{i,t}}$ using the machinery of the homoskedastic case.

Example 2.4.3 (VIX reweighting Market returns). Consider the daily VIX index introduced in Example 1.2.4 as a volatility indicator. We lag it by one trading day, then join to the daily returns of the Market introduced in Example 1.2.1. We rescale the inverse VIX to have unit mean in our sample period, which ranges from 1990-01-03 through 2020-12-31.

Over this period, the regular Sharpe ratio of the Market component is $0.64\,\mathrm{yr}^{-1/2}$, again assuming 252 trading days per year. When we rescale the daily Market returns by the inverse VIX, then compute the Sharpe ratio we compute a value of $0.76\,\mathrm{yr}^{-1/2}$. This is essentially the 'pessimistic model' where quietude weighted returns are modeled as having a constant expected return, conditional on the quietude. Under the 'vol clock' model we compute the ex-factor Sharpe ratio as $0.72\,\mathrm{yr}^{-1/2}$. Note that using an Information Criterion based approach, we slightly prefer the 'vol clock' model, but the difference in residual sum of squares between them is so very small that this is hardly conclusive proof.

$\dashv$

2.5. † Should you maximize Sharpe ratio?

In some circles the question 'should you maximize Sharpe ratio?' has an obviously affirmative answer: professional money managers are routinely asked to provide their achieved (or, more dubiously, backtested) Sharpe ratio to prospective clients. To a first order approximation, it is the *only* quantitative aspect of a trading strategy that clients care about. While volatility, capacity, leverage, fees, lockup terms, *etc.* are also important, the mythical 'utility' underpinning most investment theory [35, 39, 113, 82] has zero currency in these discussions: the author has *never* heard (nor heard of) an investor quote their utility function or risk-aversion parameter, while discussion of strategy Sharpe ratio is routine[8].

Pragmatics aside, there are two distinct questions regarding the investment decision and the Sharpe ratio, *viz.* 1. if all the population parameters of investment strategies were somehow known, should one maximize the signal-noise ratio? 2. is the sample Sharpe ratio a good predictor of whatever function of the population parameters one

[8] Often without any acknowledgement of the sample variance of the Sharpe ratio.

wishes to optimize? While the first question contains a glaring counterfactual, it is important nonetheless, since the second question relies on one's answer to the first.

It is generally assumed that investor's preferences are aligned with the *stochastic dominance* relations. We say that z (first-order) stochastically dominates x if, for all c, $\Pr\{x \geq c\} \leq \Pr\{z \geq c\}$, with inequality holding for some c. First order stochastic dominance implies second order stochastic dominance. The definition of second order stochastic dominance is too involved for our purposes here, but it should be noted that if z second-order stochastically dominates x, then z has no smaller mean and no larger variance than x.

While the stochastic dominance relations are typically considered sancrosanct, *i.e.*, all investors would prefer a returns stream to another that it stochastically dominates, stochastic dominance does not form a total ordering.[9] That is, it is possible to construct two returns streams such that neither stochastically dominates the other. For purposes of selecting an investment, one would like to be able to compare any two investments, so a total ordering is required. Here the term 'objective' will be used to mean some real-valued function of the population parameters of a returns stream (*e.g.*, μ, σ, *etc.*) which one wishes to maximize. An objective trivially forms a total ordering.

At a bare minimum one expects an objective should be consistent with stochastic dominance. That is, if z stochastically dominates x, it should have a larger objective. Regretably, the signal-noise ratio as an objective function is *not* consistent with stochastic dominance. It is not consistent with second order stochastic dominance, as most clearly demonstrated by considering the case $\mu < 0$, where the signal-noise ratio 'prefers' higher σ^2. This wrinkle is *not* fixed by assuming away the $\mu < 0$ case[10], since the signal-noise ratio is not even consistent with the stronger first-order stochastic dominance relation, as shown by Hodges' counterexample, *cf.* Exercise 2.33. [71, 154, 178]

Despite these failings, it is taken as axiomatic in this text that one should optimize the signal-noise ratio, μ/σ. There are numerous reasons to optimize the signal-noise ratio. Among them: 1. the signal-noise ratio bounds the probability of a loss against the disastrous rate, r_0; 2. a large signal-noise ratio approximately bounds the probability of experiencing a large drawdown, as measured in units of volatility; 3. a large signal-noise ratio increases the probability of experiencing a large Sharpe ratio in a fixed period of trading.[11] 4. since the signal-noise ratio is simply defined, it is a relatively easy objective to optimize analytically (again, assuming one knew the population parameters; noisy estimation thereof adds extra complication). Moreover, the deficiencies of the signal-noise ratio are somewhat mitigated for the long term investor, since the central limit theorem guarantees convergence to normality.

If the goal is maximization of the signal-noise ratio, then the Sharpe ratio is a decent

[9] The example of 'non-transitive dice' illustrate that some forms of probabilistic dominance do not even form partial orders. [54, 45]

[10] No investor should consider a returns stream with $\mu < 0$ if they can keep their wealth as cash.

[11] This is a bit (a)circular: a large achieved Sharpe ratio is supposed to instill confidence in prospective investors that $\mu > r_0$, but it is proposed that one achieve a large Sharpe ratio by maximizing μ/σ, rather than simply maximizing μ.

yardstick by which to measure investments, modulo a number of provisos: the Sharpe ratio must be measured over the same length of time, as in *e.g.*, Sharpe's original paper [149], or some correction should be made for different standard errors; if the returns are measured over common time intervals, one must take into account possibly correlated errors; the Sharpe ratio measured on a backtest should be viewed with great suspicion, as backtests are often riddled with methodological and coding errors, and the product of great amounts of data snooping. While, taken together, these make the Sharpe ratio seem like an awful objective on realized returns streams, all other metrics share the same problems. Moreover, many of the proposed alternative metrics defined on achieved returns streams lack any kind of theoretical justification, and many have worse sample variances. See Exercise 2.38.

Example 2.5.1 (Sample variance of the Sortino ratio). The Sortino ratio is defined as the sample mean divided by the downside semivariation. While this sounds like a perfectly reasonable metric, the downside semivariation has slightly higher sample variation than the standard deviation. For the case of symmetric returns, the Sortino ratio should, then, have slightly higher variance than the Sharpe ratio.

To test this, 52 weeks of weekly returns were drawn from a population with $\mu = 0\text{wk.}^{-1}, \sigma = 0.02\text{wk.}^{-1/2}$, and the Sortino ratio was computed. The signal-noise ratio was also computed. This was repeated 5000 times. The same computations were then repeated using $\mu = 0.003\text{wk.}^{-1}, \sigma = 0.02\text{wk.}^{-1/2}$. Empirical critical values for Frequentist tests based on both statistics were computed by taking the 0.95 quantiles of each statistic under the first population. The empirical powers were then computed by computing the proportion of statistics from the second population exceeding the critical value. The Sortino ratio was found to have a power of approximately 0.2552, while the Sharpe ratio had a power of approximately 0.2732. ⊣

2.6. † Probability of a loss, revisited

The crux of Roy's justification for the 'Safety-First' criterion, which is just the signal-noise ratio, is that it bounds the probability of a loss, defined as a return less than r_0, *cf.* Equation 2.7. [157] This argument is based on Chebyshev's inequality, and so is a fairly rough upper bound. There are some situations where the signal-noise ratio is exactly monotonic in the probability of a loss. For example, if the returns are drawn from a scale-location family, *i.e.*, if the probability density of x is $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, for some function $f(\cdot)$. In this case the probability that the return is less than r_0 is exactly equal to $F\left(\frac{r_0-\mu}{\sigma}\right)$ for some function $F(\cdot)$. In this case the signal-noise ratio is consistent with first order stochastic dominance.

The central limit theorem tells us that, conditional on finite variance, the mean return of x will converge to a normal distribution. Noting that for log returns, by the telescoping property, the mean return is just the total return rescaled, thus the long term log return is approximately drawn from a scale-location family. (*cf.* Exercise 2.35.)

One can construct tighter approximations than given by Chebyshev's inequality by considering the classical approximations to the central limit theorem. These will result in objectives which are sensitive to the investor's time horizon and the disastrous rate, r_0. [154, 178, 20]

Suppose that one will observe n independent draws from the returns stream, x. The disastrous event is that the observed mean sample return, $\hat{\mu}$, is less than r_0. This is equivalent to

$$\sqrt{n}\frac{\hat{\mu}-\mu}{\sigma} \leq \sqrt{n}\frac{r_0-\mu}{\sigma}.$$

The cumulative distribution function of the quantity on the left hand side can be approximated via some truncation of the *Edgeworth expansion*. [33]

Define[12] $\delta =_{\mathrm{df}} \sqrt{n}\left(\mu - r_0\right)/\sigma$. The Edgeworth expansion is [2, 26.2.48]

$$\begin{aligned}\Pr\left\{\sqrt{n}\frac{\hat{\mu}-\mu}{\sigma} \leq -\delta\right\} &= \Phi\left(-\delta\right) - \phi\left(\delta\right)\left[\frac{\gamma_1}{6\sqrt{n}} He_2\left(\delta\right)\right] \\ &+ \phi\left(\delta\right)\left[\frac{\gamma_2}{24n} He_3\left(\delta\right) + \frac{{\gamma_1}^2}{72n} He_5\left(\delta\right)\right] \\ &- \phi\left(\delta\right)\left[\frac{\gamma_3}{120n^{3/2}} He_4\left(\delta\right) + \frac{\gamma_1\gamma_2}{144n^{3/2}} He_6\left(\delta\right) + \frac{{\gamma_1}^3}{1296n^{3/2}} He_8\left(\delta\right)\right]\ldots \end{aligned} \tag{2.15}$$

where $\Phi\left(x\right)$ and $\phi\left(x\right)$ are the cumulative distribution and density functions of the standard unit normal, $He_i\left(x\right)$ is the probabilist's Hermite polynomial [2, 26.2.31], and γ_i is the standardized $i-2^{\mathrm{th}}$ cumulant, defined as the $i+2^{\mathrm{th}}$ cumulant of the distribution divided by σ^{i+2}. It happens to be the case that γ_1 is the skewness, and γ_2 is the excess kurtosis of the distribution.

Truncating beyond the $n^{-1/2}$ term, and applying basic facts of probability, gives

$$\Pr\left\{\hat{\mu} \geq r_0\right\} \approx \Phi\left(\delta\right) + \frac{\phi\left(\delta\right)}{\sqrt{n}}\left[\frac{\gamma_1}{6}\left(\delta^2 - 1\right)\right]. \tag{2.16}$$

The implication is that the probability that $\hat{\mu}$ exceeds r_0 will be increased if δ is large. Moreover, for a fixed δ, the probability that $\hat{\mu}$ exceeds r_0 is increased for large positive skew if $\delta^2 > 1$, but for large negative skew when when $\delta^2 < 1$. The implication is that when δ^2 is 'large' (compared to the unit), one would buy lottery tickets, otherwise one would sell lottery tickets[13]. This is asymptotically compatible, as $n \to \infty$, with the commonly held belief that investors universally prefer positive skew.

The Edgeworth expansion suggests a 'higher order signal-noise ratio', defined as [125]

$$\zeta_h =_{\mathrm{df}} -\frac{1}{\sqrt{n}}\Phi^{-1}\left(\Pr\left\{\sqrt{n}\frac{\hat{\mu}-\mu}{\sigma} \leq -\delta\right\}\right). \tag{2.17}$$

When all the higher order cumulants are zero, *i.e.*, when returns are normally distributed, the probability of a loss is exactly $\Phi\left(-\delta\right)$, so in this case $\zeta_h = \delta/\sqrt{n} =$

[12] In the sequel, we will see that δ so defined is the non-centrality parameter of the t-distribution associated with the Sharpe ratio.

[13] This illustrates the possible mismatch between the objectives of an investor and the money manager: the former is likely concerned about the long term, while the latter may be focused on the short term.

$(\mu - r_0)/\sigma$, which is just the signal-noise ratio with a risk-free rate r_0. This definition gives an objective which is consistent with first-order stochastic dominance (Exercise 2.36), and generalizes the signal-noise ratio in an intuitive way that is easy to understand. However, it is essentially investor-dependent in that it is parametrized by n and r_0. [154]

The implicit definition of Equation 2.17 is a bit unwieldy for use as an objective. One would prefer a definition in terms of the cumulants of the returns stream. Rather than use the Taylor series expansion of $\Phi^{-1}(x)$, one can instead use the *Cornish Fisher expansion* of the sample quantile. [91, 75, 160]

Let $Y = \sqrt{n}(\hat{\mu} - \mu)/\sigma$. This is a random variable with zero mean and unit standard deviation. Let γ_i be the $i+2^{\text{th}}$ standardized cumulant of x. The $i+2^{\text{th}}$ standardized cumulant of Y is $n^{1-i/2}\gamma_i$. The Cornish Fisher expansion is [2, 26.2.49]

$$\Pr\{Y \le w\} = \Phi(z),$$

where

$$\begin{aligned} w = z &+ \frac{1}{\sqrt{n}}\left[\frac{\gamma_1}{6}He_2(z)\right] \\ &+ \frac{1}{n}\left[\frac{\gamma_2}{24}He_3(z) - \frac{{\gamma_1}^2}{36}\left[2He_3(z) + He_1(z)\right]\right] \\ &+ \frac{1}{n^{3/2}}\left[\frac{\gamma_3}{120}He_4(z) - \frac{\gamma_1\gamma_2}{24}\left[He_4(z) + He_2(z)\right] + \frac{{\gamma_1}^3}{324}\left[12He_4(z) + 19He_2(z)\right]\right] \\ &+ \ldots \end{aligned} \tag{2.18}$$

To estimate the higher-order signal-noise ratio, recognize that $z = -\sqrt{n}\zeta_h$, and $w = -\delta$. This gives

$$\begin{aligned} \zeta_h = \frac{(\mu - r_0)}{\sigma} &+ \frac{1}{n}\left[\frac{\gamma_1}{6}He_2\left(\sqrt{n}\zeta_h\right)\right] \\ &- \frac{1}{n^{3/2}}\left[\frac{\gamma_2}{24}He_3\left(\sqrt{n}\zeta_h\right) - \frac{{\gamma_1}^2}{36}\left[2He_3\left(\sqrt{n}\zeta_h\right) + He_1\left(\sqrt{n}\zeta_h\right)\right]\right] \\ &+ \ldots \end{aligned} \tag{2.19}$$

While this defines ζ_h implicitly, truncation gives polynomial equations, whose roots can be found analytically or numerically. See Exercise 2.37.

Example 2.6.1 (Higher order signal-noise ratio of the Market). Here we consider what might be called the higher order Sharpe ratio of the Market, introduced in Example 1.2.1. The mean simple return of the Market was estimated by sampling, with replacement, 12 months of returns. This process was repeated 5000 times, and the empirical probability that the mean monthly return was less than $0\%\text{mo.}^{-1}$ was computed. The normal quantile was then computed, yielding an approximate higher order Sharpe ratio of $0.4394\text{yr}^{-1/2}$. The computation was repeated for $r_0 = 0.25\%\text{mo.}^{-1}$, yielding a higher order Sharpe ratio of $0.2997\text{yr}^{-1/2}$. Compare these with the Sharpe ratio computed on monthly returns for these two choices of r_0, which were found to be $0.4194\text{yr}^{-1/2}$ and $0.2574\text{yr}^{-1/2}$, respectively. ⊣

2.7. † Drawdowns and the signal-noise ratio

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 or prolonged drawdown. Moreover, drawdowns are statistically nebulous: the sample maximum drawdown does not correspond in an obvious way to some population parameter; the variance of sample maximum drawdown is usually large, and depends on sample size in a subtle way; traded strategies are typically cherry-picked to not have a large maximum drawdown in backtests; the distribution of maximum drawdowns is affected by skew and kurtosis, heteroskedasticity, omitted variable bias and autocorrelation. Even assuming *i.i.d.* Gaussian returns, modeling drawdowns is non-trivial. [103, 14]

Given n observations of the mark to market of a single asset, p_i, the drawdown from the high water mark, as a time series, is defined as

$$D_j =_{\text{df}} \max_{1 \le i \le j} \log\left(\frac{p_i}{p_j}\right). \tag{2.20}$$

As so defined, the drawdown is negative the most extreme peak to point log return, and so is always non-negative[14]. The drawdown can be expressed as a a percent loss from the high watermark as $100\left(1 - e^{-D_j}\right)\%$.

Example 2.7.1 (drawdowns of the Market). Consider the Market portfolio, introduced in Example 1.2.1. Its drawdown from high watermark is shown in Figure 2.1. The maximum drawdown over the period is around 80%. ⊣

The *maximum drawdown* is a commonly computed statistic of backtested and live strategy returns. It is literally the maximum of the drawdown series, or

$$M_n =_{\text{df}} \max_{1 \le j \le n} D_j = \max_{1 \le i < j \le n} \log\left(\frac{p_i}{p_j}\right). \tag{2.21}$$

There is a connection between drawdowns and the signal-noise ratio, which is made obvious when one uses volatility as the units in which drawdown is measured. 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

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

where y_i is a zero-mean, unit-variance random variable that is a shifted, rescaled version of x_i.

[14] Under this definition a large drawdown is a bad event, which matches common informal usage of the term.

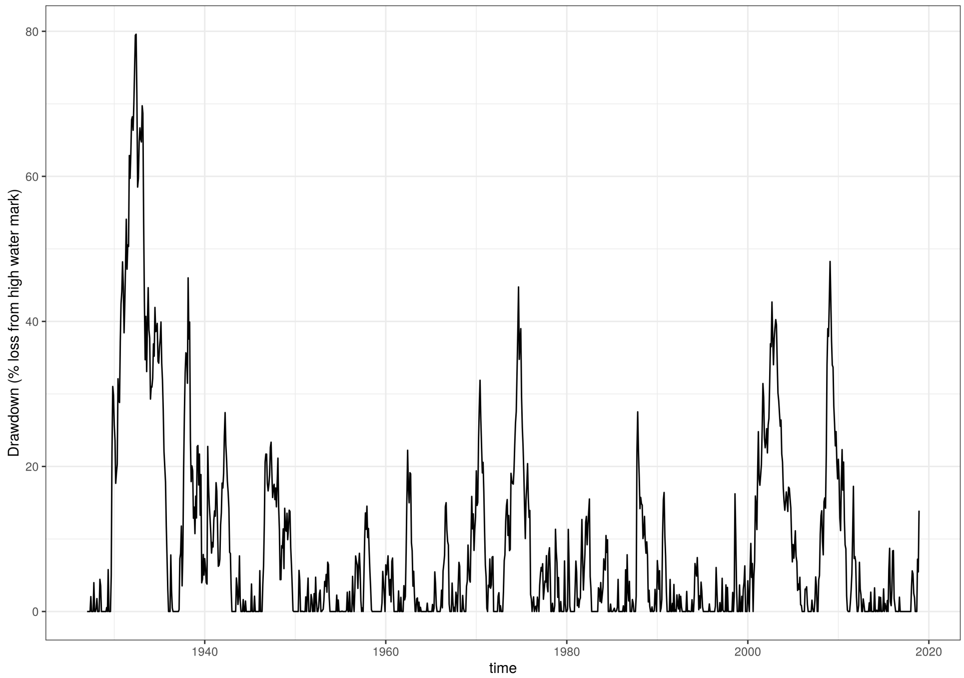

Figure 2.1.: The drawdown from high watermark of the Market is shown, as a percent loss.

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

$$\frac{M_n}{\sigma} = -\min_{1 \leq i < j \leq n} \left([j - i - 1] \, \zeta + \sum_{i < k \leq j} y_k \right). \tag{2.22}$$

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. It should be clear, moreover, that within this one parameter family, stochastic dominance of the drawdown distribution is monotonic in the signal-noise ratio. That is, a higher signal-noise ratio leads to a lower probability of a drawdown of a fixed size, *ceterus paribus*.

2.7.1. Worst statistic ever?

Equation 2.22 illustrates one reason why the sample maximum drawdown is a terrible statistic for estimating future performance of a strategy: it depends on volatility and signal-noise ratio in ways that are hard to disambiguate. That is, if one performed a hypothesis test based solely on the sample maximum drawdown, one would reject the null[15] if either the signal-noise ratio were high or the volatility were low.

Secondly, it is not at all clear that the variance of the sample maximum drawdown statistic is actually decreasing with sample size. That is, if one considers a longer time history, the sample maximum drawdown for any backtest can only increase, but the variance of that statistic may increase as well. For an arbitrarily long backtest history (longer than one could sensibly ever produce) one suspects that all sample maximum drawdowns should be nearly 100%. Depending on the signal-noise ratio there is likely to be an 'unsweet spot' of sample length where the variance of the maximum drawdown is actually *maximized*.

Another problem with sample maximum drawdown is that the statistic typically has a higher variance than other statistics. To illustrate this, 8192 Monte Carlo simulations of 5 years of weekly returns data were drawn from two populations. Returns are normally distributed with an equivalent daily volatility of around $0.013\,\mathrm{day}^{-1/2}$. The two populations correspond to the null, with $\zeta = 0\mathrm{yr}^{-1/2}$ and the alternative, with $\zeta = 0.75\mathrm{yr}^{-1/2}$, where signal-noise ratio is measured on log returns. For each draw of historical data, the sample maximum drawdown and Sharpe ratio were computed. The 0.95 empirical quantile for sample maximum drawdown under the null was around 22%. However, it is not the case that a large number of simulations under the alternative have smaller maximum drawdown. Only 2524 do, thus the empirical power is around 0.31. In comparison, the empirical power for the Sharpe ratio statistic in this experiment is around 0.53. (The theoretical power in this case is closer to 0.51. See Section 3.5.3.)

The empirical cumulative distributions of the two statistics for the two populations are shown in Figure 2.2, with vertical lines plotted at the 0.95 empirical quantile for the null population. The lack of power for the sample drawdown is apparent in this plot. See also Exercise 2.38 and Exercise 2.39.

2.7.2. Controlling drawdowns via the signal-noise ratio

While the sample maximum drawdown is an inherently flawed statistic, real portfolio drawdowns are nevertheless a serious occupational hazard. How can drawdowns be controlled? Ignoring the contributions to drawdown caused by autocorrelation or skewed distributions, the decomposition in Equation 2.22 indicates that, on a per-vol basis, drawdowns are stochastically monotonic in the signal-noise ratio when choosing from a one parameter family.

One reasonable way a portfolio manager might approach drawdowns is to define a 'knockout' drawdown from which she will certainly not recover[16] 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.[17]

[15] Presumably the null is that the strategy is 'not good', but this would have to be made precise in terms of the population parameters to form an actual hypothesis test.

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

[17] Another problem with this formulation is that humans typically have poor intuition for rare events.

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

$$\mathcal{C}\left(\epsilon, \delta, n\right) =_{\mathrm{df}} \left\{(\zeta, \sigma) \,|\, \sigma > 0,\, \mathrm{Pr}\left\{M_n \geq \sigma\epsilon\right\} \leq \delta\right\}. \tag{2.23}$$

These are the points in the two dimensional signal-noise ratio and volatility space that have strictly positive volatility and for which the probability of a drawdown exceeding $\sigma\epsilon$ is less than δ. 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. The contants σ_0 and b have to be approximated numerically, as in Example 2.7.2. It should be noted that a linear approximation of this form is easy to encode as a constraint in a portfolio optimizer.

Example 2.7.2 (halfspace for drawdown control). 5 years of weekly returns are drawn from a Gaussian distribution with a fixed signal-noise ratio and volatility. The maximum drawdown is computed in volatility units. This is replicated 50000 times for a fixed value of signal-noise ratio, and the $1 - 0.005$ quantile is computed. Then the volatility such that this quantile is equal to a 40% loss is computed, abusing the rescaling relationship of Equation 2.22. This is repeated for signal-noise ratio ranging from $0\mathrm{yr}^{-1/2}$ to $2.5\mathrm{yr}^{-1/2}$. This estimated cutoff volatility of log returns, in annualized units (*i.e.*, log return $\mathrm{yr}^{-1/2}$) is plotted versus the signal-noise ratio (in units of $\mathrm{yr}^{-1/2}$), along with a linear fit, in Figure 2.3. The linear fit is $\sigma_0 \approx 0.0856\mathrm{logret}\,\mathrm{yr}^{-1/2}$ and $b \approx 0.0906\mathrm{logret}$. As indicated in the figure, a quadratic model is likely to give a better fit, a finding supported by an F test. See also Exercise 2.40.

$\dashv$

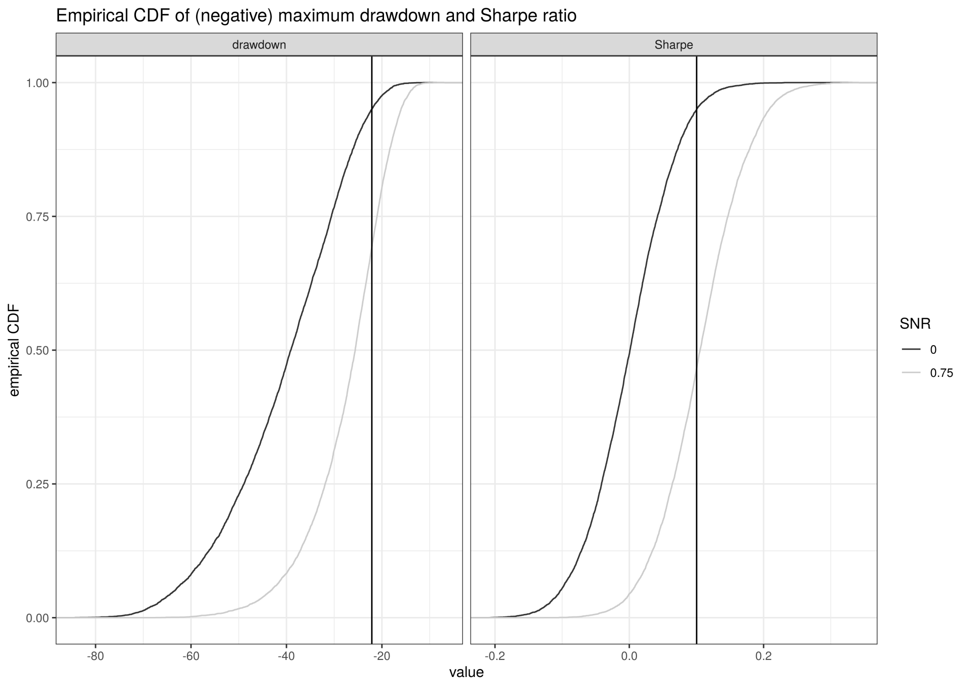

Figure 2.2.: 8192 Monte Carlo simulations of 5 years of weekly returns data were drawn from two populations. The first has a signal-noise ratio of $0\text{yr}^{-1/2}$, the other $0.75\text{yr}^{-1/2}$, where signal-noise ratio is measured on *log* returns. Returns are normally distributed with a daily volatility of around 0.013. The sample maximum drawdown and the Sharpe ratio were computed for each series. On the left empirical CDFs of the *negative* maximum drawdown (in percent) are shown for the two populations; on the right the empirical CDFs of the Sharpe ratio are plotted. For both these statistics, larger values is considered better. Vertical lines are plotted at the 0.95 empirical quantile for the $\zeta = 0\text{yr}^{-1/2}$ case. The power of the implicit Sharpe ratio test is much higher than for the sample maximum drawdown.

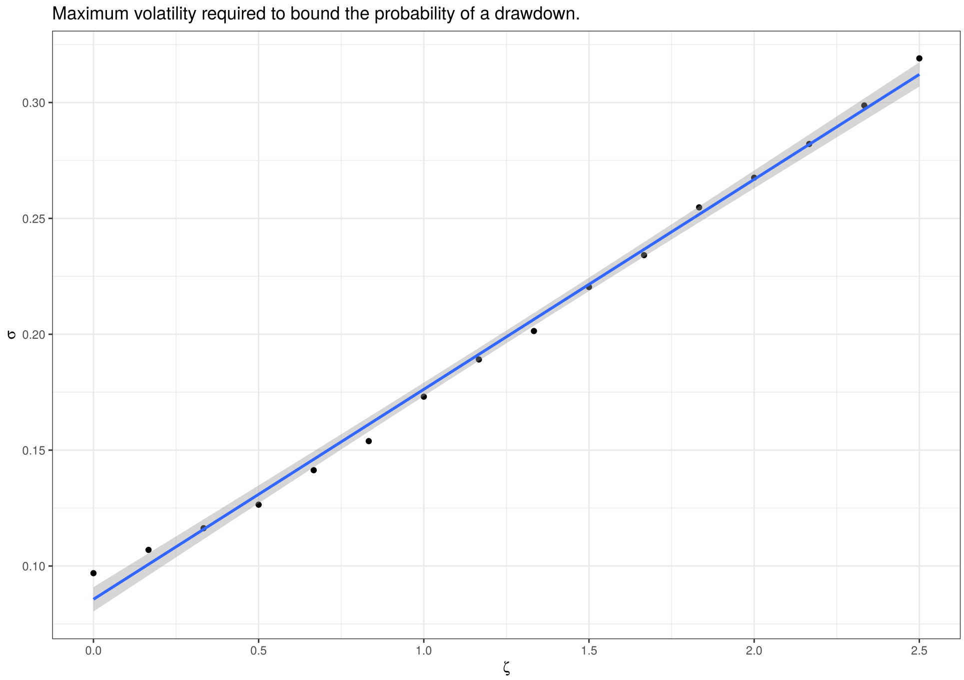

Figure 2.3.: The maximum annualized volatility of log returns such that the probability of a drawdown of 40% is less than 0.005, as estimated by 50,000 Monte Carlo simulations of weekly Gaussian returns is plotted versus the annualized signal-noise ratio. The fit is given by $\sigma \approx 0.086 \text{logret}\,\text{yr}^{-1/2} + 0.091 \text{logret} \zeta$.

Exercises

Ex. 2.1 Units conversion

1. Convert a Sharpe ratio of $0.2\,\mathrm{mo.}^{-1/2}$ computed on monthly returns to annual units.
2. Convert an annual Sharpe ratio of $0.8\,\mathrm{yr}^{-1/2}$ to daily units.

Ex. 2.2 Compute the Sharpe ratio Select some publicly traded company you consider successful. Compute the Sharpe ratio of its log returns over the past five years.

Ex. 2.3 Returns aggregation How do returns aggregate?

1. Let x_i be the relative return of an asset over some period. Show that the relative return of the asset over n periods is

$$\prod_{1\le i\le n} (1 + x_i) - 1.$$

2. Now let x_i be log returns of an asset over some period. Show that the total log return of the asset over n periods is the sum $\sum_{1\le i\le n} x_i$.

Ex. 2.4 Leverage By borrowing money, one can purchase an asset *with leverage*: One borrows l proportion of one's wealth, purchases an asset worth $1+l$ proportion of one's wealth, waits a single period, sells the asset, and returns the original loan, with interest paid. (In the simple formulation, we ignore the borrow costs.)

1. What is the relative return of a leveraged long holder of an asset?
2. What is the log return of a leveraged long holder?

Ex. 2.5 Shorting Suppose that the relative return of an asset is x. Suppose that you *short* the asset: you borrow (say, without interest) from a long holder, sell the borrowed asset, wait one period, buy the asset (preferrably at a lower price), and return to the lender.

1. What is the one-period relative return of the short holder?
2. Suppose $-1 \le x \le M$, where $M > 1$ is some large maximum return. What are the bounds of the returns to the short holder?
3. What is the one-period log return of shorting? Is it necessarily well-defined?
4. Suppose one holds the asset short for multiple periods, without any rebalancing. The relative returns of the asset for each period are $x_1, x_2, \ldots, x_n$. What is the relative return of the short holder?

Ex. 2.6 Returns of a portfolio Show that relative returns are laterally additive. That is, suppose you hold $a_i > 0$ proportion of your wealth in asset i, which experiences return x_i in a given period. Show that the relative return of your wealth is $\sum_i a_i x_i$. Does this still hold when one uses $a_i < 0$ to represent holding an asset short? Does this still hold when allowing $a_i > 1$ to represent buying with leverage?

Ex. 2.7 Concavity of log returns Use calculus to show that log returns are always less than the equivalent relative returns. That is show that $\log(1 + x) \le x$, where x is the relative return.

Ex. 2.8 The Fundamental Law of Asset Management Suppose you could invest an equal proportion of your wealth in n independent assets, each with identical mean and variance, μ and σ^2, respectively. Show that the signal-noise ratio of your portfolio, measured on relative returns, scales as $\sqrt{n}$.

This is an example of the 'Fundamental Law of Asset Management', which vaguely states that a constant 'edge' aggregates as the square root of the number of independent opportunities to apply the edge. [61] Note that the same scaling applies in the translation of time units of the signal-noise ratio.

Ex. 2.9 The opposite of a bad strategy Suppose that Alice and Bob have antithetical views of the market. If Alice holds a_i proportion of her wealth in an asset at a given time, Bob shorts a_i proportion of his wealth in that asset, and vice versa. (That is, imagine that at each day's close, Alice and Bob instantaneously trade so that their portfolios have this property, incur no trade costs, get the same trade price, *etc.*)

1. If Alice's relative return over some period is x, what is Bob's relative return over the same period?
2. Can it be the case that *both* Alice and Bob lose money over a single rebalance period?
3. Can it be the case that *both* Alice and Bob lose money over the span of multiple rebalance periods?
4. Suppose you traded at random, *i.e.*, used a random number generator to allocate your wealth among some fixed set of assets. Ignoring trading costs, why should it be the case that your expected *log* returns are negative?

Ex. 2.10 Leveraged ETFs Certain *leveraged ETFs* purport to have daily relative returns which are some multiple of the daily relative returns of some index (or some other ETF). [180] It is noted that these products tend to underperform in 'flat' markets.

Let x be the relative return of an ETF tracking an index, and let y be the relative returns of a $3\times$ levered ETF on the same index. Suppose that the ETF makes a two period 'round trip': $x_1 = \epsilon, x_2 = -\epsilon/(1 + \epsilon)$.

1. Show that the total relative return is zero.
2. What is the total relative return of y? Plot this as a function of ϵ. Include negative values of ϵ.

3. Is there an arbitrage opportunity if one can trade an ETF on an index and a leveraged ETF on the same index?

Ex. 2.11 Leverage and log return signal-noise ratio Let μ, σ^2 be the mean and variance of the relative returns of a strategy. By levering up the strategy by a factor of k, the mean becomes $k\mu$ and the variance $k^2\sigma^2$. Let ζ_g be the signal-noise ratio measured on the *log* returns. Using the approximation

$$\zeta_g \approx \frac{\mu - \frac{1}{2}\sigma^2}{\sigma},$$

what happens to ζ_g as k becomes large? How do you interpret these results?

Ex. 2.12 Translation to log returns The transformation from relative returns to log returns is given by $f_{\text{log}}\left(x\right) = \log\left(1 + x\right)$.

1. Compute the Taylor series expansion of $f_{\text{log}}\left(\cdot\right)$ around zero.
2. Let x be the relative returns, and let $\alpha_i = \operatorname{E}\left[x^i\right]$. Find the series representation of $\operatorname{E}\left[f_{\text{log}}\left(x\right)\right]$ in terms of α_i. What is the two term approximation of $\operatorname{E}\left[f_{\text{log}}\left(x\right)\right]$?
3. Compute the first three terms of the expansion of $\left(f_{\text{log}}\left(\cdot\right)\right)^2$. What is the two term approximation of $\operatorname{E}\left[\left(f_{\text{log}}\left(x\right)\right)^2\right]$?
4. What is the two term approximation of $\operatorname{E}\left[\left(f_{\text{log}}\left(x\right)\right)^2\right] - \left(\operatorname{E}\left[f_{\text{log}}\left(x\right)\right]\right)^2$?
5. Using the fact that
$$\frac{1}{\sqrt{\sigma^2 + \epsilon}} \approx \frac{1}{\sigma}\left(1 - \frac{1}{2}\frac{\epsilon}{\sigma^2}\right),$$
express the signal-noise ratio computed on log returns in terms of that computed on relative returns and some of the moments α_i.
6. Suppose the relative returns are normally distributed with mean $\mu \neq 0$ and variance σ^2. Then we have $\alpha_1 = \mu$, $\alpha_2 = \mu + \sigma^2$, $\alpha_3 = \mu^3 + \mu\sigma^2$, $\alpha_4 = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$. Write an approximation to the signal-noise ratio measured on log returns in terms of μ and σ.

Ex. 2.13 Translation to relative returns Repeat the previous exercise, but make the reverse translation. That is, let the transformation from log returns to relative returns be given by $f_{\text{exp}}\left(x\right) = e^x - 1$. Let x be the log returns, and let $\alpha_i = \operatorname{E}\left[x^i\right]$. Express the signal-noise ratio measured on relative returns in terms of the α_i using Taylor's series expansions.

Ex. 2.14 Fractal returns Suppose that daily *log* returns, x, are normally distributed. mean μ and variance σ^2.

1. Show that the weekly returns are normally distributed.
2. Assume hourly returns are independent and identically distributed. Argue that they are normally distributed. (*i.e.*, go look up Cramér's theorem.)

* 3. If x are normally distributed relative returns, are weekly returns normally distributed? Hourly returns?

Ex. 2.15 **Fractal returns II** Compare the daily relative returns of the Market component of the Fama French factors to 'fortnightly' returns. That is, first scale up the daily returns by multiplying them by 10, calling this sample A. Then convert relative returns to log returns, sum them in groups of 10, and convert back to relative returns to get sample B.

1. Perform a two-sample empirical Q-Q plot of the two.
2. Perform a two-sample non-parametric test of equality of the two distributions, like the Kolmogorov-Smirnov test, or the Baumgartner-Weiss-Schindler test. [13, 127]
3. Which returns 'look more normal'? That is, Q-Q plot both in turn against a normal law.
4. Perform the same comparisions on random data: let the fake daily returns be drawn from a fat-tailed distribution, say a $t(4)$ distribution, but independent of each other. Draw a sample of the same size as the daily market returns, and with the same standard deviation. Do you see similar changes to the fortnightly returns? Can we blame the apparent 'smoothing' of real fortnightly returns to mean reversion, or is it purely an effect of the central limit theorem?

* **Ex. 2.16** **Boost from relative returns** Suppose ζ_r is the signal-noise ratio measured on relative returns, and ζ_g is the signal-noise ratio measured using log returns.

1. Show that
$$\frac{\zeta_r - \zeta_g}{\zeta_g} \approx \frac{1}{2}\frac{\alpha_{2,g}}{\mu_g},$$
where $\mu_g = \mathrm{E}\left[x\right]$ and $\alpha_{2,g} = \mathrm{E}\left[x^2\right]$, and x are the log returns.
2. Suppose $\mu_g = 3 \times 10^{-4}\mathrm{day}^{-1}$, and $\sigma_g = 1.3 \times 10^{-2}\mathrm{day}^{-1/2}$ are the mean and standard deviation of daily log returns. Compute the boost from computing signal-noise ratio on relative returns instead of log returns, expressed as a percent, the left hand side of the above approximation. Also evaluate the right hand side.
3. Write an approximation for ζ_g, the signal-noise ratio on log returns, in terms of ζ_r and σ_r, the signal-noise ratio and volatility of relative returns. What is the order of the error? Does the approximation improve in the case where log returns have zero skew?

Ex. 2.17 **Cantelli's inequality** Describe a random variable for which Cantelli's inequality is an equality. Is this a reasonable model for the returns of *e.g.*, a mutual fund?

* **Ex. 2.18 Higher order moments** It is often said that investors have positive appetite for odd order moments of returns (mean and skew), and negative appetite for even order moments (volatility, kurtosis). Here you will expand Roy's argument to include higher order moments. [157] Let $r_0 \le \mu$ be the 'disastrous rate' of return. Let $p > 2$.

1. Show that
$$\Pr\left\{x \le r_0\right\} \le \frac{\mu_p}{\left|\mu - r_0\right|^p},$$
where $\mu_p =_{\mathrm{df}} \mathrm{E}\left[(x-\mu)^p\right]$. (*Hint:* Express the probability as an integral of a dummy indicator, and multiply the numerator and denominator by $|x-\mu|^p$, then bound the integral. This should be similar to the proof of Chebyshev's inequality.)
2. For even p, rewrite this as
$$\Pr\left\{x \le r_0\right\} \le \left(\frac{{\gamma_p}^{1/p}}{\zeta}\right)^p,$$
where ζ includes the 'risk-free' rate r_0, and
$$\gamma_p =_{\mathrm{df}} \frac{\mathrm{E}\left[(x-\mu)^p\right]}{\sigma^p}$$
is the standardized p^{th} moment. The conclusion is that one should maximize $\zeta/{\gamma_p}^{1/p}$.
3. Ignoring the fact that this ignores odd-order moments, do you think this line of argument will improve upon Roy's original argument? (*Hint:*
$$\Pr\left\{x \le r_0\right\} \le \frac{1}{\zeta} \wedge \bigwedge_{p=2,4,\ldots} \left(\frac{{\gamma_p}^{1/p}}{\zeta}\right)^p,$$
and thus only the smallest bound matters. Estimate the bounds for $p = 2, 4, 6$ using empirical data from returns of *e.g.*, a broad market ETF.)

Ex. 2.19 Contours of Sharpe ratio In the (μ, σ)-space, plot contours of constant signal-noise ratio, assuming $r_0 = 0$. Plot them again assuming $r_0 = 0.01$.

Ex. 2.20 The Sharpe ratio as yardstick The Sharpe ratio is used as a metric to compare funds. However, it is often lamented that when the sample mean is negative, use of this metric 'prefers' higher volatility funds.

1. Suppose you observe, for two different funds, $\hat{\zeta}_1 < \hat{\zeta}_2 < 0$. Is there justification to prefer the second fund to the first?
2. Suppose you are omniscient, and know that $\zeta_1 < \zeta_2 < 0$. Is there any justified preference?

Ex. 2.21 Generalized Sharpe ratio Confirm that the ex-factor Sharpe ratio defined in Equation 2.10, for the case where the factors are only the constant one, corresponds to the traditional Sharpe ratio defined in Equation 2.1

* **Ex. 2.22 Approximate Sharpe ratio** For a given return, x, let $\alpha_2 = \mathrm{E}\left[x^2\right]$ be the uncentered second moment. Define

$$f_{\text{tas}}\left(x\right) =_{\text{df}} \tan\left(\arcsin\left(x\right)\right) = \frac{x}{\sqrt{1-x^2}}. \tag{2.24}$$

1. What is the domain of $f_{\text{tas}}\left(\cdot\right)$? Show that $\left|\mu\right| \le \sqrt{\alpha_2}$
2. Show that $\zeta = f_{\text{tas}}\left(\mu/\sqrt{\alpha_2}\right)$.
3. Compute the first derivative of $f_{\text{tas}}\left(\cdot\right)$. Via Taylor's theorem, derive the linear approximation to $f_{\text{tas}}\left(\cdot\right)$; call it $\bar{f}_{\text{tas}}\left(\cdot\right)$.
4. Derive the Lagrange form of the remainder to the linear approximation.
5. Assuming $\left|x\right| \le 0.1$, find an upper bound on the absolute error, $\left|f_{\text{tas}}\left(x\right) - \bar{f}_{\text{tas}}\left(x\right)\right|$. Find an upper bound on the absolute relative error $\left|1 - \bar{f}_{\text{tas}}\left(x\right)/f_{\text{tas}}\left(x\right)\right|$.
6. Having observed n returns $x_1, \ldots, x_n$, let the first two sample moments be

$$\hat{\mu} = \frac{1}{n}\sum_{1\le i\le n} x_i, \quad \hat{\alpha}_2 = \frac{1}{n}\sum_{1\le i\le n} {x_i}^2.$$

Show that $\hat{\zeta} = \sqrt{\frac{n-1}{n}} f_{\text{tas}}\left(\hat{\mu}/\sqrt{\hat{\alpha}_2}\right)$.

* **Ex. 2.23 Market Timing** Suppose that a fund engages in *market timing* on an underlying asset. That is, for some ϵ with $\left|\epsilon\right| \le 1$, the fund correctly guesses the sign of the next month's return on the market with probability $\frac{1+\epsilon}{2}$, and takes long or short positions accordingly.

1. What is the signal-noise ratio of the fund, computed on relative returns? Your answer should be expressed in terms of ϵ and

$$\kappa = \frac{\mathrm{E}\left[\left|x\right|\right]}{\sqrt{\mathrm{E}\left[x^2\right]}}.$$

2. Is the signal-noise ratio infinite for $\epsilon = 1$? Why or why not?
3. Compute the value of κ for the case where x follows a normal distribution with mean μ and standard deviation σ. For this value of κ plot ζ as a function of ϵ.
* 4. For monthly relative returns of the Market factor, $\kappa = 0.7166$. For this value of κ, plot ζ as a function of ϵ. Express ζ in annualized terms. What value of ϵ is required to achieve $\zeta = 1\mathrm{yr}^{-1/2}$? (See also, Sharpe on market timing. [150])
5. Compute the value of κ for the case where x follows a normal distribution with mean μ and standard deviation σ. For this value of κ plot ζ as a function of ϵ.

* **Ex. 2.24** **Correlation and Market Timing** Suppose you observe a random variable f_i, called 'a signal', prior to the time needed to make an investment decision to capture the relative return x_{i+1}. Suppose that x is zero mean, as is f_i, and the correlation between f_i and x_{i+1} is ρ.

1. Suppose that f_i takes only values ± 1. If you allocate f_i of your capital (long or short, depending on the sign of f_i) to the asset, show that the signal-noise ratio of your trading strategy, measured on relative returns, is $f_{\mathsf{tas}}(\rho)$, where $f_{\mathsf{tas}}(\cdot)$ is defined as in Equation 2.24.
* 2. Suppose instead that f_i and x_{i+1} are jointly normal, zero mean variables with correlation ρ. Suppose that you allocate cf_i, long or short, in the asset, where c is chosen to maintain risk constraints. Show that the signal-noise ratio of your trading strategy, measured on relative returns, is $\rho/\sqrt{1+\rho^2}$. (*Hint:* If $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, then $\operatorname{E}\left[\boldsymbol{y}^\top \mathsf{A} \boldsymbol{y}\right] = \operatorname{tr}(\mathsf{A}\Sigma) + \boldsymbol{\mu}^\top \mathsf{A} \boldsymbol{\mu}$, and $\operatorname{Var}\left(\boldsymbol{y}^\top \mathsf{A} \boldsymbol{y}\right) = 2\operatorname{tr}(\mathsf{A}\Sigma\mathsf{A}\Sigma) + 4\boldsymbol{\mu}^\top \mathsf{A}\Sigma\mathsf{A}\boldsymbol{\mu}$ for symmetric A.)
* 3. Suppose instead that f_i and x_{i+1} are general zero mean variables with finite fourth moments, and with correlation ρ, but make no further assumptions. Suppose that you allocate cf_i, long or short, in the asset, where c is chosen to maintain risk constraints. Give a lower bound on the signal-noise ratio of your trading strategy, measured on relative returns, in terms of the first four moments of f_i and x_{i+1}. How small can your bound on signal-noise ratio be for a given $\rho > 0$? (*Hint:* apply the Cauchy-Schwarz inequality to bound the second moment of your returns from above.)
** 4. Construct zero mean, finite variance random variables f_i and x_{i+1} with correlation $\rho > 0$, but such that the signal-noise ratio of the market timing portfolio (*i.e.*, the one with returns $cf_i x_{i+1}$, for some positive c) is arbitrarily small?
** 5. § Suppose that f_i and x_{i+1} are jointly normal, zero mean variables with correlation ρ, and the variance of f_i is known. Can you construct a trading strategy with higher signal-noise ratio than the one suggested in the previous question? You may assume the variance of both f_i and x_{i+1} are known.
** 6. § Suppose that f_i and x_{i+1} are zero mean random variables with correlation ρ, and known variances. Can you find the 'maximin' trading strategy, and its signal-noise ratio? That is, find the trading strategy conditional on f_i that has maximal worst-case signal-noise ratio over all random variables with the given properties.

* **Ex. 2.25** **Bid/Ask Bounce** One commonly rediscovered, putative trading strategy is the phantom bid/ask bounce trade. This supposed anomaly is found by assuming that one can simultaneously observe and trade in *e.g.*, an auction. This odd assumption typically follows from using low frequency data for backtest simulations.

Assume that the price quoted in the data feed is either at the bid or the ask, that is, from a market participant crossing the spread to sell to, or buy the asset from, a liquidity provider. Let the price p_t be given by $\log p_t = \log m_t + b_t$, where the log of midpoint m_t is a zero mean random walk, and b_t, the (log) bid/ask bounce, takes values of $\pm\epsilon_b$. The presence of the bid/ask bounce causes negative autocorrelation of the log returns $x_{t+1} = \log p_{t+1}/p_t$. If it was assumed one could observe and trade on price p_t, a mean reverting strategy would appear profitable, when in reality the effect cannot be captured.

1. Let the midpoint log returns, $\log m_{t+1} - \log m_t$ be *i.i.d.* and have zero mean and variance σ_m^2. Let the b_t be *i.i.d.*, and let b_t be independent of $m_1, \ldots, m_{t+1}$. Show that the total log returns $x_{t+1} = \log p_{t+1} - \log p_t$ have autocorrelation of

$$\rho = \frac{-\epsilon_b^2}{2\epsilon_b^2 + \sigma_m^2}$$

2. Although it is only an approximation, and not completely correct (see Exercise 2.24, and also note this question deals with log returns, not relative), assume that the trading strategy which allocates $-cx_t$ capital to the asset achieves a signal-noise ratio of $-\rho$. How does the annualized signal-noise ratio scale with the time period Δt between trades? You may assume that the bid/ask bounce are *i.i.d.* at all scales.
3. Estimate σ_m^2 and ϵ_b from bid/ask data for a real asset. (For example, use the data sample for OEX options provided by Market Data Express, `http://www.marketdataexpress.com/User_Data/Files/mdr_20070516_OEB.csv.gz` .) At what time scale would the mean reverting strategy have a signal-noise ratio of $4\text{yr}^{-1/2}$?
4. Compute ρ in the case that the bid/ask variable is correlated to the midpoint log returns. Let η be the correlation of $\log m_t - \log m_{t-1}$ and b_t. For more on microstructure models see *e.g.*, Hasbrouck. [67]

* **Ex. 2.26 The signal-noise ratio of a sum** Let x and y be independent random variables. Let μ_x, σ_x, ζ_x be the mean, standard deviation and signal-noise ratio of x, and so on.

1. For some constants a, b, let $z = ax + by$. Derive an expression for ζ_z in terms of μ_x, σ_x, μ_y, σ_y, a and b.
2. Show that ζ_z is scale invariant with respect to a and b. That is, for any $c > 0$ $ax + by$ and $cax + cby$ have the same signal-noise ratio.
3. Find the a, b which maximize ζ_z. Because of the scale invariance, these can be determined only up to scale, so choose a and b such that $a^2 + b^2 = 1$. (This is a basic *portfolio optimization* problem, of which more in the sequel.)
4. Find ζ_z for the optimal a, b.
5. In the above, a and b were denominated in dollar-proportional units. Instead, we could denominate them in units of risk. That is, define $z = (a/\sigma_x)\,x + (b/\sigma_y)\,y$. Find the a and b which maximize ζ_z. To bypass the scale invariance, express your answer as the ratio a/b. What is the takeaway rule of thumb for portfolio optimization?

* **Ex. 2.27 The signal-noise ratio of a product** Let $z = xy$, where x and y are independent random variables. Let ζ_i be the signal-noise ratio of variable i.

1. Derive an expression for ζ_z entirely in terms of ζ_x and ζ_y.
2. Suppose that y is the level of the VIX index. [27] Then $\zeta_y \approx 2.4677\,\mathrm{day}^{-1/2}$, based on data from 1990-01-02 to 2018-12-31. [28] Suppose $\zeta_x = 0.1\,\mathrm{day}^{-1/2}$. Compute ζ_z.
3. Compute the derivative
$$\frac{\mathrm{d}\zeta_z}{\mathrm{d}\zeta_y}.$$
Does ζ_z have a local maximum with respect to ζ_y?
4. Find
$$\lim_{\zeta_y \to \infty} \zeta_z$$
5. Viewing y as some kind of random 'leverage' on the strategy with returns x, what is the best case for an investor in z?

* **Ex. 2.28 Autocorrelated returns and higher order moments** Assume that daily log returns, x, follow an AR(1) process: $x_{t+1} - \mu = \rho\,(x_t - \mu) + \epsilon_{t+1}$, where ϵ_t are independent zero-mean error terms with standard deviation σ. Let y be the annualized[18] log returns of the same process. So, for some fixed n, the number of days in an epoch, let $y_t = \sum_{(t-1)n<i\le tn} x_i$.

1. What is the mean of y in terms of the mean of x, n and ρ?
2. What is the standard deviation of y in terms of the standard deviation of x, n and ρ?
3. * What is the third moment of y in terms of the first three moments of x, n and ρ? What is the skew of y?
4. ** What is the fourth moment of y in terms of the first four moments of x, n and ρ? What is the excess kurtosis of y?

* **Ex. 2.29 Autocorrelated returns and signal-noise ratio** In Section 2.2, the method for annualizing the Sharpe ratio was given *for independent returns.* As in Exercise 2.28, assume daily log returns, x, follow an AR(1) process: $x_{t+1} - \mu = \rho\,(x_t - \mu) + \epsilon_{t+1}$, where ϵ_t are independent zero-mean error terms with standard deviation σ. So, for some fixed n, the number of days in an epoch, let $y_t = \sum_{(t-1)n<i\le tn} x_i$.

1. Compute the signal-noise ratio of x. [99]
2. Compute the signal-noise ratio of y.
3. Compute the ratio of the signal-noise ratio of y to $\sqrt{n}$ times the signal-noise ratio of x. Plot this ratio versus the signal-noise ratio of x for $n = 252$, using $\rho = 0.1$. Plot it against ρ assuming the signal-noise ratio of x is $0.05\mathrm{day}^{-1/2}$.

[18] Or weekly, or monthly, whatever.

Ex. 2.30 Attribution fits Consider the monthly returns of the Fama-French factor data from **aqfb.data**, using code as given in Example 1.2.1. Each of the following asks you to perform a factor attribution. For each of them, compute the ex-factor Sharpe ratio.

1. The 'Market' returns have the risk-free rate subtracted from them, but there may be residual exposure of these *excess returns* to the risk-free rate. Perform an attribution of the excess returns of the market to the risk free rate, computing the ex-factor Sharpe ratio.
2. Perform a CAPM attribution on the SMB portfolio returns. For simplicity, use the excess returns of the Market portfolio as the market.
3. Perform a CAPM attribution on the HML portfolio returns.
4. Perform a Chow-test like attribution on the excess returns of the Market portfolio, with an indicator factor that is one prior to 1960.
5. Perhaps it is the case that the small cap premium is disappearing, and does not currently exist. Test this by attributing the monthly returns of the SMB portfolio against a factor which is linearly decreasing over time, taking value of 1 at the start of the data, and value 0 at the end.
6. Can all the excess returns of the Market be attributed to the 'January Effect'? Perform an attribution against an indicator variable that is 1 exactly in January, and 0 elsewhere.

Ex. 2.31 Geometric utility Suppose μ and σ^2 are the mean and variance of the simple returns of a strategy, x. Show that $\mathrm{E}\left[\log\left(1+x\right)\right] \approx \mu - \frac{1}{2}\sigma^2$, the mean-variance objective function.

Ex. 2.32 Other objectives Construct several objective functions on the population parameters of returns distributions which sound like plausible objectives for an investor. Which of them are consistent with first order stochastic dominance?

* **Ex. 2.33 The signal-noise ratio and mean-variance preferences** Suppose x is a returns stream with mean $\mu > 0$ and standard deviation σ. Let y be a contingent claim (a 0/1 Bernoulli random variable) which pays out with probability p, independent of x. Suppose $z = x + y$.

1. Show that z (first order) stochastically dominates x. That is, show that, for all c, $\Pr\left\{x \ge c\right\} \le \Pr\left\{z \ge c\right\}$.
2. Find the signal-noise ratio of z.
3. Find conditions on μ, σ^2, p such that the signal-noise ratio of z is *less* than that of x.

* **Ex. 2.34 Is the signal-noise ratio monotonic?** Let x and y be paired random variables.

1. Suppose that $\Pr\left\{x_i \le y_i\right\} = 1$ for all i. Let ζ_x, and ζ_y be the signal-noise ratios of these variables. Can you prove that $\zeta_x \le \zeta_y$? If not, can you find a

counterexample?

2. Suppose you observe n paired observations of x and y and notice that $x_i \le y_i$ for all i. You measure their Sharpe ratios, $\hat{\zeta}_x$ and $\hat{\zeta}_y$. Must it be the case that $\hat{\zeta}_x \le \hat{\zeta}_y$? If not, can you find a counterexample?

Ex. 2.35 Approximate normality Load the daily returns of 'the Market' from **aqfb.data**. Resample, with replacement, 250 days of market returns, and compute the mean. Repeat this 5000 times, and Q-Q plot the means against normality.

1. What implications does this have for the argument that signal-noise ratio is an inappropriate objective because it assumes returns are normal?

Ex. 2.36 Higher order signal-noise ratio Consider the higher order signal-noise ratio, defined in Equation 2.17.

1. Show that the higher order signal-noise ratio is consistent with first-order stochastic dominance.
2. Is it necessarily consistent with second-order stochastic dominance?

Ex. 2.37 Solving for higher order signal-noise ratio Consider the approximate higher order signal-noise ratio given by Equation 2.19.

1. Find the roots of the equation

$$\zeta_h = \frac{(\mu - r_0)}{\sigma} + \frac{\gamma_1}{6}\left(\zeta_h^2 - \frac{1}{n}\right).$$

2. Which root do you suspect gives a closer approximation to the higher order signal-noise ratio defined in Equation 2.17?
3. Suppose $(\mu - r_0)/\sigma = 0.1\mathrm{day}^{-1/2}$, $\gamma_1 = 1.5$ and $n = 252\mathrm{day}$. Find the approximate higher order signal-noise ratio.

* **Ex. 2.38 Sample variance of alternative metrics** Explore the sample variation of some alternative metrics, as compared to the Sharpe ratio. For example, the Calmar ratio: draw 1 year of daily returns from a population with Gaussian log returns, using $\mu = 0\mathrm{day}^{-1}, \sigma = 0.01\mathrm{day}^{-1/2}$, then compute the Calmar ratio on the returns. Also compute the Sharpe ratio. Repeat this 1000 times. Then do the same using $\mu = 0.001\mathrm{day}^{-1}, \sigma = 0.01\mathrm{day}^{-1/2}$. Estimate, empirically, the power of the test with 0.05 type I rate via the samples of the statistics on these two populations. The Calmar ratio, is provided in **R** by the **PerformanceAnalytics** package, as are all the alternative metrics mentioned below, while the Sharpe ratio can be computed via the **SharpeR** package. [135, 122]

1. Repeat this exercise for the Sterling ratio.
2. Repeat this exercise for the 'upside potential ratio'.
3. Repeat this exercise for the maximum drawdown. (You will have to turn the log returns series into a price series to compute drawdown.)

4. Repeat this exercise for the so-called "Fano ratio," defined as $\hat{\mu}/\hat{\sigma}^2$. [78] (See also Exercise 2.42.)

* **Ex. 2.39 Drawdown and distribution** Draw 2 years of daily log returns from the Gaussian distribution with $\mu = 0\text{day}^{-1}, \sigma = 0.01\text{day}^{-1/2}$. Turn the returns series into a price series, then compute the maximum drawdown. Repeat this 10,000 times. Compute the 0.95 empirical quantile.

1. Repeat this exercise but using draws from a shifted, rescaled t-distribution with 4 degrees of freedom. You need to make sure you have achieved the proper mean and variance.
2. Repeat this exercise but using draws from a shifted, rescaled, 'Lambert W $\times$ Gaussian' distribution using $\beta = 0.1$. [57] (The **LambertW** package is recommended for this task. [59])
3. How sensitive is the critical maximum drawdown to the distribution from which returns are drawn?

* **Ex. 2.40 Drawdown and distribution II** Repeat the experiment of Example 2.7.2, but draw returns from a shifted, rescaled t-distribution with 4 degrees of freedom. How do the half space constraint regression coefficients change?

Ex. 2.41 Drawdown and distribution III Rej *et al.* note a connection between the length and depth of the current drawdown and the signal-noise ratio. [141] Perform simulations to confirm their findings. Here the 'current drawdown' is not the same as the maximum drawdown, rather it is the difference between the maximum fund value over a history and the recent value. So let x_i be *i.i.d.* Gaussian log returns and let

$$p_i = e^{\sum_{1 \le j \le i} x_j}.$$

The current drawdown at time t is defined as

$$\max_{1 \le i \le t} p_i - p_t.$$

Simulate 2 years of weekly data using volatility of $\sigma = 0.04\text{wk.}^{-1/2}$, and compute the current drawdown at the end of the 2 year period. Repeat this 1000 times for each choice of $\mu = 0, 0.001\text{wk.}^{-1}, \ldots, 0.010\text{wk.}^{-1}$.

1. Compute the median drawdown for each choice of μ and scatter plot against ζ^{-1}. Do you see a linear fit?

* **Ex. 2.42 A Drawdown estimator of signal-noise ratio** Challet introduced an estimator of the signal-noise ratio based on the number of drawdowns (and draw ups) of an asset's price level. [30] This estimator is purported to have lower standard error (and by implication, lower mean square error) than the Sharpe ratio when returns are drawn from leptokurtic distributions.

Test this assertion empirically. The code to compute the drawdown-based estimator is available in the **R** package, **sharpeRratio**. Draw 1000 days of returns from a t

distribution, and compute the Sharpe ratio and the drawdown estimator, then compute the error of each. Repeat this 10000 times. Perform this experiment for the case where the signal-noise ratio is $0.5\mathrm{yr}^{-1/2}$ and again when it is $2\mathrm{yr}^{-1/2}$. For the returns distribution vary the degrees of freedom of the t distribution, taking it to be 8, or 4.25. You will compute estimates of the mean square error of each of these two estimators for 4 different parameter settings.

1. Does the drawdown-estimator consistently have lower mean square error? If not, why not?
2. What do you think should happen when returns are normally distributed? Repeat the experiment with normal returns to find out.

Ex. 2.43 **Mr. Buffett's Bet** In 2007 Warren Buffett made a million dollar bet with Ted Seides that the Vanguard S&P 500 index fund would outperform the average return of five funds of hedge funds returns over a ten year period. Mr. Buffett prevailed in this bet. [25, 1] Let us consider Mr. Seides' odds.

We downloaded the BarclayHedge Equity Market Neutral Index data, then joined to the monthly Market returns data from the Fama French factor returns. In total we consider 240 months of data from 1997-01-01 to 2016-12-01. Supposing that the Hedge Fund returns were similar to those of the Market Neutral Index (they weren't), and that the S&P fund from Vanguard tracked the Market (a decent approximation), consider the relative returns of one dollar long on Market and one dollar short on the Market neutral index.

Over the sample period, this combined strategy had an empirical mean return of $0.3\%\,\mathrm{mo.}^{-1}$ and a standard deviation of $4.5\%^{-1/2}\,\mathrm{mo.}^{-1/2}$.

1. What is the probability that a normally distributed random variable with mean 0.3 and standard deviation 4.5 will be positive?
2. What is the probability that the sum of 120 independent normally distributed random variables with mean 0.3 and standard deviation 4.5 will be positive?
3. We have taken the sample standard deviation of the difference in returns of the two series, which exhibit sizeable correlation over the sample period (around 0.21). Instead consider the influence of correlation. The two returns series have volatilities of $0.88\%^{-1/2}\,\mathrm{mo.}^{-1/2}$ and $4.6\%^{-1/2}\,\mathrm{mo.}^{-1/2}$. Suppose the correlation between them is 0. What is the volatility of their difference? How would this affect the probability that their difference, assumed normal, would have positive sum over 120 months?
4. Suppose instead that the Fund of Funds has zero expected mean and that the Market has mean return of $0.74\%\,\mathrm{mo.}^{-1}$, and their difference has standard deviation $4.5\%^{-1/2}\,\mathrm{mo.}^{-1/2}$. What is the probability that their difference will be positive over 120 months assuming normal returns?

3. The Sharpe ratio for Normal Returns

If the track is tough and the hill is rough,
THINKING you can just ain't enough!

(Shel Silverstein, The Little Blue Engine*)*

An intellectual is a man who doesn't know how to park a bike.

(Spiro Agnew, attributed*)*

While normality of returns is a *terrible* model for most market instruments [36], it is a terribly convenient model. For the rest of this chapter, then, unless stated otherwise, we will adhere to this terrible model and assume returns are unconditionally *i.i.d.* Gaussian, *i.e.*, $x_t \sim \mathcal{N}\left(\mu, \sigma^2\right)$. In the sequel, we will examine the consequences of assuming normality, independence, homoskedasticity, *etc.*, and correct for them when possible.

The Sharpe ratio is, up to a scaling, the Student t-statistic for testing the mean of a (typically Gaussian) random variable. Sharpe himself never mentions this relationship, although he quotes regression fit t-statistics in his original paper. [149] This connection to t-statistics, perhaps first noted by Miller and Gehr [112], allows us to translate known statistical results from testing of the mean to testing of the signal-noise ratio.

3.1. The non-central t-distribution

Let Z be a random variable following a normal distribution with mean δ and variance 1, which we write as $Z \sim \mathcal{N}(\delta, 1)$. Let X, independent of Z take a chi-square distribution with ν degrees of freedom, written $X \sim \chi^2(\nu)$. Then the variable

$$t = \frac{Z}{\sqrt{X/\nu}}$$

follows a *non-central t distribution* with non-centrality parameter δ and ν degrees of freedom. [77, 147, 120] We write this as $t \sim t(\nu, \delta)$. As a special case, when $\delta = 0$, t follows a (central) t distribution. (The non-central t is a special case of the more general *doubly* non-central t distribution, wherein X follows a non-central chi-square distribution.)

Suppose now that $Z \sim \mathcal{N}\left(\mu_0, \sigma^2/\nu\right)$ independently of $\left(\nu - 1\right) X/\sigma^2 \sim \chi^2\left(\nu - 1\right)$. Then

$$\sqrt{\nu}\frac{Z - \mu_1}{\sqrt{X}} \sim t\left(\delta = \sqrt{\nu}\frac{\mu_0 - \mu_1}{\sigma}, \nu - 1\right). \tag{3.1}$$

The non-centrality parameter δ is $\sqrt{\nu}$ times the difference in signal-noise ratio values, μ_0/σ and μ_1/σ.

3.1.1. Distribution of the Sharpe ratio

Let $x_1, x_2, \ldots, x_n$ be *i.i.d.* draws from a normal distribution $\mathcal{N}\left(\mu, \sigma\right)$. 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 statistics. It is well known that $\hat{\mu}$ and $\hat{\sigma}$ are independent, $\hat{\mu} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$, and $\left(n-1\right)\hat{\sigma}^2/\sigma^2 \sim \chi^2\left(n-1\right)$. [11, 155] This gives the distribution of the Sharpe ratio under the assumption of Gaussian returns. That is

$$\sqrt{n}\hat{\zeta} = \sqrt{n}\frac{\hat{\mu} - r_0}{\hat{\sigma}} \sim t\left(\sqrt{n}\frac{\mu - r_0}{\sigma} = \sqrt{n}\zeta, n-1\right). \tag{3.2}$$

Note the non-centrality parameter, $\delta = \sqrt{n}\left(\mu - r_0\right)/\sigma$, looks like the sample statistic $\sqrt{n}\left(\hat{\mu} - r_0\right)/\hat{\sigma}$, but defined with population quantities. Informally, it is the 'population analogue' of the sample statistic.

3.1.2. Distribution of the ex-factor Sharpe ratio

Now consider the ex-factor Sharpe ratio introduced in Equation 2.10 of Section 2.4. Suppose that $x_1, x_2, \ldots, x_n$ are the observed returns, and that $\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_n$ are the corresponding factors in a factor model. Let F be the $n \times l$ matrix whose rows are the vectors $\boldsymbol{f}_t{}^{\top}$. Assume the model is properly specified: conditional on observing $\boldsymbol{f}_t$,

$$x_t = \boldsymbol{f}_t{}^{\top}\beta + \epsilon_t,$$

where the errors are *i.i.d.* normal, $\epsilon_t \sim \mathcal{N}\left(0, \sigma\right)$. The multiple linear regression estimates are

$$\hat{\boldsymbol{\beta}} =_{\mathrm{df}} \left(\mathsf{F}^{\top}\mathsf{F}\right)^{-1}\mathsf{F}^{\top}\boldsymbol{x}, \qquad \hat{\sigma} =_{\mathrm{df}} \sqrt{\frac{\left(\boldsymbol{x} - \mathsf{F}\hat{\boldsymbol{\beta}}\right)^{\top}\left(\boldsymbol{x} - \mathsf{F}\hat{\boldsymbol{\beta}}\right)}{n - l}}. \tag{3.3}$$

Then, conditional on observing F,

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N}\left(\boldsymbol{\beta}, \sigma^2\left(\mathsf{F}^{\top}\mathsf{F}\right)^{-1}\right), \quad \left(n - l\right)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2\left(n - l\right),$$

and the estimators are independent. [11, 140]

The ex-factor Sharpe ratio is then distributed as

$$\hat{\zeta}_g =_{\mathrm{df}} \frac{\hat{\boldsymbol{\beta}}^\top \boldsymbol{v} - r_0}{\hat{\sigma}} \sim \frac{\mathcal{N}\left(\boldsymbol{\beta}^\top \boldsymbol{v} - r_0, \sigma^2 \boldsymbol{v}^\top \left(\mathsf{F}^\top \mathsf{F}\right)^{-1} \boldsymbol{v}\right)}{\sqrt{\sigma^2 \chi^2\left(n-l\right)/\left(n-l\right)}},$$
$$\sim \sqrt{\boldsymbol{v}^\top \left(\mathsf{F}^\top \mathsf{F}\right)^{-1} \boldsymbol{v}} \frac{\mathcal{N}\left(\left(\boldsymbol{v}^\top \left(\mathsf{F}^\top \mathsf{F}\right)^{-1} \boldsymbol{v}\right)^{-1/2} \frac{\boldsymbol{\beta}^\top \boldsymbol{v} - r_0}{\sigma}, 1\right)}{\sqrt{\chi^2\left(n-l\right)/\left(n-l\right)}},$$

and therefore

$$\left(\boldsymbol{v}^\top \left(\mathsf{F}^\top \mathsf{F}\right)^{-1} \boldsymbol{v}\right)^{-1/2} \hat{\zeta}_g \sim t\left(\left(\boldsymbol{v}^\top \left(\mathsf{F}^\top \mathsf{F}\right)^{-1} \boldsymbol{v}\right)^{-1/2} \zeta_g, n-l\right),$$
$$\text{where } \zeta_g =_{\mathrm{df}} \frac{\boldsymbol{\beta}^\top \boldsymbol{v} - r_0}{\sigma}. \tag{3.4}$$

Compare this with Equation 3.2. Most of the results we will explore concerning the Sharpe ratio have equivalent results for the ex-factor Sharpe ratio. However, keeping track of the scaling parameter $\left(\boldsymbol{v}^\top \left(\mathsf{F}^\top \mathsf{F}\right)^{-1} \boldsymbol{v}\right)^{-1/2}$ would be unweildy, so generalizing the results will be left as an exercise for the reader.

While many of the results in this chapter have analagous forms for the ex-factor Sharpe ratio, strictly speaking they only apply to the case where the factors $\boldsymbol{f}_t$ are *deterministic*, and controlled by the experimenter. This is decidedly *not* the case for many models in which attribution will be performed, *e.g.*, CAPM, the Fama-French models, Henriksson-Merton, *etc.* Indeed, few, if any, of the models listed in Section 2.4.1 feature deterministic $\boldsymbol{f}_t$. The issue is that the randomness of the $\boldsymbol{f}_t$ contributes to estimation error. Methods for dealing with attribution for random covariates will be considered in Chapter 4.

3.2. † Density and distribution of the Sharpe ratio

Since the Sharpe ratio is distributed as a non-central t up to scaling, its density, cumulative distribution, and quantile functions can be defined in terms of their non-central t counterparts, as follows:

$$\begin{aligned} f_{SR}\left(\hat{\zeta}; \zeta, n\right) &= n^{-1/2} f_t\left(\sqrt{n}\hat{\zeta}; \sqrt{n}\zeta, n-1\right), \\ F_{SR}\left(\hat{\zeta}; \zeta, n\right) &= F_t\left(\sqrt{n}\hat{\zeta}; \sqrt{n}\zeta, n-1\right), \\ SR_p\left(\zeta, n\right) &= n^{-1/2} t_p\left(\sqrt{n}\zeta, n-1\right). \end{aligned} \tag{3.5}$$

Here $f_{SR}\left(\hat{\zeta}; \zeta, n\right)$ is the density (or PDF) of the Sharpe ratio distribution with signal-noise ratio ζ on n samples; $F_{SR}\left(\hat{\zeta}; \zeta, n\right)$ is the cumulative distribution (or CDF); and

$SR_p(\zeta, n)$ is the p^{th} quantile. The functions on the right hand side of the equations are the PDF, CDF, and quantile of the non-central t distribution.

The PDF, CDF, and quantile functions of the non-central t have stock implementations. For example, in **R**, these are available as `dt`, `pt`, and `qt`. As a practical matter, these should be used instead of writing your own. *If your interests are practical, rather than theoretical, you can safely skip the rest of this section.*

3.2.1. The PDF of the Sharpe ratio

First, let us compute the density of the Sharpe ratio under *i.i.d.* normal returns. Following Walck [165], let $\hat{\zeta} = z/\sqrt{x}$, where $z \sim \mathcal{N}\left(\zeta, \frac{1}{n}\right)$ independently of $x \sim \chi^2(n-1)$. The joint density of x, z is

$$f(x, z; \zeta, n) = \left[\frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} x^{\frac{n-1}{2}-1} e^{-\frac{x}{2}}\right]\left[\sqrt{\frac{n}{2\pi}} e^{-\frac{n(z-\zeta)^2}{2}}\right]. \tag{3.6}$$

Now we make the transformation $[x, z]^\top \to [x, z/\sqrt{x}]^\top = \left[x, \hat{\zeta}\right]^\top$. The determinant of the Jacobian of the inverse transform is $\sqrt{x}$. Thus

$$f\left(x, \hat{\zeta}; \zeta, n\right) = \sqrt{x} f(x, z; \zeta, n) = \sqrt{x} f\left(x, \hat{\zeta}\sqrt{x}; \zeta, n\right).$$

To get the density of $\hat{\zeta}$, integrate out x. This gives the integral form of the density as

$$\begin{aligned} f_{SR}\left(\hat{\zeta}; \zeta, n\right) &= \int_0^\infty \sqrt{x}\left[\frac{x^{\frac{n-1}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}\right]\left[\sqrt{\frac{n}{2\pi}} \exp\left(-\frac{n\left(\hat{\zeta}\sqrt{x}-\zeta\right)^2}{2}\right)\right] \mathrm{d}x, \\ &= \frac{\sqrt{n/2}}{\sqrt{2\pi}\Gamma\left(\frac{n-1}{2}\right)} \int_0^\infty \left(\frac{x}{2}\right)^{\frac{n-2}{2}} \exp\left(-\frac{x}{2} - \frac{n\left(\hat{\zeta}\sqrt{x}-\zeta\right)^2}{2}\right) \mathrm{d}x. \end{aligned} \tag{3.7}$$

3.2.2. The CDF and quantile of the Sharpe ratio

Finding the CDF of the Sharpe ratio consists of 'simply' integrating the PDF given in Equation 3.7. Witkovsky gives a few different forms for the CDF of the non-central t distribution. [172] The equivalent formulations for the Sharpe ratio distribution are as follows, for $\hat{\zeta} > 0$,

$$\begin{aligned}F_{SR}\left(\hat{\zeta};\zeta,n\right) &= \frac{1}{\sqrt{n-1}}\int_0^{\infty}\Phi\left(\hat{\zeta}\sqrt{\frac{n}{n-1}}z - \sqrt{n}\zeta\right)f_{\chi^2}\left(z;n-1\right)\,\mathrm{d}z,\\
&= \Phi\left(-\sqrt{n}\zeta\right) + \int_{-\sqrt{n}\zeta}^{\infty}\left(1 - F_{\chi^2}\left(\frac{(n-1)\left(z+\sqrt{n}\zeta\right)^2}{n\hat{\zeta}^2};n-1\right)\right)\phi\left(z\right)\,\mathrm{d}z,\\
&= \Phi\left(-\sqrt{n}\zeta\right) + \frac{1}{2}\sum_{i=0}^{\infty}\left[\left(\frac{n\zeta^2}{2}\right)^i\frac{e^{-n\zeta^2/2}}{i!}\mathrm{I}_{\frac{n\hat{\zeta}^2}{n-1+n\hat{\zeta}^2}}\left(i+\frac{1}{2},\frac{n-1}{2}\right)\right]\\
&+ \frac{1}{2}\sum_{i=0}^{\infty}\left[\left(\frac{n\zeta^2}{2}\right)^{i+1/2}\frac{e^{-n\zeta^2/2}}{\Gamma\left(i+3/2\right)}\mathrm{I}_{\frac{n\hat{\zeta}^2}{n-1+n\hat{\zeta}^2}}\left(i+1,\frac{n-1}{2}\right)\right].\end{aligned} \tag{3.8}$$

Here $\mathrm{I}_x\left(a,b\right)$ is the incomplete beta function [2, 6.6.2], and $F_{\chi^2}\left(x;\nu\right)$ and $f_{\chi^2}\left(x;\nu\right)$ are, respectively, the CDF and PDF of the chi-square distribution with ν degrees of freedom. The connection between the non-central t distribution and the lambda prime distribution (*cf.* Section 3.4) are evident in some of these forms. The last form is essentially that used in the standard computation of the CDF of the t distribution via 'AS 243'. [94, 62]

The inverse CDF, or quantile function, of the non-central t-distribution is not easily expressed in compact notation. The 'exact' computation, provided by *e.g.*,`qt` in **R**, is due to Hill. [69, 70] Akahira *et al.* present an approximation to the quantile function, first described by Akahira, based on the Cornish Fisher expansion[1]. [3, 4] The Akahira approximation generalizes the quantile approximation of Johnson and Welch. [77]

3.3. Moments of the Sharpe ratio

The moments of the non-central t-distribution are known, and can easily be translated into those of the Sharpe ratio. [72, 165, 171] Suppose that $t \sim t\left(\delta, n-1\right)$. Then, for $0 \le i < n-1$,

$$\mathrm{E}\left[t^i\right] = \left(\frac{n-1}{2}\right)^{i/2}\frac{\Gamma\left(\frac{n-1-i}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\left(e^{-\delta^2/2}\frac{\mathrm{d}^i e^{\delta^2/2}}{\mathrm{d}\delta^i}\right). \tag{3.9}$$

The quantity in parenthesis is a variant of the 'probabilist's Hermite polynomial', but has all positive coefficients in δ. [2] Since $\sqrt{n}\hat{\zeta} \sim t\left(\sqrt{n}\zeta, n-1\right)$ under our assumption of normal returns, the moments of the Sharpe ratio are

$$\alpha_i\left(\hat{\zeta}\right) =_{\mathrm{df}} \mathrm{E}\left[\hat{\zeta}^i\right] = \left(\frac{n-1}{2}\right)^{i/2}\frac{\Gamma\left(\frac{n-1-i}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\frac{1}{n^{i/2}}e^{-n\zeta^2/2}\frac{\mathrm{d}^i e^{n\zeta^2/2}}{\mathrm{d}\left(\sqrt{n}\zeta\right)^i}, \tag{3.10}$$

[1] Akahira *et al.* use the inversion that connects the t-distribution to the Lambda prime distribution, see Section 3.4.

and so

$$\begin{aligned}
\alpha_1\left(\hat{\zeta}\right) &= \left(\frac{n-1}{2}\right)^{1/2}\frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{n^{\frac{1}{2}}\zeta}{n^{\frac{1}{2}}}\right) = d_n\zeta,\\
\alpha_2\left(\hat{\zeta}\right) &= \left(\frac{n-1}{2}\right)\frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{1+n\zeta^2}{n}\right) = \frac{n-1}{n-3}\left(\frac{1+n\zeta^2}{n}\right),\\
\alpha_3\left(\hat{\zeta}\right) &= \left(\frac{n-1}{2}\right)^{3/2}\frac{\Gamma\left(\frac{n-4}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{3n^{\frac{1}{2}}\zeta + n^{\frac{3}{2}}\zeta^3}{n^{\frac{3}{2}}}\right) = \frac{n-1}{n-4}d_n\frac{\zeta}{n}\left(3+n\zeta^2\right),\\
\alpha_4\left(\hat{\zeta}\right) &= \left(\frac{n-1}{2}\right)^{2}\frac{\Gamma\left(\frac{n-5}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{3+6n\zeta^2+n^2\zeta^4}{n^2}\right) = \frac{(n-1)^2}{(n-3)(n-5)}\left(\frac{3+6n\zeta^2+n^2\zeta^4}{n^2}\right),\\
\alpha_5\left(\hat{\zeta}\right) &= \left(\frac{n-1}{2}\right)^{5/2}\frac{\Gamma\left(\frac{n-6}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{15n^{\frac{1}{2}}\zeta + 10n^{\frac{3}{2}}\zeta^3 + n^{\frac{5}{2}}\zeta^5}{n^{\frac{5}{2}}}\right),\\
&= \frac{(n-1)^2}{(n-4)(n-6)}d_n\frac{\zeta}{n^2}\left(15+10n\zeta^2+n^2\zeta^4\right),
\end{aligned} \tag{3.11}$$

where we have defined

$$d_n =_{\mathrm{df}} \sqrt{\frac{n-1}{2}}\frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}, \tag{3.12}$$

as the 'bias term'. The even moments enjoy cancellation due to definition of the Gamma function, while the odd moments have this irreducible constant, d_n. Thus we have

$$\begin{aligned}
\mathrm{E}\left[\hat{\zeta}\right] &= d_n\zeta,\\
\mathrm{Var}\left(\hat{\zeta}\right) &= \frac{(1+n\zeta^2)(n-1)}{n(n-3)} - \left(d_n\zeta\right)^2,\\
\mathrm{skew}\left(\hat{\zeta}\right) &= \frac{\alpha_3\left(\hat{\zeta}\right) - 3\alpha_2\left(\hat{\zeta}\right)\alpha_1\left(\hat{\zeta}\right) + 2\alpha_1\left(\hat{\zeta}\right)^3}{\left(\alpha_2\left(\hat{\zeta}\right) - \alpha_1\left(\hat{\zeta}\right)^2\right)^{3/2}},\\
\mathrm{ex\,kurtosis}\left(\hat{\zeta}\right) &= \frac{\alpha_4\left(\hat{\zeta}\right) - 4\alpha_3\left(\hat{\zeta}\right)\alpha_1\left(\hat{\zeta}\right) + 6\alpha_2\left(\hat{\zeta}\right)\alpha_1\left(\hat{\zeta}\right)^2 - 3\alpha_1\left(\hat{\zeta}\right)^4}{\left(\alpha_2\left(\hat{\zeta}\right) - \alpha_1\left(\hat{\zeta}\right)^2\right)^2} - 3.
\end{aligned} \tag{3.13}$$

Example 3.3.1 (Cumulants of the Sharpe ratio). Suppose you observe 60 months of Gaussian log returns with $\zeta = 0.3\,\mathrm{mo.}^{-1/2}$. Then $\mathrm{E}\left[\hat{\zeta}\right] = 0.3039\,\mathrm{mo.}^{-1/2}$, $\mathrm{Var}\left(\hat{\zeta}\right) = 0.0181\,\mathrm{mo.}^{-1}$, $\mathrm{skew}\left(\hat{\zeta}\right) = 0.1202$, $\mathrm{ex\,kurtosis}\left(\hat{\zeta}\right) = 0.1288$. The skewness and excess kurtosis are unitless quantities by definition.

To check these calculations, 10^7 samples from the Sharpe ratio distribution with $\zeta = 0.3$, mo.$^{-1/2}$ and $n = 60$ months were drawn. The empirical mean of the Sharpe ratio was $0.3039\,\mathrm{mo.}^{-1/2}$, the empirical variance was $0.0181, \mathrm{mo.}^{-1}$, and the empirical skew and excess kurtosis were measured to be 0.1213, and 0.1291. ⊣

3.3.1. The Sharpe ratio is biased

The geometric bias term, d_n, is related to the constant c_4 from the statistical control literature via

$$d_n = \frac{n-1}{n-2} c_4\left(n\right).$$

The bias term does not equal one, thus the Sharpe ratio is a biased estimator of the signal-noise ratio (when it is nonzero), as first described by Miller and Gehr. [112, 76] The bias is multiplicative and larger than one, so the Sharpe ratio will overestimate the signal-noise ratio when the latter is positive, and underestimate it when it is negative. The bias term is a function of sample size only, and approaches one fairly quickly. A A decent asymptotic approximation [38] to d_n is given by

$$d_{n+1} = 1 + \frac{3}{4n} + \frac{25}{32n^2} + \mathcal{O}\left(n^{-3}\right). \tag{3.14}$$

In Figure 3.1, $1 - d_n$ is plotted versus n, along with this approximation. Higher order formulæ for the bias of the Sharpe ratio for non-Gaussian returns are given in Section 4.2.3, and suggest the approximation

$$d_{n+1} \approx 1 + \frac{3}{4n} + \frac{49}{32n^2}.$$

In Section 4.2.3, we will give the bias of the Sharpe ratio for general, non-Gaussian returns.

Example 3.3.2 (Bias in Sharpe ratio). Looking at one year's worth of returns with monthly marks, the bias is fairly large: $d_{12} = 1.0753$, *i.e.*, almost 8%. When looking at one year's worth of weekly marks, the bias is more modest: $d_{52} = 1.015$; for a year of daily marks $d_{252} = 1.003$. When $n > 100$, say, this bias is negligible. ⊣

3.3.2. Moments under up-sampling

Suppose, as a prospective investor in a fund, you are given one year's worth of daily log returns of the fund, say $n = 252$. You are offered, for a small fee, the option of viewing the log returns of every minute the market is open, a 390-fold increase in n. Should you accept this offer? Again, in this chapter we are assuming returns are *i.i.d.* normally distributed. How will this up-sampling affect the moments of the sample Sharpe ratio? Your intution should tell you that under the *i.i.d.* assumption, not much can be gleaned from higher resolution data, even though it appears to be a huge increase in the sample size. Indeed, because the daily returns and minutely log returns have the same sum the mean over minute marks is exactly the daily mean divided by

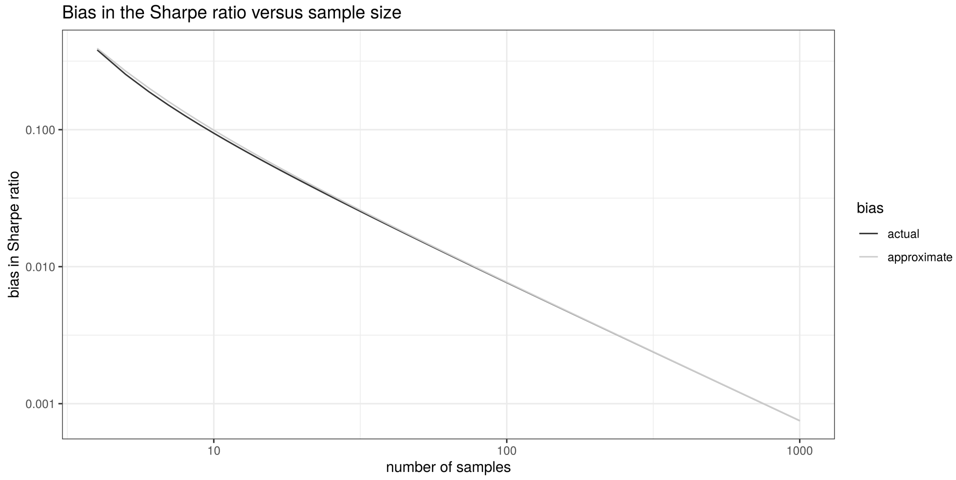

Figure 3.1.: The relative bias of the Sharpe ratio, $d_n - 1$ is plotted versus sample size. The approximation given by $d_{n+1} \approx 1 + \frac{3}{4n} + \frac{25}{32n^2}$ is also plotted.

390; only a change in the sample variance could cause the annualized Sharpe ratio to change.

To be concrete, let $\hat{\zeta}$ be the Sharpe ratio computed based on n marks, and let k be some positive integer. Let $\hat{\zeta}_k$ be the Sharpe ratio computed over the same time period, but with each log return divided into k pieces. Under the assumption of normality,

$$\sqrt{kn}\hat{\zeta}_k \sim t\left(\sqrt{kn}\zeta_k, kn - 1\right),$$

where $\zeta_k = \zeta/\sqrt{k}$ to make the units match. (That is, we are measuring returns over a higher frequency time scale, so the signal-noise ratio decreases in the usual way.) Similarly, because of how the units are defined, we should compare $\hat{\zeta}$ to $\sqrt{k}\hat{\zeta}_k$, since they have the same units. The moments of this estimate, with limits as $k \to \infty$, are

$$\begin{aligned}
\mathrm{E}\left[\left(\sqrt{k}\hat{\zeta}_k\right)\right] &= \sqrt{k} d_{kn}\zeta_k = d_{kn}\zeta \to \zeta,\\
\mathrm{E}\left[\left(\sqrt{k}\hat{\zeta}_k\right)^2\right] &= k\frac{kn-1}{kn-3}\left(\frac{1+kn\zeta_k^2}{kn}\right) \to \left(\frac{1+n\zeta^2}{n}\right),\\
\mathrm{E}\left[\left(\sqrt{k}\hat{\zeta}_k\right)^3\right] &= k^{3/2}\frac{kn-1}{kn-4} d_{kn}\frac{\zeta_k}{kn}\left(3+kn\zeta_k^2\right) \to \frac{\zeta}{n}\left(3+n\zeta^2\right),\\
\mathrm{E}\left[\left(\sqrt{k}\hat{\zeta}_k\right)^4\right] &= k^2\frac{(kn-1)^2}{(kn-3)(kn-5)}\left(\frac{3+6kn\zeta_k^2+k^2n^2\zeta_k^4}{k^2n^2}\right),\\
&\to \left(\frac{3+6n\zeta^2+n^2\zeta^4}{n^2}\right),
\end{aligned} \tag{3.15}$$

and so on. All terms in Equation 3.11 containing $\sqrt{n}\hat{\zeta}$ remain unchanged (they are unitless), while the ratios of terms in n converge to 1. The decrease in bias and variance in using $\sqrt{k}\hat{\zeta}_k$ instead of $\hat{\zeta}$ will be small, indeed, even when n is modest, say bigger than only 100. As $k \to \infty$, in fact, we have $\sqrt{k}\hat{\zeta}_k \xrightarrow{d} \mathcal{N}\left(\zeta, \frac{1}{n}\right)$. Thus the variance in the Sharpe ratio statistic is limited by the *length of time* over which we measure returns, and only weakly dependent on sample size (when sample size is reasonably large). It is not clear that changes in the higher order moments will have any meaningful impact on inference.

Example 3.3.3 (Up-sampling). Suppose you observe 252 days of Gaussian log returns with $\zeta = 0.05\,\mathrm{day}^{-1/2} = 0.7937\,\mathrm{yr}^{-1/2}$. You also observe the up-sampled data with $k = 390$. Then the percent change in the mean is

$$100\% \frac{\mathrm{E}\left[\sqrt{k}\hat{\zeta}_k\right] - \mathrm{E}\left[\hat{\zeta}\right]}{\mathrm{E}\left[\hat{\zeta}\right]} = -0.2984\%,$$

which is very small. The similarly defined percent change in the variance is -0.92%; in the skewness -99.7%; in the excess kurtosis -99.7%. ⊣

Example 3.3.4 (Down-sampling Market returns). The Sharpe ratio of the Market portfolio, introduced in Example 1.2.1, was computed on relative returns from Jan 1927 to Dec 2018. Returns were resampled at frequencies up to yearly, then the Sharpe ratios were computed, and tabulated in Table 3.1. The signal-noise ratios are all around $0.6\,\mathrm{yr}^{-1/2}$. We do not know which estimate is more accurate, but we also do not see radically different Sharpe ratios from different resampling frequencies.

daily	weekly	monthly	quarterly	yearly
0.63	0.62	0.60	0.53	0.58

Table 3.1.: The Sharpe ratio of the Market factor is shown, in units of $\mathrm{yr}^{-1/2}$, computed at different sampling frequencies, based on relative returns from Jan 1927 to Dec 2018.

⊣

3.3.3. Unbiased estimation and efficiency

While $\hat{\zeta}$ is a biased estimator for ζ, it is *asymptotically* unbiased. Moreover, we can easily construct an unbiased estimator

$$\tilde{\zeta} =_{\mathrm{df}} \frac{\hat{\zeta}}{d_n}. \tag{3.16}$$

This estimator is only of academic interest, as any fund manager reporting this statistic would be putting themselves at a relative disadvantage, since it is smaller than the biased estimator[2].

[2] Assuming their Sharpe ratio is positive!

It is well known that $\left[\hat{\mu}, \hat{\sigma}^2\right]^\top$ is a sufficient statistic for the parameters $\left[\mu, \sigma^2\right]^\top$ of a normal distribution. [11] Because $\left[\tilde{\zeta}, \hat{\sigma}\right]^\top$ is a one-to-one transformation of this sufficient statistic, it is sufficient as well.

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

$$\mathcal{I}\left(\zeta, \sigma\right) = n \left(\begin{array}{cc} 1 & \frac{\zeta}{2\sigma^2} \\ \frac{\zeta}{2\sigma^2} & \frac{2+\zeta^2}{4\sigma^4} \end{array} \right) \tag{3.17}$$

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

$$\mathcal{I}^{-1}\left(\zeta, \sigma\right) = \frac{1}{n} \left(\begin{array}{cc} 1 + \zeta^2/2 & -\zeta\sigma^2 \\ -\zeta\sigma^2 & 2\sigma^4 \end{array} \right) \tag{3.18}$$

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

$$\mathrm{Var}\left(\left[\tilde{\zeta}, \hat{\sigma}^2\right]^\top\right) = \left(\begin{array}{cc} \frac{(1+n\zeta^2)(n-1)}{{d_n}^2 n(n-3)} - \zeta^2 & \zeta\sigma^2\left(\frac{1}{d_n} - 1\right) \\ \zeta\sigma^2\left(\frac{1}{d_n} - 1\right) & \frac{2\sigma^4}{n-1} \end{array} \right). \tag{3.19}$$

The variance of $\tilde{\zeta}$ follows from Equation 3.13. 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.

3.4. † The lambda prime distribution

Now we turn our attention to Lecoutre's *lambda prime* distribution, which is, in some sense, 'dual' to the t-distribution. It is defined as follows: let $Z \sim \mathcal{N}\left(0, 1\right)$ independently of $\chi^2 \sim \chi^2\left(\nu\right)$. Then $\lambda = Z + t\sqrt{\chi^2/\nu}$ follows a lambda prime distribution with parameter t and degrees of freedom ν. [87, 88, 137] Let us write this as $\lambda \sim \lambda'\left(t, \nu\right)$, and we write the density, cumulative distribution, and quantile functions of the lambda prime distribution as, respectively, $f_{\lambda'}\left(x; t, \nu\right), F_{\lambda'}\left(x; t, \nu\right)$, and ${\lambda'}_p\left(t, \nu\right)$.

The connection between the lambda prime and the non-central t distribution shows up in the construction of frequentist confidence intervals (see Section 3.5.1), hypothesis tests on Sharpe ratio for independent samples (Section 3.5.3), and the Bayesian analysis

(Section 3.7). It is also the 'confidence distribution' of the non-central t, and appears in Fisher's work on Fiducial inference.

The connection between the two distributions is apparent. Suppose that $t \sim t(\delta, \nu)$. This means that there are independent random variables $Z \sim \mathcal{N}(0, 1)$ and $\chi^2 \sim \chi^2(\nu)$ such that

$$t = \frac{\delta + Z}{\sqrt{\chi^2/\nu}}.$$

This can be rearranged as

$$\delta = t\sqrt{\chi^2/\nu} - Z.$$

Because the normal distribution is symmetric, $-Z$ has the same distribution as Z, so conditional on observing t, we have $\delta \sim \lambda'(t, \nu)$.

This also connects the cumulative distribution functions (equivalently, the quantile functions) of the two distributions. For example, we have

$$F_t(x; \delta, \nu) = 1 - F_{\lambda'}(\delta; x, \nu), \tag{3.20}$$

and thus

$$F_{SR}(x; \zeta, n) = F_t\left(\sqrt{n}x; \sqrt{n}\zeta, n-1\right) = 1 - F_{\lambda'}\left(\sqrt{n}\zeta; \sqrt{n}x, n-1\right). \tag{3.21}$$

See Exercise 3.10. We can also define confidence intervals on δ in terms of the quantile function of the lambda prime distribution. An α confidence bound for the non-centrality parameter, δ, conditional on observing t, is ${\lambda'}_{1-\alpha}(t, \nu)$. Thus the confidence intervals given in Equation 3.27 (see later) can be expressed as

$$\frac{1}{\sqrt{n}}\left[{\lambda'}_{\alpha/2}\left(\sqrt{n}\hat{\zeta}, n-1\right), {\lambda'}_{(2-\alpha)/2}\left(\sqrt{n}\hat{\zeta}, n-1\right)\right]. \tag{3.22}$$

The distribution and quantile functions of the lambda prime can be evaluated directly [137], but they can be more easily implemented via off-the-shelf implementations of the distribution and quantile of the t distribution. For our purposes, we will at times need the distribution and quantile of a more general distribution, described below.

3.5. Frequentist inference on the signal-noise ratio

3.5.1. Confidence intervals

The variance of $\hat{\zeta}$ is given in Equation 3.13. Using the asymptotic expansion $d_n = 1 + \frac{3}{4(n-1)} + \mathcal{O}\left(n^{-2}\right)$, the variance of $\hat{\zeta}$ is approximated by

$$\mathrm{Var}\left(\hat{\zeta}\right) = \frac{(1+n\zeta^2)(n-1)}{n(n-3)} - \left(d_n\zeta\right)^2 \approx \frac{n-1}{n(n-3)} + \frac{\zeta^2}{2(n-3)} + \zeta^2\mathcal{O}\left(n^{-2}\right). \tag{3.23}$$

This is often quoted in the following more convenient form, which is asymptotically equivalent,

$$se\left(\hat{\zeta}\right) =_{\mathrm{df}} \sqrt{\mathrm{Var}\left(\hat{\zeta}\right)} \approx \sqrt{\frac{1 + \frac{\zeta^2}{2}}{n}}. \tag{3.24}$$

This is the most widely used form of the standard error of the Sharpe ratio. This standard error was described by Jobson and Korkie, and later by Lo [99, 76]. The equivalent result concerning the non-central t-distribution (which is the Sharpe ratio up to scaling by $\sqrt{n}$) was published in 1940 by Johnson and Welch. [77] We will refer to Equation 3.24 informally as the "vanilla" standard error, or the "Johnson and Welch" formula.

Since the signal-noise ratio, $\hat{\zeta}$, is unknown, it is typically approximated with the Sharpe ratio (the so-called 'plug-in' method), giving the following approximate $1-\alpha$ confidence interval on the signal-noise ratio:

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

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

$$\hat{\zeta} \pm z_{\alpha/2}\sqrt{\frac{1+\frac{\hat{\zeta}^2}{2}}{n-1}} \tag{3.25}$$

has better small sample coverage for normal returns.

We can take this one step further for the small sample case by adjusting for the bias in the Sharpe ratio. Using the Taylor approximation ${d_n}^{-1} \approx 1 - \frac{3}{4(n-1)}$ gives the approximate $1-\alpha$ confidence interval

$$\hat{\zeta}\left(1 - \frac{3}{4\left(n-1\right)}\right) \pm z_{\alpha/2}\sqrt{\frac{1+\frac{\hat{\zeta}^2}{2}}{n-1}} \tag{3.26}$$

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 $1-\alpha$ confidence interval on ζ has endpoints $[\zeta_l, \zeta_u]$ defined implicitly by

$$1-\alpha/2 = F_t\left(\sqrt{n}\hat{\zeta}; \sqrt{n}\zeta_l, n-1\right), \quad \alpha/2 = F_t\left(\sqrt{n}\hat{\zeta}; \sqrt{n}\zeta_u, n-1\right), \tag{3.27}$$

where $F_t\left(x; \delta, n-1\right)$ is the CDF of the non-central t-distribution with non-centrality parameter δ and $n-1$ degrees of freedom. Computationally, this method requires one to invert the CDF (*e.g.*, by Brent's method [21]), which is slower than approximations based on asymptotic normality. The endpoints of this confidence interval, ζ_l, ζ_u are actually quantiles of the lambda prime distribution. (*cf.* Section 3.4)

Mertens gives the form of standard error

$$se\left(\hat{\zeta}\right) \approx \sqrt{\frac{1+\frac{2+\gamma_2}{4}\zeta^2 - \gamma_1\zeta}{n}}, \tag{3.28}$$

where γ_1 is the skew, and γ_2 is the *excess* kurtosis of the returns distribution. [111, 119, 10] These are both zero for normally distributed returns, and so Mertens' form reduces to the Johnson & Welch form of the standard error. 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. We will consider Mertens' form further in the sequel, see Chapter 4.

Example 3.5.1 (Confidence intervals, Market returns). Consider the monthly relative returns of the Market, introduced in Example 1.2.1, which span from Jan 1927 to Dec 2018. The Sharpe ratio was measured to be $0.5982\text{yr}^{-1/2}$. 95% confidence intervals, based on the 'exact' method were computed as $[0.3922, 0.8039]\,\text{yr}^{-1/2}$. ⊣

Example 3.5.2 (Confidence intervals, UMD returns, attribution model). Consider the monthly relative returns of UMD, under an attribution against intercept, Market, SMB and HML, using the data introduced in Example 1.2.1, 1104 months of data from Jan 1927 to Dec 2018. The ex-factor Sharpe ratio was measured to be $0.2521\text{mo.}^{-1/2}$. 95% confidence intervals, based on the 'exact' method were computed as $[0.1911, 0.3129]\,\text{mo.}^{-1/2}$. ⊣

3.5.2. † Symmetric confidence intervals

The implicit confidence intervals on ζ given above are 'symmetric' in the sense that they are typically computed with equal type I error rates on both sides. That is, a $1-\alpha$ confidence interval $[\zeta_l, \zeta_u]$ is usually computed such that

$$\Pr\left\{\zeta < \zeta_l\right\} = \alpha/2 = \Pr\left\{\zeta > \zeta_u\right\}.$$

However these are typically not symmetric around the observed Sharpe ratio, $\hat{\zeta}$, rather they are usually slightly imbalanced. However, symmetric confidence intervals can easily be constructed numerically by finding δ such that

$$F_t\left(\sqrt{n}\hat{\zeta}; \sqrt{n}\left(\hat{\zeta}-\delta\right), n-1\right) - F_t\left(\sqrt{n}\hat{\zeta}; \sqrt{n}\left(\hat{\zeta}+\delta\right), n-1\right) = 1-\alpha.$$

The advantage of a symmetric interval is that we can express it as

$$\Pr\left\{\left|\zeta-\hat{\zeta}\right| \le \delta\right\} = 1-\alpha,$$

a fact which we will abuse later.

Example 3.5.3 (Confidence intervals, Symmetric). Consider the case of 504 daily observations of some hypothetical asset's returns, which result in a measured Sharpe ratio of exactly $0.6\text{yr}^{-1/2}$, assuming 252 days per year. The 'exact' 95% confidence intervals are computed as $[-0.7867, 1.9861]\,\text{yr}^{-1/2}$, which we can write as $\left[\hat{\zeta}-1.3867, \hat{\zeta}+1.3861\right]\text{yr}^{-1/2}$. These are *almost* symmetric.

Numerically we compute the symmetric intervals as approximately

$$\hat{\zeta} \pm 1.3864\text{yr}^{-1/2}.$$

These are very close to the exact intervals. They are also close to, but smaller than, the confidence intervals computed by the standard error approximation, Equation 3.25, which we compute as

$$\hat{\zeta} \pm 1.3878\text{yr}^{-1/2}.$$

⊣

The symmetry condition is useful in the following situation. Suppose you observe the Sharpe ratio of an asset; if it is positive, you will hold the asset long, otherwise you will hold the asset short. This simple "opportunistic" strategy seems perfectly natural, though perhaps somewhat naïve. It is not obvious how you would compute confidence intervals on your achieved returns, since they depend not just on the population parameter, ζ, but also on the statistic $\hat{\zeta}$, which you will also use to compute your confidence intervals.

We can easily compute such intervals. Starting from the symmetric confidence interval condition, $\Pr\left\{\left|\zeta - \hat{\zeta}\right| \le \delta\right\} = 1 - \alpha$, multiply the inside of the absolute value by a ± 1 in the form of $\operatorname{sign}\left(\hat{\zeta}\right)$ to arrive at

$$\Pr\left\{\left|\operatorname{sign}\left(\hat{\zeta}\right)\zeta - \left|\hat{\zeta}\right|\right| \le \delta\right\} = 1 - \alpha.$$

And thus

$$\left|\hat{\zeta}\right| \pm \delta \tag{3.29}$$

form a $1 - \alpha$ confidence intervals on the returns of this simple opportunistic strategy.

While the confidence intervals are symmetric about $\left|\hat{\zeta}\right|$, the rates of type I errors are typically not balanced on the two sides. Moreover, the imbalance depends on the unknown signal-noise ratio: when $|\zeta|$ is near 0, the type I errors will mostly be found below the lower bound; conversely when $|\zeta|$ is 'large', the type I errors will be more evenly balanced.

To illustrate this effect, we consider the rate of type I errors separately for "lower" and "upper" violations of the symmetric confidence bound. These can be defined respectively as

$$\Pr\left\{\operatorname{sign}\left(\hat{\zeta}\right)\zeta \le \left|\hat{\zeta}\right| - \delta\right\}, \quad \text{and} \quad \Pr\left\{\operatorname{sign}\left(\hat{\zeta}\right)\zeta \ge \left|\hat{\zeta}\right| + \delta\right\}.$$

We construct the symmetric confidence intervals using the standard error approximation of Equation 3.25, with the actual signal-noise ratio, *i.e.*, we are setting $\delta = \left|z_{\alpha/2}\right|\sqrt{\frac{1+\frac{\hat{\zeta}^2}{2}}{n-1}}$. We consider the case of 3 years of daily data, at 252 days per year. We compute the lower and upper type I rates, and plot them against ζ ranging from $0\text{yr}^{-1/2}$ to $2.5\text{yr}^{-1/2}$ in Figure 3.2. When $\zeta \le \left|z_{\alpha/2}\right|\sqrt{\frac{1+\frac{\hat{\zeta}^2}{2}}{n-1}}$ we see that there are no upper type I errors. (See Exercise 3.24.)

Note this imbalance in type I rates is a function of the unknown signal-noise ratio, and not of the observed Sharpe ratio. It is not possible to say much about the balance of type I errors conditional on the Sharpe ratio without leaning on a prior distribution for ζ, which would take us out of the Frequentist framework.

Note that one can turn the symmetric confidence interval $\left|\hat{\zeta}\right| \pm \delta$ into a one-sided 'symmetric' confidence interval

$$\left[\left|\hat{\zeta}\right| - \delta, \infty\right] \tag{3.30}$$

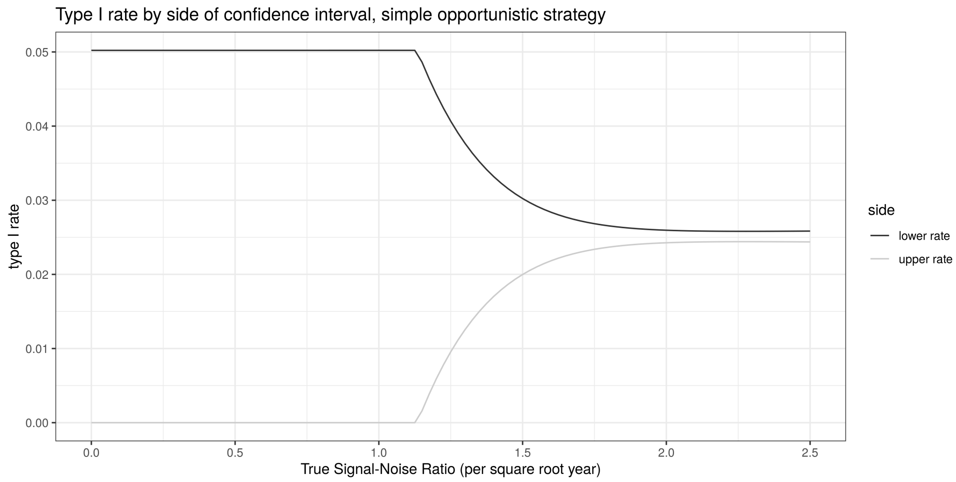

Figure 3.2.: Rates of upper and lower type I errors are plotted versus ζ for the case of 3 years of daily data, using the approximate standard error from Equation 3.25. The symmetric confidence interval has a total type I rate of approximately 0.05.

by extending one side. This confidence interval should have coverage between $1 - \alpha$ and $1 - \alpha/2$ depending on the unknown signal-noise ratio.

One can also analyze the opportunistic strategy via conditional inference, *cf.* Section 5.1.5.

3.5.3. Hypothesis tests

There are a few statistical tests for hypotheses involving the signal-noise ratio. The classical t-test for the mean can be considered a hypothesis test on the signal-noise ratio with a disastrous rate of return. In each of these, the sample Sharpe ratio is used.

1 sample mean test The classical one-sample test for mean involves a t-statistic which is like a Sharpe ratio with constant benchmark. Thus to test the null hypothesis:

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu > \mu_0,$$

we reject if the statistic

$$t_0 = \sqrt{n}\frac{\hat{\mu} - \mu_0}{\hat{\sigma}}$$

is greater than $t_{1-\alpha}\left(n - 1\right)$, the $1-\alpha$ quantile of the (central) t-distribution with $n-1$ degrees of freedom.

If $\mu = \mu_1 > \mu_0$, then the power of this test is

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

where $\delta_1 = \sqrt{n}\left(\mu_1 - \mu_0\right)/\sigma$ and $F_t\left(x; \delta, n-1\right)$ is the cumulative distribution function of the non-central t-distribution with non-centrality parameter δ and $n-1$ degrees of freedom. [147, 120]

1 sample signal-noise ratio test A one-sample test for signal-noise ratio can also be interpreted via the t-statistic. To test:

$$H_0 : \zeta = \zeta_0 \quad \text{versus} \quad H_1 : \zeta > \zeta_0,$$

we reject if the statistic $t = \sqrt{n}\hat{\zeta}$ is greater than $t_{1-\alpha}\left(\delta_0, n-1\right)$, the $1-\alpha$ quantile of the non-central t-distribution with $n-1$ degrees of freedom and non-centrality parameter $\delta_0 = \sqrt{n}\zeta_0$. Equivalently we reject if $\hat{\zeta} > SR_{1-\alpha}\left(\zeta_0, n\right)$.

If $\zeta = \zeta_1 > \zeta_0$, then the power of this test is

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

where $\delta_1 = \sqrt{n}\zeta_1$ and $F_t\left(x; \delta, n-1\right)$ is the cumulative distribution function of the non-central t-distribution with non-centrality parameter δ and $n-1$ degrees of freedom. [147, 120] Equivalently the power can be expressed as

$$1 - F_{SR}\left(SR_{1-\alpha}\left(\zeta_0, n\right); \zeta_1, n\right).$$

Hypothesis tests of the signal-noise ratio involving dependent returns will be considered later, see Section 4.3 and Equation 4.44.

Example 3.5.4 (Hypothesis testing, the Market). Consider the monthly relative returns of the Market, introduced in Example 1.2.1. To test the null hypothesis, $H_0 : \mu = 0.5\%\text{mo.}^{-1}$, we essentially compute the Sharpe ratio with benchmark rate of $0.5\%\text{mo.}^{-1}$, which is computed as $0.2736\text{yr}^{-1/2}$. Multiplying this by the square root of $n = 1104\text{mo.}$ gives the t-stat of 9.0895. We reject the null at the 0.05 level.

To test the null hypothesis, $H_0 : \zeta = 0.3\text{yr}^{-1/2}$, against the alternative $H_1 : \zeta > 0.3\text{yr}^{-1/2}$, we compute $\hat{\zeta} = 0.5982\text{yr}^{-1/2}$. The 0.975 quantile of the Sharpe ratio distribution under the null for $n = 1104\text{mo.}$ is $SR_{0.975}\left(0.3\text{yr}^{-1/2}, 1104\right) = 0.5055\text{yr}^{-1/2}$, and thus we reject the null hypothesis at the 0.025 level. This is essentially the same conclusion as could be drawn from the confidence intervals given in Example 3.5.1. $\dashv$

3.5.4. † Two One-sided and other Intersection Union Tests

It is sometimes lamented that hypothesis tests only allow you to cast doubt on the null hypothesis, rather than somehow positively assert some condition on the population parameters. The use of "Two One-Sided Tests" is sometimes prescribed[3] for this situation, as they potentially allow one to reject 'towards' some relationship, rather than away from one. [86]

Suppose the goal is to establish that the signal-noise ratio of an asset is equal to some fixed value, say ζ_0. Since strict equality is unlikely to be true, one instead seeks to establish equality within some range, or up to some uncertainty. So let us say that ζ is equivalent to ζ_0 if and only if $\zeta_l \le \zeta \le \zeta_h$, for some suitably defined ζ_l, ζ_h selected to give the proper precision to the notion of equivalence.

To try to show equivalence, one performs a test under the null hypothesis of *un-equivalence*, with the hope of rejecting in favor of equivalence. That is, one tests

$$H_0 : \zeta < \zeta_l \text{ or } \zeta > \zeta_h \quad \text{versus} \quad H_1 : \zeta_l \le \zeta \le \zeta_h.$$

To perform this test, one conducts two hypothesis tests, namely

$$\begin{aligned} H_{0a} &: \zeta < \zeta_l \quad \text{versus} \quad H_{1a} : \zeta \ge \zeta_l, \text{ and} \\ H_{0b} &: \zeta > \zeta_h \quad \text{versus} \quad H_{1b} : \zeta \le \zeta_h. \end{aligned}$$

If one rejects *both* H_{0a} and H_{0b}, then effectively one is rejecting H_0 'in favor of' the alternative hypothesis of equality, H_1. To conduct these tests, as outlined in Section 3.5.3, we reject if

$$t_{1-\alpha}\left(\sqrt{n}\zeta_l, n-1\right) \le \sqrt{n}\hat{\zeta} \le t_\alpha\left(\sqrt{n}\zeta_h, n-1\right).$$

Here, as above, $t_\alpha\left(\delta, n-1\right)$ is the α quantile of the non-central t-distribution with $n-1$ degrees of freedom and non-centrality parameter δ.

If we truly have equivalence, that is if $\zeta_l \le \zeta \le \zeta_h$, then the power of this test is

$$\left[1 - F_t\left(t_{1-\alpha}\left(\sqrt{n}\zeta_l, n-1\right); \sqrt{n}\zeta, n-1\right)\right] \bigwedge \left[F_t\left(t_\alpha\left(\sqrt{n}\zeta_h, n-1\right); \sqrt{n}\zeta, n-1\right)\right].$$

That is, the power of the test is the *minimum* of the powers of the two sub-hypothesis tests.

This combined test is an example of an *Intersection Union Test*, wherein one is testing a null hypothesis that can be expressed as the *union* of multiple testable hypotheses. The critical region, wherein one rejects based on the sample statistic (the Sharpe ratio in this case) is the *intersection* of critical regions of the separate tests[4] [148, 15]

In a similar manner, any of the equality tests of Section 3.5.3 and Section 3.5.5 can be reformulated as Intersection Union tests, see Exercise 3.23.

[3] To be fair, Bayesian methods are probably more popular for this task.

[4] In contrast, in a Union Intersection test, one is testing the null which is the intersection of hypotheses, and the critical region is the union of critical regions. The order of words in the nomenclature here seems somewhat arbitrary.

Example 3.5.5 (TOST for signal-noise ratio equivalent to zero). Consider the case of testing equivalence of the signal-noise ratio to zero by means of a symmetric interval, with $\zeta_l = -0.2\mathrm{yr}^{-1/2}$, $\zeta_h = 0.2\mathrm{yr}^{-1/2}$ at the type I rate of 0.05. Suppose one observes $n = 1008$ days of data. Then, the critical region for TOST in this case is $[-0.1479, 0.1479]\,\mathrm{yr}^{-1/2}$. We reject the null hypothesis precisely when the Sharpe ratio falls in this interval. The power of this test, the probability of correctly rejecting the null when the signal-noise ratio falls within $[-0.2, 0.2]\,\mathrm{yr}^{-1/2}$ for a type I rate of 0.05, is shown in Figure 3.3. We see that this test has very high power in this case when the signal-noise ratio is between about $-0.1\mathrm{yr}^{-1/2}$ and $0.1\mathrm{yr}^{-1/2}$.

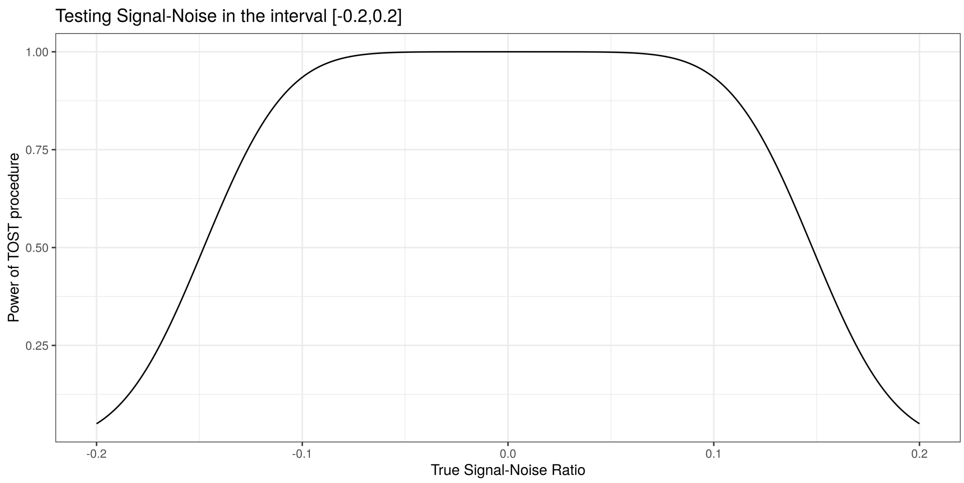

Figure 3.3.: Power of the two-sided test for $-0.2\mathrm{yr}^{-1/2} \leq \zeta \leq 0.2\mathrm{yr}^{-1/2}$ for the case of 1008 daily observations is plotted versus the true signal-noise ratio, for a type I rate of 0.05.

$\dashv$

3.5.5. Hypothesis 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. Again we should stress that the factors must be deterministic, as otherwise their variability would contribute to extra uncertainty in the test statistics. This requirement is relaxed in Section 4.4, where we consider random Gaussian returns and factors. 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 sample test for regression coefficients 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

$$t = \frac{\hat{\beta}^\top \boldsymbol{v} - c}{\hat{\sigma}\sqrt{\boldsymbol{v}^\top \left(\mathsf{F}^\top \mathsf{F}\right)^{-1} \boldsymbol{v}}}. \tag{3.31}$$

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.

1 sample test for ex-factor signal-noise ratio 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

$$t = \frac{\hat{\beta}^\top \boldsymbol{v}}{\hat{\sigma}\sqrt{\boldsymbol{v}^\top \left(\mathsf{F}^\top \mathsf{F}\right)^{-1} \boldsymbol{v}}}. \tag{3.32}$$

Under the null 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 $t_{1-\alpha}\left(\delta, n - l\right)$, the $1 - \alpha$ quantile of the non-central t-distribution with $n - l$ degrees of freedom and non-centrality parameter δ.

Equivalently, by definition,

$$\hat{\zeta}_g = \frac{\hat{\boldsymbol{\beta}}^\top \boldsymbol{v}}{\hat{\sigma}},$$

thus we should compare $\hat{\zeta}_g$ to the cutoff value of $r t_{1-\alpha}\left(c/r, n - l\right)$, where $r = \sqrt{\boldsymbol{v}^\top \left(\mathsf{F}^\top \mathsf{F}\right)^{-1} \boldsymbol{v}}$ is the rescaling parameter.

Example 3.5.6 (Hypothesis testing, UMD, attribution model). Consider the monthly relative returns of the Market, SMB, HML, and UMD portfolios, introduced in Example 1.2.1. Ignoring the error in the factors, we perform an attribution of the returns of UMD as some exposure to Market, SMB and HML. The ex-factor Sharpe ratio of the residual term under this attribution model is computed to be approximately $0.2521\text{mo.}^{-1/2}$ on $n = 1104\text{mo}$. This actually corresponds to a rather large t statistic.

To test the null hypothesis that the *residual* mean of UMD is $0.5\%\text{mo.}^{-1}$, we compute the t-statistic to have value of 4.2742, and reject the null at the 0.05 level.

To test the null hypothesis that the ex-factor signal-noise ratio of UMD is equal to $0.1\text{mo.}^{-1/2}$, we compute the 0.95 quantile of the non-central t distribution with non-centrality parameter $\delta = 3.2671$ and $n - l = 1100$ degrees of freedom. This has value 4.9221, equivalent to comparing the computed value of the ex-factor Sharpe ratio to $0.1507\text{mo.}^{-1/2}$. Since we compute the ex-factor Sharpe ratio to be around $0.2521\text{mo.}^{-1/2}$, we reject the null at the 0.05 level. ⊣

Example 3.5.7 (Hypothesis testing, Technology ex-factor signal-noise ratio). Consider the monthly relative returns of the Technology industry portfolio, introduced in Example 1.2.3. We align these with the corresponding monthly Market returns, Example 1.2.1. Joining the two series, we have 1104 months of data, ranging from Jan 1927 through Dec 2018. We compute the 'beta' of Technology to the market to be close to one, taking value 0.9476. We compute the ex-factor Sharpe ratio to be around $0.0327\text{mo.}^{-1/2}$.

To test the null hypothesis that the ex-factor signal-noise ratio of Technology is equal to $0.01\text{mo.}^{-1/2}$, we compute the 0.95 quantile of the non-central t distribution with non-centrality parameter $\delta = 0.3274$ and $n - l = 1102$ degrees of freedom. This has value 1.9742, leading to cutoff value $0.0603\text{mo.}^{-1/2} > 0.0327\text{mo.}^{-1/2}$, and we fail to reject the null at the 0.05 level. ⊣

Example 3.5.8 (Hypothesis testing, VIX reweighting Market returns). Consider the daily Market returns scaled by inverse previous day VIX weights, as introduced in Example 2.4.3. Joining the two series, we have 7808 days of data, ranging from 1990-01-03 through 2020-12-31. The inverse VIX weights have been rescaled to have unit mean in the sample period. We perform the attribution

$$\tilde{x_t} = x_t s_{t-1} = \beta_0 s_{t-1} + \beta_1 + \epsilon_t,$$

and test $H_0 : \beta^\top \boldsymbol{v} = \sigma c$ with $c = 0.3175\,\text{yr}^{-1/2}$ and $\boldsymbol{v}^\top = [1, 1]^\top$.

We compute $\hat{\zeta}_g = 0.7599\,\text{yr}^{-1/2}$. To test at the $\alpha = 0.05$ rate, we compare this to $rt_{1-\alpha}(c/r, n-l) = 0.6131\,\text{yr}^{-1/2}$, and narrowly reject the null hypothesis.

⊣

3.5.6. † Confidence intervals on the linear attribution model

The hypothesis tests given above can be converted into confidence intervals on the ex-factor signal-noise ratio. Such confidence intervals can be constructed by inverting the non-central t distribution, in analogue to the confidence intervals on the signal-noise ratio discussed in Section 3.5.1.

Again, we are assuming we observe n-vector of returns $\boldsymbol{x}$ and $n \times l$ matrix of associated factors F. The returns are Gaussian and independent, but are related to the factors which we assume here are deterministic. Let $\boldsymbol{v}$ be a given l vector. Then a $1-\alpha$ confidence interval on the quantity $\zeta_g = \beta^{\top}\boldsymbol{v}/\sigma$ has endpoints $[\zeta_{g,l}, \zeta_{g,u}]$ defined implicitly by

$$1-\alpha/2 = F_t\left(\hat{\zeta}_g/r; \zeta_{g,l}/r, n-l\right), \quad \alpha/2 = F_t\left(\hat{\zeta}_g/r; \zeta_{g,u}/r, n-l\right), \tag{3.33}$$

where $\hat{\zeta}_g = \hat{\boldsymbol{\beta}}^{\top}\boldsymbol{v}/\hat{\sigma}$ is the ex-factor Sharpe ratio, $r = \sqrt{\boldsymbol{v}^{\top}\left(\mathsf{F}^{\top}\mathsf{F}\right)^{-1}\boldsymbol{v}}$ is a rescaling parameter, and $F_t\left(x; \delta, n-l\right)$ is the CDF of the non-central t-distribution with non-centrality parameter δ and $n-l$ degrees of freedom. (*cf.* Equation 3.33.)

Additionally by using the connection to the t-distribution as we did in Section 3.5.1, we can approximate the standard error of the ex-factor Sharpe ratio as

$$se\left(\hat{\zeta}_g\right) =_{\mathrm{df}} \sqrt{\mathrm{Var}\left(\hat{\zeta}_g\right)} \approx \sqrt{\frac{1 + \frac{\zeta_g^2}{2r^2(n-l+1)}}{n-l+1}}, \tag{3.34}$$

where again $r = \sqrt{\boldsymbol{v}^{\top}\left(\mathsf{F}^{\top}\mathsf{F}\right)^{-1}\boldsymbol{v}}$ is the rescaling parameter.

Example 3.5.9 (Confidence intervals, UMD, attribution model). Consider the monthly relative returns of the Market, SMB, HML, and UMD portfolios, introduced in Example 1.2.1. We perform an attribution of the returns of UMD as some exposure to Market, SMB and HML, treating the latter returns as deterministic. The ex-factor Sharpe ratio of the residual term under this attribution model is computed to be approximately $0.2521\mathrm{mo.}^{-1/2}$ on $n = 1104\mathrm{mo}$. We compute 'exact' 95% confidence intervals on the ex-factor Sharpe ratio as $[0.1911, 0.3129]\,\mathrm{mo.}^{-1/2}$.

By plugging in the sample value, we estimate the standard error of the ex-factor Sharpe ratio via Equation 3.34 to be approximately $0.0306\mathrm{mo.}^{-1/2}$. Then we can compute an approximate 95% confidence intervals on the ex-factor Sharpe ratio as $[0.1921, 0.312]\,\mathrm{mo.}^{-1/2}$. *cf.* Example 4.4.1. $\dashv$

Example 3.5.10 (Confidence intervals, Technology ex-factor signal-noise ratio). Consider the attribution of monthly Technology industry returns to those of the Market, as described in Example 3.5.7. The ex-factor Sharpe ratio of the residual term under this attribution model is computed to be approximately $0.0327\mathrm{mo.}^{-1/2}$ on $n = 1104\mathrm{mo}$. We compute 'exact' 90% confidence intervals on the ex-factor Sharpe ratio as $[-0.0176, 0.0829]\,\mathrm{mo.}^{-1/2}$.

By plugging in the sample value, we estimate the standard error of the ex-factor Sharpe ratio via Equation 3.34 to be approximately $0.0301\mathrm{mo.}^{-1/2}$. Then we can compute an approximate 90% confidence intervals on the ex-factor Sharpe ratio as $[-0.0169, 0.0822]\,\mathrm{mo.}^{-1/2}$, which contains zero. $\dashv$

3.5.7. Type I errors, true incidence rate, and false discovery

Caution (On 'Significance'). Often, in the social sciences and elsewhere, a p-value smaller than 0.05 is deemed 'significant', a word which has come to have no meaning other than "exhibiting a p-value smaller than 0.05."

One should recognize, moreover, that the incidence rate of profitable trading strategies is certainly very low, likely lower than 0.05. In fact, some forms of the Efficient Markets Hypothesis would posit that the incidence rate is actually zero, and all trading strategies are type I errors. [107] One should exercise extreme caution, and use Bayes rule and some amount of guesstimation to avoid false discoveries.

Example 3.5.11 (Testing, incidence, and false discovery rate). Suppose that you sample randomly from trading strategies where the incidence rate of 'good' strategies is 0.001. If one employs a test for strategies with a 0.001 type I rate, and a power of 0.8, the false discovery rate will be as high as 0.5553. Which is to say around half of strategies which pass the significance test will not actually be 'good'. If, however, the incidence rate is as low as 10^{-6}, using the same test with the same type I and type II rates, the false discovery rate jumps to 0.9992; the vast majority of strategies passing the test are type I errors. ⊣

Unfortunately, the true incidence rate is unknown. Moreover, there is often no real gold standard for determining whether a trading strategy is in fact 'good': even stellar performance of a strategy in real trading for some fixed time after the analysis was performed may be the result of a type I error, and not an indication of true goodness.

It should also be noted that this simple model of randomly selecting trading strategies for significance testing is often an inaccurate description of strategy development. Typically strategies are developed by building off of well known ideas and theories, sometimes from published studies, and often involves sequential refinement of code and ideas, typically only accepting changes which improve some metric (*e.g.*, Sharpe ratio) of backtested returns. Dealing with the false discovery rate under this model of strategy development is addressed in Chapter 5. See also Exercise 3.28 to Exercise 3.29.

3.5.8. Power and sample size

Consider the test of the hypothesis $H_0 : \zeta = 0$, against the alternative $H_1 : \zeta > 0$. Note that this is equivalent to the traditional t-test for zero mean, *i.e.*, for testing the hypothesis $H_0 : \mu = 0$. A *power rule* ties together the (unknown) true effect size (ζ), sample size (n), and the type I and type II rates implicitly into a single equation. Typically, starting from three of these quantities, one infers the fourth, as illustrated by the following use cases:

1. Suppose you wanted to analyze a pairs trade on a pair of stocks which have only existed for two years. Is this enough data assuming the signal-noise ratio 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 signal-noise ratio 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 signal-noise ratio of the process is 1.5 $\mathrm{yr}^{-1/2}$, what is the probability of a type II error?

The power equation can be derived simply. The hypothesis test for $H_0 : \zeta = \zeta_0$ against the alternative $H_1 : \zeta > \zeta_0$ rejects the null precisely when

$$\hat{\zeta} \geq SR_{1-\alpha}\left(\zeta_0, n\right),$$

where the quantity on the right hand side is the $1 - \alpha$ quantile of the Sharpe ratio distribution with signal-noise ratio ζ_0 on sample size n. Now suppose that we wish the test to have power $1 - \beta$ when the true signal-noise ratio is ζ_e. This requires the cutoff value to be the β quantile when the signal-noise ratio is ζ_e. Thus the power equation is

$$SR_{1-\alpha}\left(\zeta_0, n\right) = SR_{\beta}\left(\zeta_e, n\right). \tag{3.35}$$

Because the quantile function is the inverse CDF, this can be expressed in two other, equivalent, ways, *viz.*

$$1 - \alpha = F_{SR}\left(SR_{\beta}\left(\zeta_e, n\right); \zeta_0, n\right), \quad \text{or,} \quad F_{SR}\left(SR_{1-\alpha}\left(\zeta_0, n\right); \zeta_e, n\right) = \beta. \tag{3.36}$$

From one of these two equations, one can infer either α or β as needed. If the goal is to infer the requisite ζ_e or n, numerical search using Equation 3.35 is indicated.

Example 3.5.12 (Basic power computations). Suppose you wish to test $H_0 : \zeta = 0\text{yr}^{-1/2}$ against $H_1 : \zeta > 0\text{yr}^{-1/2}$, with a 0.05 type I rate. Given 4 years of daily observations, the significance test rejects when $\hat{\zeta}$ exceeds $0.8232\text{yr}^{-1/2}$. Supposing $\zeta = 1.5\text{yr}^{-1/2}$, the power of the test is 0.912. ⊣

For sufficiently large sample size (say $n \geq 30$), the power law for the t-test of $H_0 : \zeta = 0$ is well approximated by

$$n \approx \frac{c}{\zeta^2}, \tag{3.37}$$

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. [77] 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 3.2. The case of $\alpha = 0.05, 1 - \beta = 0.80$ is known as "Lehr's rule". [162, 92]

	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 3.2.: 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.

Consider now the scaling in the rule $n \approx c\zeta^{-2}$. If the signal-noise ratio 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:

Power rule

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 $\left[\Phi^{-1}\left(0.95\right)\right]^2$. The relative error in this approximation for determining the sample size is shown in Figure 3.4, as a function of ζ; the error is smaller than one percent in the tested range.

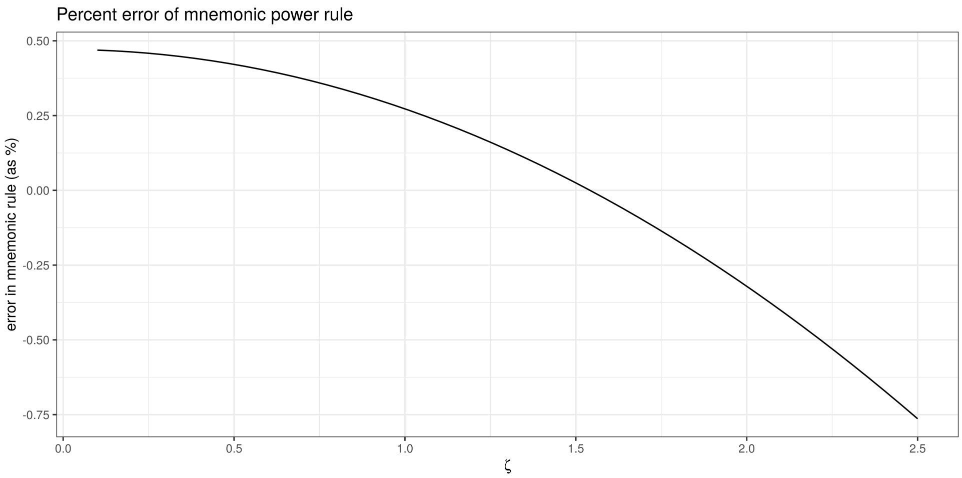

Figure 3.4.: The percent error of the power mnemonic $e \approx n\zeta^2$ is plotted versus ζ.

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 $1.6492\text{yr}^{-1/2}$ (a value which should be considered "very good"), your chances of 'passing' the t-test are less than fifty percent.

3.5.9. † Frequentist prediction intervals

Suppose, based on a sample of size n_1, you observed $\hat{\zeta}_1$ for some asset stream. What can you expect of the Sharpe ratio for n future observations? Though this is a question similar to that answered by the power rules given in Section 3.5.8, the power rules are deficient because 1. they rely on the unknown population parameter, ζ, when only a noisy estimate is available, and 2. they make statements within the dichotomy of type I and type II errors. To correct this, one may use a Frequentist *prediction interval*, which is an interval which, conditional on $\hat{\zeta}_1$ and n_1, contains the Sharpe ratio of those future observations with some specified probability, under replication.[5] The "under replication" clause here means that if you repeated the full experiment of generating $\hat{\zeta}_1$, n_1, constructing the prediction interval, then observing $\hat{\zeta}_2$, you should find that $\hat{\zeta}_2$ is within the interval with the given probability.

Suppose you observe $\hat{\zeta}_1$ on n_1 observations of normally distributed *i.i.d.* returns, then observe $\hat{\zeta}_2$ on n_2 observations from the same returns stream. We can write

$$\hat{\zeta}_1\sqrt{\chi_1^2/\left(n_1-1\right)} + Z_1/\sqrt{n_1} = \zeta = \hat{\zeta}_2\sqrt{\chi_2^2/\left(n_2-1\right)} + Z_2/\sqrt{n_2}, \tag{3.38}$$

where the $Z_i \sim \mathcal{N}\left(0,1\right)$, and the $\chi_i^2 \sim \chi^2\left(n_i-1\right)$ are independent. If there were an exact two-sample test for equal signal-noise ratios on independent samples, one would use that for this problem. Instead an approximation must be used.

Let $\hat{\zeta}_i = \zeta + s_i\epsilon_i$, for $i = 1, 2$ where s_i is the standard error of the Sharpe ratio, $\hat{\zeta}_i$ based on n_i observations, and ϵ_i is a zero mean, unit variance random variable. Usually we have $s_i = s/\sqrt{n_i}$. Then if the ϵ_i are independent, we have

$$\hat{\zeta}_2 = \hat{\zeta}_1 + \sqrt{1+\frac{n_1}{n_2}}\frac{s}{\sqrt{n_1}}\epsilon. \tag{3.39}$$

Thus the prediction interval around $\hat{\zeta}_1$ is inflated by a factor of

$$c = \sqrt{1+\frac{n_1}{n_2}} \tag{3.40}$$

compared to the equivalent confidence interval.

Example 3.5.13 (Prediction intervals, the Market). Consider the monthly relative returns of the Market, introduced in Example 1.2.1. The Sharpe ratio on $n_1 = 1128\text{mo.}$ was computed to be $\hat{\zeta}_1 = 0.1772\text{mo.}^{-1/2}$. For $n_2 = 12\text{mo.}$, the 95% prediction interval on $\hat{\zeta}_2$ was computed, via the normal approximation and standard error of Equation 3.24 to be approximately $\left[-0.4, 0.75\right]\text{mo.}^{-1/2}$.

For comparison, if one simply assumes that $\zeta = \hat{\zeta}_1 = 0.1772\text{mo.}^{-1/2}$, then the 95% prediction interval for the Sharpe ratio on $n_2 = 12\text{mo.}$ is $\left[-0.42, 0.86\right]\text{mo.}^{-1/2}$. ⊣

Example 3.5.14 (Prediction interval coverage, the Market). Consider the daily relative returns of the Market, introduced in Example 1.2.1. For each odd-numbered year from 1927 through 2019, the Sharpe ratio is computed using daily returns in the odd numbered year. Then 0.9 prediction intervals for Sharpe ratio are constructed based on the number of days in the following (even-numbered) year. The realized Sharpe ratio is then computed on the following even-numbered year, and compared to the prediction interval. The empirical coverage is approximately 0.81, much smaller than

[5] Typically 'prediction interval' is reserved for an interval around a single future observation, while 'tolerance interval' is used for multiple future observations. Our application is somewhat between these two.

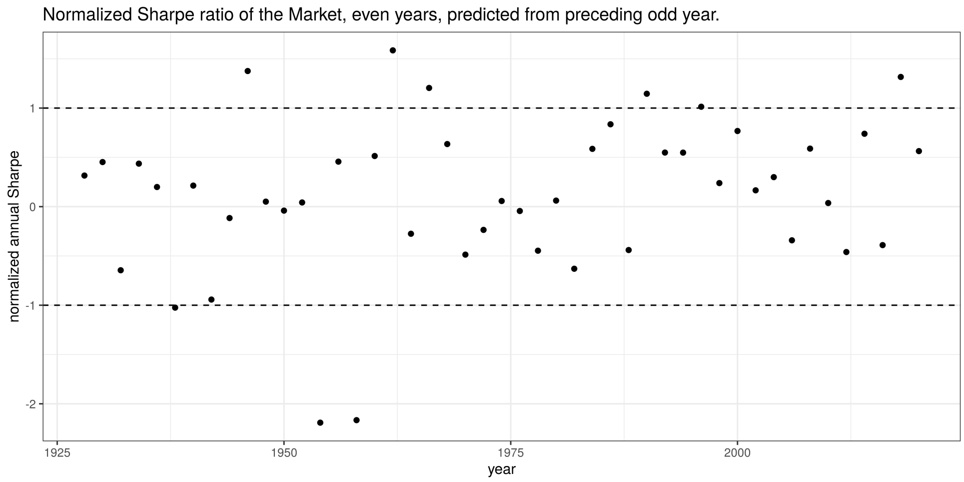

Figure 3.5.: The normalized Sharpe ratio of daily returns of the Market in even-numbered years are plotted, normalized to the 0.9 prediction interval computed using the previous year's Sharpe ratio. The normalization is such that a value of -1 corresponds to the lower limit of the prediction interval, 1 to the upper limit. The empirical coverage is approximately 0.81. See Example 3.5.14.

the nominal coverage. Thus the prediction intervals are too conservative, likely due to omitted variable bias or volatility clustering.

In Figure 3.5, the realized Sharpe ratio is plotted, for even-numbered years, relativized to the prediction intervals, so that a value of 1 corresponds to the realized Sharpe ratio exactly equaling the upper limit of the prediction interval, and a -1 corresponds to the lower limit. There are 9 values falling outside the prediction intervals.

To diagnose the conservative prediction intervals, the daily returns were permuted, and the experiment repeated: prediction intervals on Sharpe ratio computed using odd 'years', then compared to 'realized' Sharpe ratio in even 'years'. Repeating this process 100 times, the empirical coverage is approximately 0.89, equal to the nominal value. Thus it appears that non-normality and the standard error approximation are *not* to blame for the poor coverage, rather autocorrelation of returns or volatility in actual market returns. ⊣

ex-factor Sharpe ratio prediction intervals

We note that this approximate form only requires the computation and inflation of a standard error, and thus could be applied to computing prediction intervals on the ex-factor Sharpe ratio. However, the standard error of Equation 3.34 does not take into account randomness of the factors F_i, and may not deliver faithful prediction intervals. We attempt to correct for this in Section 4.5.2.

Example 3.5.15 (Prediction intervals, Technology ex-factor signal-noise ratio). Consider the attribution of monthly Technology industry returns to those of the Market, as described in Example 3.5.7. As in Example 3.5.14, for each odd-numbered year from 1927 through 2017, the ex-factor Sharpe ratio is computed using 12 monthly returns in the odd numbered year. Then 0.9 prediction intervals for ex-factor Sharpe ratio are constructed for the following (even-numbered) year. Prediction intervals are constructed by plugging in the observed ex-factor Sharpe ratio into Equation 3.34 to compute a standard error, then multiplying by $\sqrt{2}$.

The realized ex-factor Sharpe ratio is then computed on the following even-numbered year, and compared to the prediction interval. The empirical coverage is approximately 0.7, much smaller than the nominal coverage. Again the prediction intervals are too conservative, although in this case one suspects the poor coverage may be due to the rough approximation of Equation 3.34, the small sample size (12 monthly returns in-sample and out-of-sample), or due to autocorrelation, or correlated heteroskedasticity, *etc.*

To test this, as in Example 3.5.14 we permute the monthly paired returns of Technology and the Market and repeat the experiment, computing the ex-factor Sharpe ratio on the odd (shuffled) 'years' and comparing them to realized ex-factor Sharpe ratio on even 'years'. Repeating this process 100 times with different shufflings of the data, the empirical coverage is approximately 0.88, much closer to the nominal value. We blame autocorrelation of returns or volatility for the poor performance of the prediction intervals, and not non-normality of returns, small sample size, or the asymptotic approximation. ⊣

3.6. † Likelihoodist inference on the signal-noise ratio

We will now consider the likelihoodist's view of the Sharpe ratio. While philosophically distinct from the frequentist's view, the likelihoodist will arrive at similar conclusions to the frequentist.

We start with the likelihood of ζ given $\hat{\zeta}$. This is merely the density, given in Equation 3.7, but expressed as a function of ζ:

$$\mathcal{L}_{SR}\left(\zeta\,\middle|\,\hat{\zeta}, n\right) =_{\mathrm{df}} f_{SR}\left(\hat{\zeta}; \zeta, n\right).$$

The expression for the density of the Sharpe ratio given in Equation 3.7 suffices to point out there is unlikely to be a simple closed form for the maximum likelihood estimate, MLE. To find the MLE, take the derivative of the likelihood:

$$\frac{\partial \mathcal{L}_{SR}\left(\zeta\,\middle|\,\hat{\zeta}, n\right)}{\partial \zeta} = \frac{\sqrt{n/2}}{\sqrt{2\pi}\Gamma\left(\frac{n-1}{2}\right)} \int_0^\infty n\left(\hat{\zeta}\sqrt{x} - \zeta\right)\left(\frac{x}{2}\right)^{\frac{n-2}{2}} \exp\left(-\frac{x + n\left(\hat{\zeta}\sqrt{x} - \zeta\right)^2}{2}\right) \mathrm{d}x.$$

Then the MLE satisfies

$$\hat{\zeta}_{\mathrm{MLE}} = \frac{\hat{\zeta} \int_0^\infty \sqrt{x} \left(\frac{x}{2}\right)^{\frac{n-2}{2}} \exp\left(-\frac{x+n\left(\hat{\zeta}\sqrt{x}-\hat{\zeta}_{\mathrm{MLE}}\right)^2}{2}\right) \mathrm{d}x}{\int_0^\infty \left(\frac{x}{2}\right)^{\frac{n-2}{2}} \exp\left(-\frac{x+n\left(\hat{\zeta}\sqrt{x}-\hat{\zeta}_{\mathrm{MLE}}\right)^2}{2}\right) \mathrm{d}x}. \tag{3.41}$$

Thus $\hat{\zeta}_{\mathrm{MLE}}/\hat{\zeta}$ is the ratio of two integrals of positive functions, and is a positive number. Therefore the MLE estimate has the same *sign* as $\hat{\zeta}$. Unlike the unbiased estimator, which is smaller than $\hat{\zeta}$ when the latter is positive, it appears that the MLE is often *larger* than $\hat{\zeta}$ in this case; see Example 3.6.1.

For the purposes of estimating the likelihood, or the MLE, however, we are likely better off using off-the-shelf implementations of the density of the non-central t-distribution, and then finding the maximum using *e.g.*, a golden section search in the neighborhood of $\hat{\zeta}$.

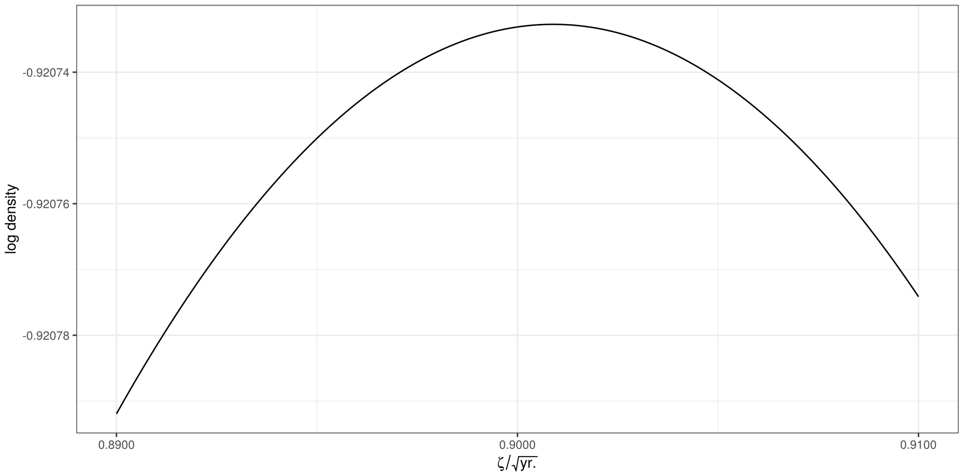

Figure 3.6.: The log density of $\hat{\zeta} = 0.9\,\mathrm{yr}^{-1/2}$ is plotted as a function of the signal-noise ratio, for $n = 253$ daily observations.

Example 3.6.1 (MLE). Suppose, based on 253 days of Gaussian log returns, you observe $\hat{\zeta} = 0.9\,\mathrm{yr}^{-1/2}$. The likelihood of $\hat{\zeta}$ as a function of ζ is plotted in Figure 3.6. The MLE is *not* equal to $\hat{\zeta}$, rather it takes a slightly larger value. We have $100\% \left(\frac{\hat{\zeta}_{\mathrm{MLE}}}{\hat{\zeta}} - 1\right) \approx 0.099\%$. ⊣

Example 3.6.2 (MLE, Market returns). Consider the monthly relative returns of the Market, introduced in Example 1.2.1. We compute $\hat{\zeta} = 0.6138\mathrm{yr}^{-1/2}$ for $n = 1128\mathrm{mo}$. The MLE for ζ is found to be approximately $\hat{\zeta} = 0.6139\mathrm{yr}^{-1/2}$. ⊣

3.6.1. Likelihood ratio test

To test $H_0 : \zeta = \zeta_0$ versus an unrestricted alternative, find the MLE, then compute the test statistic

$$D = -2 \log \frac{\mathcal{L}_{SR}\left(\zeta_0 \,\middle|\, \hat{\zeta}, n\right)}{\mathcal{L}_{SR}\left(\zeta_{\mathrm{MLE}} \,\middle|\, \hat{\zeta}, n\right)} = -2 \log \frac{f_{SR}\left(\hat{\zeta}; \zeta_0, n\right)}{f_{SR}\left(\hat{\zeta}; \zeta_{\mathrm{MLE}}, n\right)}. \tag{3.42}$$

The unrestricted model has one extra degree of freedom. For large n, under Wilk's theorem, we expect D to converge to a chi-square random variable with one degree of freedom. See Example 3.6.3.

Example 3.6.3 (LRT, simple example). Suppose, based on 1518 days of hypothetical Gaussian log returns, you observe $\hat{\zeta} = 1.5\,\mathrm{yr}^{-1/2}$. To test $H_0 : \zeta = 0.5\,\mathrm{yr}^{-1/2}$ against the unrestricted alternative, the log likelihood under H_0 is computed as -3.0136; the log likelihood for the MLE is -0.0254. The test statistic is then $D = 5.9764$. The probability that a chi-square with one degree of freedom takes a value this large is around 0.0145, so we may reject H_0 with high confidence. ⊣

To test $H_0 : \zeta = \zeta_0$ versus $H_1 : \zeta = \zeta_1$, compute the test statistic

$$\Lambda = \frac{\mathcal{L}_{SR}\left(\zeta_0 \,\middle|\, \hat{\zeta}, n\right)}{\mathcal{L}_{SR}\left(\zeta_1 \,\middle|\, \hat{\zeta}, n\right)}. \tag{3.43}$$

When Λ is small, smaller than some cutoff, we prefer H_1; when it is large, we prefer H_0. To use this test statistic in a null hypothesis test, one must find the cutoff value to reject H_0 in favor of H_1, with the cutoff value chosen to achieve the desired type I rate.

Note that the density of the Sharpe ratio is that of the non-central t distribution, up to scaling: $\mathcal{L}_{SR}\left(\zeta \,\middle|\, \hat{\zeta}, n\right) = \sqrt{n}\mathcal{L}_t\left(\sqrt{n}\zeta \,\middle|\, \sqrt{n}\hat{\zeta}, n-1\right)$. Thus

$$\Lambda = \frac{\mathcal{L}_t\left(\sqrt{n}\zeta_0 \,\middle|\, \sqrt{n}\hat{\zeta}, n-1\right)}{\mathcal{L}_t\left(\sqrt{n}\zeta_1 \,\middle|\, \sqrt{n}\hat{\zeta}, n-1\right)}.$$

Kruskal showed that this ratio is monotonic in $\sqrt{n}\hat{\zeta}$, and thus it is monotonic in $\hat{\zeta}$. [84] That is, supposing without loss of generality, that $\zeta_0 < \zeta_1$, the statistic Λ is decreasing in $\hat{\zeta}$. Thus to find a cutoff which achieves the desired type I rate under H_0, we take the $1-\alpha$ quantile under H_0, and plug it in as the Sharpe ratio. That is, the cutoff value is

$$\Lambda_c = \frac{f_{SR}\left(SR_{1-\alpha}\left(\zeta_0, n\right); \zeta_0, n\right)}{f_{SR}\left(SR_{1-\alpha}\left(\zeta_0, n\right); \zeta_1, n\right)}, \tag{3.44}$$

where $SR_q\left(\zeta_0, n\right)$ is the q quantile of the Sharpe ratio for the given signal-noise ratio and number of observations. This we can find by using off-the-shelf implementations of the quantile function of the non-central t distribution. See Example 3.6.4.

The upshot of all this is that the identity of H_1 is largely irrelevant (other than that it posits a larger signal-noise ratio than the null hypothesis, rather than a smaller one). There is, in fact, no reason to compute this likelihood ratio, and we reject H_0 in favor of *any* larger signal-noise ratio whenever $\hat{\zeta} > SR_{1-\alpha}(\zeta_0, n)$. This is merely the frequentist test for H_0 presented in Section 3.5.3. However, by the Neyman-Pearson lemma, the likelihood ratio test is the *uniformly most powerful* test of size α for H_0 against the set of alternatives $\zeta > \zeta_0$. [11] Since the frequentist test rejects exactly when the likelihood ratio test does, it, too, is the UMP for H_0.

Example 3.6.4 (LRT and Hypothesis testing). Suppose you wish to test $H_0 : \zeta = 0.25\,\mathrm{yr}^{-1/2}$ versus $H_1 : \zeta = 1.75\,\mathrm{yr}^{-1/2}$ for a Sharpe ratio observed on 1518 days of Gaussian log returns. To achieve a type I rate of $\alpha = 0.005$, the $1 - \alpha$ quantile under H_0 is computed as $SR_{1-\alpha}(\zeta_0, n) = 1.3036\,\mathrm{yr}^{-1/2}$. The cutoff value for the likelihood ratio is $\Lambda_c = 0.0656$. ⊣

Example 3.6.5 (LRT, Market returns). Consider the monthly relative returns of the Market, introduced in Example 1.2.1. Suppose you wish to test $H_0 : \zeta = 0.25\,\mathrm{yr}^{-1/2}$ versus $H_1 : \zeta = 0.75\,\mathrm{yr}^{-1/2}$ for 'the Market', assuming normality of returns. To achieve a type I rate of $\alpha = 0.01$, the $1-\alpha$ quantile under H_0 is computed as $SR_{1-\alpha}(\zeta_0, n) = 0.4913\,\mathrm{yr}^{-1/2}$. Since we observe $\hat{\zeta} = 0.6138\mathrm{yr}^{-1/2}$, we reject H_0 at the 0.01 level. As noted above, the identity of H_1 is irrelevant, other than that it posits a larger signal-noise ratio than H_0. ⊣

3.7. † Bayesian inference on the signal-noise ratio

Now we shall consider the Bayesian's view of the Sharpe ratio. In the traditional development of Bayesian inference on a Gaussian distribution with unknown parameters, prior and posterior distributions are considered on the mean and variance, μ and σ^2, or the mean and *precision*, the latter defined as σ^{-2}. [56, 114] It is a simple task to reformulate these in terms of the signal-noise ratio, ζ, and some transform of, say, the variance.

One commonly used conjugate prior is the 'Normal-Inverse-Gamma', under which one has an unconditional inverse gamma prior distribution on σ^2 (this is, up to scaling, one over a chisquare), and, conditional on σ, a normal prior on μ. [56, 3.3] These can be stated as

$$\begin{aligned} \sigma^2 &\propto \Gamma^{-1}\left(m_0/2, m_0\sigma_0^2/2\right), \\ \mu\left|\sigma^2\right. &\propto \mathcal{N}\left(\mu_0, \sigma^2/n_0\right), \end{aligned} \tag{3.45}$$

where σ_0^2, m_0, μ_0 and n_0 are the hyper-parameters. The density function for the inverse gamma law is given by

$$F_{\Gamma^{-1}}\left(x; a, b\right) = \frac{b^a}{\Gamma\left(a\right)} x^{-a-1} e^{-\frac{b}{x}}. \tag{3.46}$$

Under this formulation, an noninformative prior corresponds to $m_0 = 0 = n_0$.

After observing n *i.i.d.* draws from a normal distribution, $\mathcal{N}(\mu, \sigma)$, say $x_1, x_2, \ldots, x_n$, let $\hat{\mu}$ and $\hat{\sigma}$ be the sample estimates from Equation 2.2. The posterior is then

$$\begin{aligned} \sigma^2 &\propto \Gamma^{-1}\left(m_1/2, m_1\sigma_1^2/2\right), \\ \mu \left|\sigma^2\right. &\propto \mathcal{N}\left(\mu_1, \sigma^2/n_1\right), \end{aligned} \tag{3.47}$$

where

$$n_1 = n_0 + n, \qquad \mu_1 = \frac{n_0\mu_0 + n\hat{\mu}}{n_1}, \tag{3.48}$$

$$m_1 = m_0 + n, \qquad \sigma_1^2 = \frac{m_0\sigma_0^2 + (n-1)\,\hat{\sigma}^2 + \frac{n_0 n}{n_1}\left(\mu_0 - \hat{\mu}\right)^2}{m_1}. \tag{3.49}$$

This commonly used model can be trivially modified to one on the variance and the signal-noise ratio, where the former is a nuisance parameter. Transforming Equation 3.45, we arrive at

$$\begin{aligned} \sigma^2 &\propto \Gamma^{-1}\left(m_0/2, m_0\sigma_0^2/2\right), \\ \zeta \left|\sigma^2\right. &\propto \mathcal{N}\left(\frac{\mu_0}{\sigma}, 1/n_0\right), \end{aligned} \tag{3.50}$$

Marginalizing out σ^2, we arrive at a lambda prime prior (*cf.* Section 3.4)

$$\sqrt{n_0}\zeta \propto \lambda'\left(\sqrt{n_0}\zeta_0, m_0\right), \tag{3.51}$$

where $\zeta_0 = \mu_0/\sigma_0$. The marginal posterior can be written as

$$\sqrt{n_1}\zeta \propto \lambda'\left(\sqrt{n_1}\zeta_1, m_1\right), \tag{3.52}$$

where

$$n_1 = n_0 + n, \qquad \zeta_1 = \frac{n_0\zeta_0\sigma_0 + n\hat{\zeta}\hat{\sigma}}{n_1\sigma_1}, \tag{3.53}$$

$$m_1 = m_0 + n, \qquad \sigma_1^2 = \frac{m_0\sigma_0^2 + (n-1)\,\hat{\sigma}^2 + \frac{n_0 n}{n_1}\left(\zeta_0\sigma_0 - \hat{\zeta}\hat{\sigma}\right)^2}{m_1}, \tag{3.54}$$

where $\hat{\zeta} = \hat{\mu}/\hat{\sigma}$.

One is tempted to rearrange the equations to try to achieve an update formula that relies on $\hat{\zeta}$ alone. However, since $\hat{\zeta}$ follows, up to scaling, a non-central t distribution, which is not part of the exponential family, there is no hope of finding a conjugate prior for it. [56]

Example 3.7.1 (Basic Bayesian update). Suppose your prior is $\zeta_0 = 0.0154\text{day}^{-1/2}$, $n_0 = 10\text{day}$, $\sigma_0 = 0.013\text{day}^{-1/2}$, $m_0 = 100\text{day}$. One then observes $n = 252$ days of returns with $\hat{\zeta} = 0.0088\text{day}^{-1/2}$ and $\hat{\sigma} = 0.017\text{day}^{-1/2}$. The posterior is then $\zeta_1 = 0.0095\text{day}^{-1/2}$, $n_1 = 262\text{day}$, $\sigma_1 = 0.0159\text{day}^{-1/2}$, $m_1 = 352\text{day}$. 50000 samples were

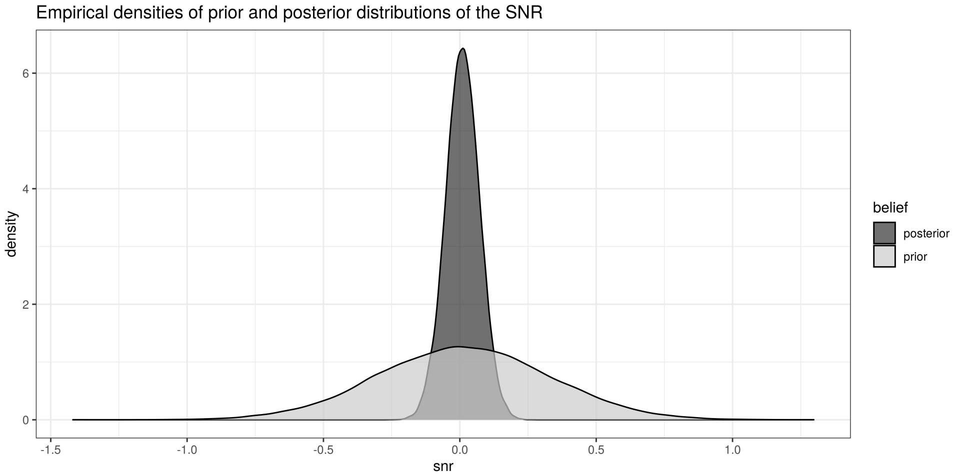

Figure 3.7.: 50,000 samples were drawn from both the prior and posterior distributions for Example 3.7.1. The empirical densities are then plotted.

drawn from the prior and posterior distributions; their empirical densities are plotted in Figure 3.7. The signal-noise ratio is plotted in units of day$^{-1/2}$. Thus while the evidence reduces the uncertainty in the prior, the posterior still contains considerable doubt regarding profitability. ⊣

Example 3.7.2 (Bayesian update, Market returns). Consider the monthly relative returns of the Market, introduced in Example 1.2.1. Suppose your prior (constructed prior to Jan 1927!) was $\zeta_0 = 0.125\text{mo.}^{-1/2}$, $n_0 = 12\text{mo.}$, $\sigma_0 = 4\%\text{mo.}^{-1/2}$, $m_0 = 72\text{mo.}$

One then observes $n = 1128$ months of returns with $\hat{\zeta} = 0.1772\text{mo.}^{-1/2}$ and $\hat{\sigma} = 5.3544\%\text{mo.}^{-1/2}$. The posterior is then $\zeta_1 = 0.1788\text{mo.}^{-1/2}$, $n_1 = 1140\text{mo.}$, $\sigma_1 = 5.2808\%\text{mo.}^{-1/2}$, $m_1 = 1200\text{mo.}$ ⊣

3.7.1. Bayesian inference on the ex-factor signal-noise ratio

The unattributed model of the previous section can be generalized to the attribution case by following the standard Bayesian regression analysis. Again, we are assuming that the factors $\boldsymbol{f}_t$ are deterministic. The Bayesian regression prior is typically stated as

$$\begin{aligned} \sigma^2 &\propto \Gamma^{-1}\left(m_0/2, m_0\sigma_0^2/2\right), \\ \boldsymbol{\beta}\left|\sigma^2\right. &\propto \mathcal{N}\left(\boldsymbol{\beta}_0, \sigma^2{\Lambda_0}^{-1}\right), \end{aligned} \tag{3.55}$$

where σ_0^2, m_0 are the Bayesian hyperparameters for the coefficient and degrees of freedom of the error term, while $\boldsymbol{\beta}_0$ is that for the regression coefficient, and Λ_0 parametrizes uncertainty in the regression coefficient.

As in Section 3.1.2, assume on has n observations of $\boldsymbol{f}_t$, stacked row-wise in the $n \times l$ matrix, F, and the corresponding returns stacked in vector $\boldsymbol{x}$. As in Equation 3.3, define $\hat{\boldsymbol{\beta}} =_{\mathrm{df}} \left(\mathsf{F}^\top\mathsf{F}\right)^{-1}\mathsf{F}^\top\boldsymbol{x}$ and $\hat{\sigma} =_{\mathrm{df}} \sqrt{\left(\boldsymbol{x} - \mathsf{F}\hat{\boldsymbol{\beta}}\right)^\top \left(\boldsymbol{x} - \mathsf{F}\hat{\boldsymbol{\beta}}\right)(n-l)^{-1}}$. The posterior distribution is

$$\begin{aligned} \sigma^2 &\propto \Gamma^{-1}\left(m_1/2, m_1\sigma_1^2/2\right), \\ \boldsymbol{\beta}\left|\sigma^2\right. &\propto \mathcal{N}\left(\boldsymbol{\beta}_1, \sigma^2{\Lambda_1}^{-1}\right), \end{aligned} \tag{3.56}$$

where

$$\Lambda_1 = \Lambda_0 + \mathsf{F}^\top\mathsf{F}, \quad \boldsymbol{\beta}_1 = {\Lambda_1}^{-1}\left(\Lambda_0\boldsymbol{\beta}_0 + \mathsf{F}^\top\mathsf{F}\hat{\boldsymbol{\beta}}\right), \tag{3.57}$$

$$m_1 = m_0 + n, \qquad \sigma_1^2 = \frac{m_0\sigma_0^2 + (n-l)\,\hat{\sigma}^2 + \hat{\boldsymbol{\beta}}^\top\mathsf{F}^\top\mathsf{F}\hat{\boldsymbol{\beta}} + {\boldsymbol{\beta}_0}^\top\Lambda_0\boldsymbol{\beta}_0 - {\boldsymbol{\beta}_1}^\top\Lambda_1\boldsymbol{\beta}_1}{m_1}. \tag{3.58}$$

A non-informative prior corresponds to $\Lambda_0 = \mathbf{0}, \boldsymbol{\beta}_0 = \mathbf{0}, \sigma_0^2 = 0, m_0 = 0$.

We can collapse the prior or posterior 'along' the direction $\boldsymbol{v}$ via

$$\begin{aligned} \sigma^2 &\propto \Gamma^{-1}\left(m_i/2, m_i\sigma_i^2/2\right), \\ \boldsymbol{v}^\top\boldsymbol{\beta}\left|\sigma^2\right. &\propto \mathcal{N}\left(\boldsymbol{v}^\top\boldsymbol{\beta}_i, \sigma^2\boldsymbol{v}^\top{\Lambda_i}^{-1}\boldsymbol{v}\right), \end{aligned} \tag{3.59}$$

where $i = 0$ for the prior and $i = 1$ for the posterior. As in the unattributed model, marginalizing out σ^2, we have a lambda prime prior and posterior:

$$\left(\boldsymbol{v}^\top{\Lambda_i}^{-1}\boldsymbol{v}\right)^{-1/2}\frac{\boldsymbol{\beta}^\top\boldsymbol{v}}{\sigma} = \left(\boldsymbol{v}^\top{\Lambda_i}^{-1}\boldsymbol{v}\right)^{-1/2}\zeta_g \propto \lambda'\left(\left(\boldsymbol{v}^\top{\Lambda_i}^{-1}\boldsymbol{v}\right)^{-1/2}\zeta_{g,i}, m_i\right), \tag{3.60}$$

where $\zeta_{g,i} =_{\mathrm{df}} {\boldsymbol{\beta}_i}^\top\boldsymbol{v}/\sigma_i$.

Example 3.7.3 (Bayesian update, UMD returns, attribution model). Consider the monthly relative returns of UMD, under an attribution against intercept, Market, SMB and HML, using the data introduced in Example 1.2.1. Let $\boldsymbol{\beta}$ (and thus the columns of F) be ordered by intercept (the idiosyncratic term), then SMB, then HML. To keep track of the units, let $U = \mathrm{Diag}\left([\mathrm{mo.}, \%, \%, \%]^\top\right)$ be the matrix of units.

Suppose your prior (constructed prior to Jan 1927!) was

$$\Lambda_0 = U\begin{bmatrix} 60.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 12.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 12.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 12.00 \end{bmatrix}U,$$

$$\boldsymbol{\beta}_0 = U^{-1}[0.5, 0, 0, 0]^\top\%, \quad m_0 = 12\mathrm{mo.}, \quad \sigma_0 = 3\%\mathrm{mo.}^{-1/2}.$$

One then observes $n = 1128$ months of returns with

$$\mathsf{F}^{\top}\mathsf{F} = U \begin{bmatrix} 1128.00 & 1070.17 & 234.60 & 363.32 \\ 1070.17 & 33325.49 & 6435.78 & 5419.69 \\ 234.60 & 6435.78 & 11492.63 & 1689.78 \\ 363.32 & 5419.69 & 1689.78 & 13984.86 \end{bmatrix} U,$$

$$\hat{\boldsymbol{\beta}} = U^{-1}[1.0152, -0.2173, -0.0499, -0.4691]^{\top}\%, \quad n = 1128\text{mo.}, \quad \hat{\sigma} = 4.1219\%\text{mo.}^{-1/2}.$$

The posterior is then

$$\Lambda_1 = U \begin{bmatrix} 1188.00 & 1070.17 & 234.60 & 363.32 \\ 1070.17 & 33337.49 & 6435.78 & 5419.69 \\ 234.60 & 6435.78 & 11504.63 & 1689.78 \\ 363.32 & 5419.69 & 1689.78 & 13996.86 \end{bmatrix} U,$$

$$\boldsymbol{\beta}_1 = U^{-1}[0.9882, -0.2165, -0.0499, -0.4684]^{\top}\%, \quad m_1 = 1140\text{mo.}, \quad \sigma_1 = 4.1064\%\text{mo.}^{-1/2}.$$

The ex-factor signal-noise ratio for idiosyncratic returns under the attribution model corresponds to $\boldsymbol{v} = U[1, 0, 0, 0]^{\top}$. Thus the posterior belief on ζ_g, marginalizing out σ, can be expressed as

$$33.9246\text{mo.}^{1/2}\%^{-1}\zeta_g \propto \lambda'\left(8.164, 1140\right).$$

⊣

3.7.2. Credible intervals on the signal-noise ratio

One can construct credible intervals on the signal-noise ratio based on the posterior, so-called *posterior intervals*, via quantiles of the lambda prime distribution. [56] For example, a $(1-\alpha)$ credible interval on ζ is given by

$$\frac{1}{\sqrt{n_1}}\left[{\lambda'}_{\alpha/2}\left(\sqrt{n_1}\zeta_1, m_1\right), {\lambda'}_{(2-\alpha)/2}\left(\sqrt{n_1}\zeta_1, m_1\right)\right]. \tag{3.61}$$

For the case of noninformative priors (corresponding to $n_0 = m_0 = 0$), this is equivalent to

$$\frac{1}{\sqrt{n}}\left[{\lambda'}_{\alpha/2}\left(\sqrt{n}\sqrt{\frac{n}{n-1}}\hat{\zeta}, n\right), {\lambda'}_{(2-\alpha)/2}\left(\sqrt{n}\sqrt{\frac{n}{n-1}}\hat{\zeta}, n\right)\right] \tag{3.62}$$

This is equivalent to the Frequentist confidence intervals given in Equation 3.22, but replacing n for $n-1$ in the degrees of freedom for σ. Asymptotically, the Bayesian credible interval for an noninformed prior is the same as the Frequentist confidence interval. Alternatively, a Bayesian quant could argue that her Frequentist cousin is a confused Bayesian with prior $n_0 = 0, m_0 = -1, \sigma_0^2 = 0$.

For the ex-factor signal-noise ratio, the $(1-\alpha)$ credible interval on ζ_g is

$$\frac{1}{\sqrt{\boldsymbol{v}^{\top}\Lambda_1\boldsymbol{v}}}\left[{\lambda'}_{\alpha/2}\left(\sqrt{\boldsymbol{v}^{\top}\Lambda_1\boldsymbol{v}}\zeta_{g,1}, m_1\right), {\lambda'}_{(2-\alpha)/2}\left(\sqrt{\boldsymbol{v}^{\top}\Lambda_1\boldsymbol{v}}\zeta_{g,1}, m_1\right)\right]. \tag{3.63}$$

Example 3.7.4 (Noninformative prior, Market returns). Consider the monthly relative returns of the Market, introduced in Example 1.2.1. Starting from a noninformative prior ($n_0 = m_0 = 0\text{mo.}$), one computes $n = 1128$, $\hat{\zeta} = 0.1772\text{mo.}^{-1/2}$ and $\hat{\sigma} = 5.3544\%\text{mo.}^{-1/2}$. The posterior is then $\zeta_1 = 0.1773\text{mo.}^{-1/2}$, $n_1 = 1128\text{mo.}$, $\sigma_1 = 5.352\%\text{mo.}^{-1/2}$, $m_1 = 1128\text{mo}$. From the posterior, a 95% credible interval on ζ is $[0.4102, 0.8177]\,\text{yr}^{-1/2}$. The Frequentist 95% confidence interval is $[0.4099, 0.8174]\,\text{yr}^{-1/2}$. The two agree to three decimal places. ⊣

Example 3.7.5 (Noninformative prior, UMD returns, attribution model). Consider the monthly relative returns of UMD, under an attribution against intercept, Market, SMB and HML, using the data introduced in Example 1.2.1. Letting $\boldsymbol{\beta}$ be ordered by intercept (the idiosyncratic term), then SMB, then HML, for a noninformative prior, the posterior belief on ζ_g, marginalizing out σ, can be expressed as

$$33.0283\text{mo.}^{1/2}\%^{-1}\zeta_g \propto \lambda'\left(8.1492, 1128\right).$$

From the posterior, a 95% credible interval on ζ_g is $[0.1865, 0.3069]\,\text{mo.}^{-1/2}$. Compare this to the frequentist confidence intervals, found in Example 3.5.2 to be $[0.186, 0.3065]\,\text{mo.}^{-1/2}$. ⊣

3.7.3. Posterior prediction intervals on the Sharpe ratio

A Bayesian prediction interval is an interval which contains some fixed proportion of our posterior *belief* about the Sharpe ratio of some future observations. This is very similar to the Frequentist prediction interval introduced in Section 3.5.9, but dances around the issues of frequency and belief that separate the Frequentist and Bayesian.

As in the Frequentist case, our belief is that ζ_2, based on n_2 future observations, will be drawn from a compound non-central Sharpe ratio distribution with non-centrality parameter drawn from the posterior distribution. Effectively this is a 't of lambda prime' distribution, as was the case in the Frequentist setting. We can summarize this as

$$\begin{aligned} \sqrt{n_1}\zeta &\propto \lambda'\left(\sqrt{n_1}\zeta_1, m_1\right),\\ \sqrt{n_2}\hat{\zeta}_2\,|\zeta &\propto t\left(\sqrt{n_2}\zeta, n_2 - 1\right), \end{aligned} \tag{3.64}$$

although this jumbles up the usual notation, since $\hat{\zeta}_2$ is a quantity one can eventually observe, not a population parameter. Nevertheless, the intent of these equations should be clear. As a consequence, one could find posterior prediction intervals, by using the the *upsilon distribution.* [126] However, code to compute quantiles or the cumulative distribution of the upsilon distribution is not widely available. Instead the prediction interval may be more easily found via direct Monte Carlo simulations: first draw a value of ζ from the posterior, then simulate a value of $\hat{\zeta}_2$ from a Sharpe ratio distribution with that signal-noise ratio. That is, simulate ζ, then $\hat{\zeta}_2$ using the generative forms from Equation 3.64.

Example 3.7.6 (Predictive intervals, Market returns). Consider the monthly relative returns of the Market, introduced in Example 1.2.1. Based on a prior of

$\zeta_0 = 0.125\text{mo.}^{-1/2}$, $n_0 = 12\text{mo.}$, $\sigma_0 = 4\%\text{mo.}^{-1/2}$, $m_0 = 72\text{mo.}$, one observes $n = 1128$ months of returns with $\hat{\zeta} = 0.1772\text{mo.}^{-1/2}$ and $\hat{\sigma} = 5.3544\%\text{mo.}^{-1/2}$. The posterior is then $\zeta_1 = 0.1788\text{mo.}^{-1/2}$, $n_1 = 1140\text{mo.}$, $\sigma_1 = 5.2808\%\text{mo.}^{-1/2}$, $m_1 = 1200\text{mo.}$

For $n_2 = 12\text{mo.}$, the 95% prediction interval on $\hat{\zeta}_2$ was computed, by 10^6 Monte Carlo simulations, to be approximately $\left[-0.43, 0.87\right]\text{mo.}^{-1/2}$. Compare this to the Frequentist prediction interval found in Example 3.5.13. ⊣

Exercises

Ex. 3.1 **Units** What are the units of the t-statistic?

Ex. 3.2 **t form for ex-factor Sharpe ratio** Derive the equivalents of Equation 3.5 for the ex-factor Sharpe ratio.

Ex. 3.3 **Form of the t statistic** Suppose that $Z \sim \mathcal{N}\left(\mu_0, \sigma^2/\nu\right)$ independently of $(\nu - 1)\, X/\sigma^2 \sim \chi^2\left(\nu - 1\right)$. Show that

$$\sqrt{\nu}\frac{Z - \mu_1}{\sqrt{X}} \sim t\left(\sqrt{\nu}\frac{\mu_0 - \mu_1}{\sigma}, \nu - 1\right).$$

Ex. 3.4 **Regarding c_4** The statistical quality control literature defines

$$c_4\left(n\right) =_{\mathrm{df}} \sqrt{\frac{2}{n-1}}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}.$$

Prove that $d_n = \frac{n-1}{n-2}c_4\left(n\right)$, for d_n defined as in Equation 3.12.

Ex. 3.5 **'Robust' tests** The t test tests the null hypothesis of zero *mean* return. In social sciences it is often argued that the *median* is a more representative measure of some effect. Why would it be a bad idea to rely on tests of median return?

Ex. 3.6 **Probability of a loss** Suppose you observe n observations of *i.i.d.* Gaussian log returns $x \sim \mathcal{N}\left(\mu, \sigma\right)$.

1. Show that the probability of a loss (*i.e.*, $\sum_i x_i \le 0$) is $\Phi\left(-\sqrt{n}\zeta\right)$, where $\Phi\left(x\right)$ is the probability distribution function of the normal distribution.
2. Show that, as a consequence, $f_{SR}\left(0; \zeta, n\right) = \Phi\left(-\sqrt{n}\zeta\right)$.
3. * Is the function $f_{SR}\left(x; \zeta, n\right) - \Phi\left(x - \sqrt{n}\zeta\right)$ monotonic in x?
4. Plot the probability of a loss over one year ($n = 253$ days) when ζ ranges from $0.5\mathrm{yr}^{-1/2}$ to $2.5\mathrm{yr}^{-1/2}$, in logscale. Estimate a power law for the probability of a loss.

Ex. 3.7 **Boring Frequentist Computations** Suppose you observe n daily observations of *i.i.d.* Gaussian log returns $x \sim \mathcal{N}\left(\mu, \sigma\right)$.

1. Suppose that $\zeta = 0.1\mathrm{day}^{-1/2}$. Compute the probability of observing $\hat{\zeta} \ge 0.02\mathrm{day}^{-1/2}$ given $n = 500$.
2. Suppose that $\zeta = 0.1\mathrm{day}^{-1/2}$. Compute the expected value of $\hat{\zeta}$ given $n = 500$.
3. Supposing that $\hat{\zeta} = 0.02\mathrm{day}^{-1/2}$ based on $n = 500$ days of observations. Compute 95% confidence intervals on ζ based on Equation 3.25. Compute them again based on Equation 3.27.

Ex. 3.8 More Boring Frequentist Computations

1. Compute the ex-factor Sharpe ratio of the Consumer industry returns using an attribution against the Fama French Market, SMB, HML, and UMD returns. You can easily join the data together in **R** as follows:
   Test the hypothesis that the ex-factor signal-noise ratio is zero against the alternative that it is larger than zero.
2. Repeat that test using the Manufacturing industry returns.

Ex. 3.9 Frequentist approximate test for equal signal-noise ratio Derive the normal approximation to the hypothesis test of equality of signal-noise ratios for two independent samples.

1. Apply your procedure to test the null hypothesis that the returns of the Market have the same signal-noise ratio before and since January 1970.

Ex. 3.10 Lambda prime Consider the lambda prime distribution defined in Section 3.4.

1. Show that the probability distribution of the non-central t distribution is the survival function of the lambda prime, and vice versa, *i.e.*, Equation 3.20.
2. Show that the lambda prime is the *confidence distribution* associated with the non-central t distribution. That is, show that if $t \sim t(\delta, \nu)$, then $F_{\lambda'}(\delta; t, \nu) \sim \mathcal{U}([0, 1])$. [152, 175]
3. Derive the probability density function of the lambda prime distribution. (*Hint:* it is a convolution of a Nakagami distribution density with a normal density.)
4. How does the PDF of the lambda prime relate to the PDF of the non-central t?
5. Suppose $t \sim t(\delta, \nu)$. Is it the case that the MLE of δ conditional on observed t is equal to the *mode* of the lambda prime distribution with parameter t and ν degrees of freedom?

Ex. 3.11 CDF of Sharpe ratio Write code to compute the cumulative distribution function of the Sharpe ratio. It should take observed Sharpe ratio, $\hat{\zeta}$, the signal-noise ratio, ζ, and sample size n, and return the CDF. A simple test of your code is to randomly generate $\hat{\zeta}$ values for a fixed value of ζ and n, apply your CDF function to those $\hat{\zeta}$ values, and Q-Q plot them against the uniform.

Ex. 3.12 CDF of the lambda prime Write code to compute the CDF of the lambda prime distribution. Test it via randomly generating variates.

Ex. 3.13 Moments of Sharpe ratio Given 1000 daily observations of *i.i.d.* normal log returns with $\zeta = 0.06\,\mathrm{day}^{-1/2}$, compute the first four non-central moments of $\hat{\zeta}$ computed in daily units. Also compute the variance, skewness, and excess kurtosis of $\hat{\zeta}$.

Ex. 3.14 Up-sampling Suppose you are given 260 weekly observations of *i.i.d.* normal log returns with $\zeta = 0.10\,\mathrm{wk.}^{-1/2}$.

1. Compute the first four non-central moments of $\hat{\zeta}$, computed in weekly units.
2. Compute the variance, skewness, and excess kurtosis of $\hat{\zeta}$, computed in weekly units where appropriate.
3. Suppose you upsample to daily units, thus $k = 5$. Compute the percent changes in mean, variance, skewness, and excess kurtosis of $\sqrt{k}\hat{\zeta}_k$ versus $\hat{\zeta}$.

Ex. 3.15 Up-sampling FF3 Consider the monthly and daily returns of the Fama-French factors, available in **aqfb.data**. [128] Compute the Sharpe ratio for the three portfolios, Market, SMB, and HML, for the monthly and the daily returns.

Ex. 3.16 HFT: not about sample size It is often clumsily argued that high frequency trading is to be preferred to daily frequency trading 'because the sample size is larger.' Critique this argument.

Ex. 3.17 Standard error of the Sharpe ratio Derive the approximation of Equation 3.23.

Ex. 3.18 Empirical verification of hypothesis tests under the null Via Monte Carlo simulation verify the frequentist tests of Section 3.5.3 by drawing samples under the null and computing p-values. Perform 10^5 simulations for each of the following:

1. Confirm the one sample test for signal-noise ratio. Draw $n = 1024\mathrm{day}$ daily returns from a normal random generator with $\zeta = 0.1\mathrm{day}^{-1/2}$. Confirm that $\sqrt{n}\hat{\zeta} \sim t\left(\sqrt{n}\zeta, n-1\right)$. One easy way to perform this task is to compute p-values and confirm they are uniformly distributed by Q-Q plotting them against a uniform law, or by plotting their empirical CDF.
2. Confirm the independent two sample test for equality of signal-noise ratio. For $i = 1, 2$, draw $n_i = 512i\,\mathrm{day}$ daily returns from two independent normal asset streams, both with $\zeta = 0.075\mathrm{day}^{-1/2}$. Let the population mean of the i^{th} population be $\mu_i = i \times 10^{-4}\mathrm{day}^{-1}$.
3. Confirm the independent k sample test for an equation on signal-noise ratio. For $i = 1, \ldots, 7$, draw $n_i = 512i\,\mathrm{day}$ daily returns from seven independent normal asset streams, with the signal-noise ratio of the i^{th} population $\zeta_i = (0.01 + 0.01i)\mathrm{day}^{-1/2}$. Let the population mean of the i^{th} population be $\mu_i = i \times 10^{-4}\mathrm{day}^{-1}$. The null hypothesis being tested is that $\sum_i \zeta_i = 0.35\mathrm{day}^{-1/2}$.

Ex. 3.19 Verification of hypothesis tests under the alternative Via Monte Carlo simulation, check that the frequentist tests of Section 3.5.3 can reject the null, by drawing samples under an alternative hypothesis, computing p-values, and confirming that the p-values are not uniform. (*cf.* Exercise 3.18.) Perform 10^5 simulations for each of the following:

1. Check the one sample test for signal-noise ratio. Draw $n = 1024\mathrm{day}$ daily

returns from a normal random generator with $\zeta = 0.25\text{day}^{-1/2}$. Check the hypothesis $H_0 : \zeta = 0$ against the alternative $H_1 : \zeta > 0$. Again, compute p-values and Q-Q plot them against a uniform law.

2. Check the independent two sample test for equality of signal-noise ratio. For $i = 1, 2$, draw $n_i = 512\,i\,\text{day}$ daily returns from two independent normal asset streams, with $\zeta_i = 0.05\,i\,\text{day}^{-1/2}$. Let the population mean of the i^{th} population be $\mu_i = \sqrt{i} \times 10^{-4}\text{day}^{-1}$. Check the hypothesis $H_0 : \zeta_1 = \zeta_2$ against the alternative $H_1 : \zeta_1 < \zeta_2$.
3. Check the independent k sample test for an equation on signal-noise ratio. For $i = 1, \ldots, 7$, draw $n_i = 512\,i\,\text{day}$ daily returns from seven independent normal asset streams, with $\zeta_i = (0.01 + 0.01\,i)\text{day}^{-1/2}$. Let the population mean of the i^{th} population be $\mu_i = i \times 10^{-4}\text{day}^{-1}$. Check the hypothesis $H_0 : \sum_i \zeta_i = 0\,\text{day}^{-1/2}$ against the alternative $H_1 : \sum_i \zeta_i > 0\,\text{day}^{-1/2}$.

Ex. 3.20 **Rescaling parameter in linear attribution model** The rescaling factor $r = \sqrt{\boldsymbol{v}^\top (\mathsf{F}^\top \mathsf{F})^{-1} \boldsymbol{v}}$ appears often in testing of the ex-factor Sharpe ratio, in Section 3.5.6. The factor r appears in the denominator, while in the analagous results for the Sharpe ratio, we often instead see $\sqrt{n}$ in the numerator. We wish to show that r is "morally equivalent" to $n^{-1/2}$.

1. For the case where the linear attribution model expresses Sharpe's model, we take F to be the $n \times 1$ matrix of all ones, and $\boldsymbol{v}$ is a scalar one. In this case show that $r = n^{-1/2}$.
2. Although in this chapter we usually consider the matrix F to be deterministic, assume here that the rows of F are independent draws of a zero-mean l-dimensional Gaussian with identity covariance. Let $\boldsymbol{v}$ be the vector consisting of a single one, and $l - 1$ zeros. Then $\boldsymbol{v}^\top \left(\mathsf{F}^\top \mathsf{F}\right)^{-1} \boldsymbol{v}$ follows an inverse Gamma distribution with shape parameter $\alpha = (n - l + 1)/2$ and scale parameter $\beta = 1/2$, and so $\mathrm{E}\left[1/r^2\right] = n - l + 1$. Confirm this relationship via simulation.

Ex. 3.21 **No TOST for you** Consider the two one-sided test of Section 3.5.4.

1. Is it possible that the critical region is empty? What implications does this have?
2. Under what conditions might the critical region be empty?
3. For the equivalence test and type I rate considered in Example 3.5.5, find sample size n such that the critical region is empty.

Ex. 3.22 **TOST the Market** Use the TOST approach of Section 3.5.4 to test the hypothesis

$$\left|\zeta - 0.5\text{yr}^{-1/2}\right| > 0.1\text{yr}^{-1/2}$$

on the Market returns.

Ex. 3.23 Intersection Union Tests Construct some Intersection Union tests as described in Section 3.5.4.

1. For $i = 1, 2$, given n_i *i.i.d.* draws from Gaussian returns from two assets with signal-noise ratios ζ_i, consider the test for
$$H_0 : |\zeta_2 - m\zeta_1 - b| \geq \epsilon \quad \text{versus} \quad H_1 : |\zeta_2 - \zeta_m - b| < \epsilon,$$
for given a, ϵ. The hope is that one will reject 'towards' the alternative which puts the signal-noise ratios within a band of the line $y = mx + b$. Under what conditions will you reject the null?
2. For the case of ex-factor signal-noise ratio, consider the test for
$$H_0 : \left|\beta^\top \boldsymbol{v} - \sigma c\right| \geq \epsilon \quad \text{versus} \quad H_1 : \left|\beta^\top \boldsymbol{v} - \sigma c\right| < \epsilon,$$
for given $\boldsymbol{v}$ and c. Under what conditions will one reject the null hypothesis?

Ex. 3.24 One sided violations of symmetric confidence interval Prove that when $\zeta \leq \delta$,
$$\Pr\left\{\operatorname{sign}\left(\hat{\zeta}\right)\zeta \geq \left|\hat{\zeta}\right| + \delta\right\} = 0.$$

Ex. 3.25 Narrowest confidence interval on signal-noise ratio Just as the confidence intervals given in Equation 3.27 may not be symmetric about $\hat{\zeta}$, they might not be the narrowest confidence intervals for ζ based on the exact method. By "width", we mean $\zeta_u - \zeta_l$. This question is hard to study theoretically, but easy to consider empirically. Assume $\alpha = 0.05$. Take x from 0.001 to 0.0499. Compute ζ_l and ζ_u via
$$0.95 + x = F_t\left(\sqrt{n}\hat{\zeta}; \sqrt{n}\zeta_l, n - 1\right), \quad x = F_t\left(\sqrt{n}\hat{\zeta}; \sqrt{n}\zeta_u, n - 1\right),$$
then plot $\zeta_u - \zeta_l$ versus x. Assume you have observed $\hat{\zeta} = 0.75\text{yr}^{-1/2}$ based on 4 years of daily data at a rate of 253 days per year. Where do you observe the narrowest confidence interval?

Ex. 3.26 Power I Assume log returns follow an *i.i.d.* Gaussian distribution. Fix the type I rate, α at 0.01, and set the 'combined effect size', $\sqrt{n}\zeta_e = 2.0$. Compute the power of the hypothesis test for $H_0 : \zeta = 0$ against the one sided alternative, $H_0 : \zeta > 0$ via Equation 3.36, for various values of n. Does the power vary much?

Ex. 3.27 Boring power computations Suppose you observe n daily observations of *i.i.d.* Gaussian log returns $x \sim \mathcal{N}(\mu, \sigma)$, assuming 253 days per year. Consider the power of the hypothesis test for $H_0 : \zeta = \zeta_0$ against the one sided alternative, $H_0 : \zeta > \zeta_0$.

1. For $n = 759$, and taking $\alpha = 0.01$, compute the power for testing $\zeta_0 = 1.0\text{yr}^{-1/2}$ assuming $\zeta = 1.5\text{yr}^{-1/2}$.

2. For $n = 759$, and taking $\alpha = 0.01$, how large must ζ be to achieve a power of 90% when testing for $\zeta_0 = 0.75\text{yr}^{-1/2}$?
3. Suppose you wish to test at a type I rate of $\alpha = 0.001$, and achieve a power of 50%, when testing for $\zeta_0 = 0.5\text{yr}^{-1/2}$. Assuming $\zeta = 1.0\text{yr}^{-1/2}$, how large must n be?

* **Ex. 3.28** **Toy hedge fund** Suppose you are starting a hedge fund. After one year, your investors will pull all their money if your fund has not achieved a Sharpe ratio of $1.0\text{yr}^{-1/2}$ over that year, which we will call 'failure'. There are two possible causes of failure: bad luck of an actually good strategy, and the 'normal' poor performance from trading a strategy which had been selected due to a type I error.

1. What signal-noise ratio do you need to bound the probability of failure due to bad luck to be no greater than ϵ? What value do you get for $\epsilon = 0.10$? Let this number be ζ_0.
2. Suppose that you randomly select a trading strategy, backtest it over a 5 year period, then perform a hypothesis test for $H_0 : \zeta = 0$ against $H_1 : \zeta > 0$. If the strategy 'passes' the test, you trade it for a year, otherwise you randomly select another until you find one which does. What is the cutoff for the hypothesis test to achieve a type I rate of $\alpha = 0.05$?
3. To estimate the total probability of failure, you must compute the false discovery rate of the hypothesis test. This requires knowledge of the distribution of signal-noise ratio in the population from which you randomly sample. For simplicity, assume the population is bimodal: with probability 10^{-6}, you select a strategy with $\zeta = \zeta_0$, otherwise you select a strategy with $\zeta = 0\text{yr}^{-1/2}$. Assume that the strategies' returns are independent in the backtest period (this is very unrealistic). What is the false discovery rate of your process? (Here, the false discovery rate is the probability that a strategy which passes the hypothesis test is a type I error, rather than a truly good strategy.)
4. What is the total probability of failure? Note that this should include the possibility that a strategy which is a type I error has actually performed well when traded live.
5. Suppose your fund does not fail, rather your achieved Sharpe ratio passes the hurdle rate. What is the probability that this was due to a type I error strategy?

* **Ex. 3.29** **Toy hedge fund II** Suppose you are starting a hedge fund. You have at your disposal a random stream of strategies which have independent returns.

1. Suppose your sampling process draws strategies with signal-noise ratio $\zeta \sim \mathcal{N}\left(-0.25\text{yr}^{-1/2}, 0.5\text{yr}^{-1/2}\right)$. You draw a strategy, observe 4 years of backtest, perform a hypothesis test for $H_0 : \zeta = 1.0\text{yr}^{-1/2}$ against the alternative $H_1 : \zeta > 1.0\text{yr}^{-1/2}$. Those strategies for which you reject the null, with $\alpha = 0.05$, 'pass' to live trading. The distribution of signal-noise ratio of strategies which pass does not have a closed form. Via simulation, estimate the mean and standard deviation of signal-noise ratio of this distribution. Is it nearly normal? What

is the approximate probability that a randomly drawn strategy will pass the significance test filter? Why is the mean signal-noise ratio so much lower than the hypothesis test cutoff?

2. To improve the quality of passed strategies, one is tempted to use a more stringent test. Perform your simulations again, but testing for $H_0 : \zeta = 1.5\mathrm{yr}^{-1/2}$ against the alternative $H_1 : \zeta > 1.5\mathrm{yr}^{-1/2}$. Does the mean of passed strategies increase? How is the probability of passing the filter affected?

* 3. Can you imagine a case where increasing the cutoff of the hypothesis test (either via requiring a smaller type I rate, or via testing for a more stringent null hypothesis) could lead to a *decrease* in the expected signal-noise ratio of strategies which pass the test?

* **Ex. 3.30 Distribution of optimum Sharpe ratio** Let $p_1, p_2, \ldots, p_k$ be independent random variables each uniform on $[0, 1]$. Consider the j^{th} *order statistic* of the p_i, which is the j^{th} largest of the p_i, call it $p_{(j)}$. It takes a beta distribution:

$$p_{(j)} \sim \mathcal{B}\left(j, k + 1 - j\right).$$

(See also Section 5.1.1.)

1. Suppose you observe k series of returns, each of length n, and each independent of each other. Suppose all returns are normally distributed. Furthermore suppose that the signal-noise ratios are all zero. Compute their Sharpe ratios and consider their order statistics: $\hat{\zeta}_{(1)} \leq \hat{\zeta}_{(2)} \leq \ldots \leq \hat{\zeta}_{(k)}$. What is the distribution of $\hat{\zeta}_{(j)}$?
2. Write code to compute the CDF of $\hat{\zeta}_{(j)}$ under these assumptions.
3. Perform a simulation: generate 1000 days of independent normal returns with zero mean, and compute the Sharpe ratio; repeat this 100 times and record the largest Sharpe ratio observed; repeat that 2500 times and feed these 2500 maximal Sharpe ratio values into your CDF code. The resultant values should be uniform. Q-Q plot then against a uniform law.
4. Write code to compute the quantile function of the j^{th} largest of k independent realizations of the Sharpe ratio for normal returns and a common (possibly nonzero) signal-noise ratio.
5. Suppose you have an army of 100 quants, each of whom generates 10 strategies, all of which are independent and have normally distributed returns. You will pick the strategy which demonstrates the highest Sharpe ratio based on a three year backtest of daily returns, with 252 trading days per year. Your goal for live trading is a signal-noise ratio of $0.9\mathrm{yr}^{-1/2}$. How large should the largest Sharpe ratio be to trade it assuming a type I rate of 0.10?
6. Consider the test of the hypothesis on the maximal signal-noise ratio of a set of strategies with independent returns. That is, consider testing

$$H_0 : \bigvee_{1 \leq i \leq k} \zeta_i = 0$$

against the alternative

$$H_1 : \bigvee_{1 \le i \le k} \zeta_i > 0.$$

This can be tested by computing the CDF of $\zeta_{(k)}$, having observed independent series of normal returns each of size n, and rejecting the null if the CDF exceeds $1 - \alpha$. Does this achieve the nominal type I rate of α? What is the power of this test?

Ex. 3.31 **Freakout intervals** Losses during the early days of a quantitative strategy's live trading often leads neophyte portfolio managers and quants to freak out. Thus one can think of prediction intervals as 'freakout intervals'. Suppose you observe Sharpe ratio of $\hat{\zeta}_1$ based on $n = 1250$day of returns of a strategy. Assume there are 252 trading days per year.

1. Compute approximate 98% prediction intervals for the Sharpe ratio of a further 21day of returns if $\hat{\zeta}_1 = 1\text{yr}^{-1/2}$.
2. Write a function that computes approximate lower 1% prediction bounds for the Sharpe ratio for 21day of future returns, and plot this for values of $\hat{\zeta}_1$ ranging from 0 to $2\text{yr}^{-1/2}$
3. You observed $\hat{\zeta}_1 = 1.5\text{yr}^{-1/2}$ based on $n = 1250$day of returns in backtests. In the first 15 days of trading live, the strategy has achieved a Sharpe ratio of $\hat{\zeta}_2 = -2.0\text{yr}^{-1/2}$. Do you freak out and pull the plug?

* 4. Suppose you observe the past returns of 200 strategies with independent returns over a period of $n = 1250$day, and pick the one with the maximal Sharpe ratio over that period. It achieved a Sharpe ratio of $\hat{\zeta}_1 = 1.75\text{yr}^{-1/2}$. Write a function that computes approximate lower 1% prediction bounds for the Sharpe ratio of this strategy for 21day of future returns.

Ex. 3.32 **Prediction intervals on the Market** Download the *daily* returns of the Market factor, as described in Example 1.2.1. Using the daily data from January 01, 1930 through November 31, 1930, construct a 0.90 prediction interval for the Sharpe ratio for the period December 01, 1930 through December 31, 1930. Compute the actual Sharpe ratio for that period and record whether it was within the 0.90 prediction interval. Repeat this for each year from 1930 through 2020. What proportion of December Sharpes fell within the prediction interval?

* **Ex. 3.33** **Prediction intervals, attribution model** Repeat Example 3.5.14, but compute prediction intervals on the ex-factor Sharpe ratio of the UMD factor attributed against the returns of the Market, SMB and HML. Perform computations on daily returns, and find the empirical coverage.

Ex. 3.34 **LRT** Consider the likelihood ratio test of $H_0 : \zeta = 0.8\,\text{yr}^{-1/2}$ against the unrestricted alternative. Suppose, given 1000 days of Gaussian log returns, you observe $\hat{\zeta} = 1.4\,\text{yr}^{-1/2}$.

1. Compute the MLE, ζ_{MLE}.

2. Compute the LRT statistic, D.
3. Interpret D as a chi-square random variable with 1 degree of freedom. Would you reject H_0 at the $\alpha = 0.01$ level?

Ex. 3.35 **LRT II** Consider the likelihood ratio test of $H_0 : \zeta = 0.5\,\mathrm{yr}^{-1/2}$ versus $H_1 : \zeta = 0.8\,\mathrm{yr}^{-1/2}$ for a Sharpe ratio observed on 2000 days of Gaussian log returns. What is the cutoff Λ_c needed to achieve a type I rate of $\alpha = 0.01$? What is the smallest Sharpe ratio for which this test rejects H_0?

Ex. 3.36 **Likelihood intervals** An α *likelihood interval* for ζ is defined as a set

$$\left\{ \zeta \;\middle|\; \frac{\mathcal{L}_{SR}\left(\zeta \middle| \hat{\zeta}, n\right)}{\mathcal{L}_{SR}\left(\zeta_{\mathrm{MLE}} \middle| \hat{\zeta}, n\right)} \geq \alpha \right\}$$

It is said that a 14.7% likelihood interval will be equivalent to the 95% confidence interval in certain cases. Continue Example 3.7.4 by finding the 14.7% likelihood interval for the signal-noise ratio, given $\hat{\zeta} = 0.1772\mathrm{mo.}^{-1/2}$ for $n = 1128$ months of observations.

Ex. 3.37 **Boring Bayesian computations** Assume a prior of $\zeta_0 = 0.2\mathrm{mo.}^{-1/2}$, $n_0 = 24\mathrm{mo.}$, $\sigma_0 = 1.5\%\mathrm{mo.}^{-1/2}$, $m_0 = 72\mathrm{mo.}$

1. Suppose you observe $n = 60\mathrm{mo.}$ of returns with $\hat{\zeta} = 0.31\mathrm{mo.}^{-1/2}$ and $\hat{\sigma} = 2\%\mathrm{mo.}^{-1/2}$. Compute the posterior hyperparameters.
2. Construct 95% credible intervals on ζ based on that posterior.
3. Compute approximate 99% posterior prediction intervals on $\hat{\zeta}_2$ drawn from a future $n_2 = 12\mathrm{mo.}$ of returns.

Ex. 3.38 **Approximate conjugate prior for signal-noise ratio** Using the normal approximation for Sharpe ratio given in Equation 3.24, construct a approximate conjugate prior for signal-noise ratio. The prior and posterior densities should be normal. That is, the prior should take the form

$$\zeta \propto \mathcal{N}\left(\zeta_0, \gamma_0^2\right), \tag{3.65}$$

where ζ_0 and γ_0^2 are the prior hyperparameters.

1. Assuming normal likelihood for ζ given an observed $\hat{\zeta}$ on n observations, what is the posterior belief?
2. What settings of the prior hyperparameters correspond to a non-informative prior?
3. Continue Example 3.7.4 by finding the 95% credible interval on the Market returns under this approximate posterior, using a non-informative prior.
4. Continue Example 3.7.6 by finding the 95% posterior prediction interval for $\hat{\zeta}_2$ over $n_2 = 12\mathrm{mo.}$, assuming a non-informative prior. (An 'informative' prior

was used in that example, so one does not expect the results to match exactly, unless the data overwhelm the prior.)

Ex. 3.39 Bayesian inference on sum of signal-noise ratios In Section 3.5.3, a Frequentist test is quoted for the null hypothesis

$$H_0 : \sum_i a_i\zeta_i = b \quad \text{versus} \quad H_1 : \sum_i a_i\zeta_i > b,$$

given n_i independent draws from Gaussian returns from k assets with signal-noise ratios ζ_i, and fixed $a_1, a_2, \ldots, a_k, b$. Construct a conjugate Bayesian prior and posterior for the sum $\sum_i a_i\zeta_i$.

Ex. 3.40 Bayesian inference on sum of ex-factor signal-noise ratios In Section 3.5.5, a Frequentist test is quoted for the null hypothesis

$$H_0 : \sum_i \frac{{\beta_i}^\top \boldsymbol{v_i}}{\sigma_i} = c \quad \text{versus} \quad H_1 : \sum_i \frac{{\beta_i}^\top \boldsymbol{v_i}}{\sigma_i} > c,$$

given n_i independent draws from factor models with Gaussian errors on k assets, and fixed $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}, c$. Construct a conjugate Bayesian prior and posterior for the sum $\sum_i {\beta_i}^\top \boldsymbol{v_i}/\sigma_i$.

Ex. 3.41 Market Winter/Summer Consider the hypothesis that the signal-noise ratio of the Market is higher in the summer than the winter. Define the summer as the returns of the months of May through August, inclusive, and the winter as November through February. Use the monthly returns of the Fama-French factor data from **aqfb.data**, using code as given in Example 1.2.1.

1. Compute the Sharpe ratio of the Market for the Summer and Winter, along with confidence intervals.
2. Perform the frequentist hypothesis test for equality of signal-noise ratio of the two periods.

Ex. 3.42 Market takes a weekend Consider the hypothesis that the signal-noise ratio of the Market is higher from Friday close to Monday close than from Wednesday close to Thursday close. Use the daily returns of the Fama-French factor data from **aqfb.data**.

1. Compute the Sharpe ratio of the Market for EOD Mondays and EOD Thursdays, along with confidence intervals.
2. Perform the frequentist hypothesis test for equality of signal-noise ratio of the two periods.

* **Ex. 3.43 Test for bounded non-stationarity** One objection to estimating the signal-noise ratio using long returns histories is that "returns are not stationary." (Indeed, this seems to be the lesson of Example 3.5.13, wherein prediction intervals

failed to achieve nominal coverage for returns of the Market.) One response would be to give up and embrace your own mortality and powerlessness.
An optimist, however, might allow for a small amount of non-stationarity in returns. Suppose you measure returns over k years. Let the signal-noise ratio in the i^{th} year be denoted by ζ_i. Suppose for a given $V, \epsilon > 0$, you wanted to test the null hypothesis

$$H_0 : \zeta_k = V,\ |\zeta_{i+1} - \zeta_i| \leq \epsilon, \text{ for } \ i = 1, 2, \ldots, k-1.$$

1. Using the normal approximation for Sharpe ratio, how would you test for this null hypothesis?
2. How would you construct confidence intervals for ζ_k under the assumption that $|\zeta_{i+1} - \zeta_i| \leq \epsilon$ for $i = 1, 2, \ldots, k-1$?
3. Taking $\epsilon = 0.05\text{yr}^{-1/2}$, construct 95% confidence intervals for the most recent year's signal-noise ratio of the Market.

Ex. 3.44 Prediction intervals, volatility reweighted Market Perform the analysis of Example 3.5.13, but rescaling Market returns by an inverse volatility, as outlined in Section 2.4.2 and Equation 2.14. Compute the rolling twelve month mean absolute return of the Market, delay it by one month, invert it, then normalize to mean 1. You can perform this in **R** as follows:

```
library(aqfb.data)
data(mff4)
rollsum <- function(x, window = 1) {
    cx <- cumsum(x)
    cx - lag(as.numeric(cx), window)
}
mff4$vol_like <- rollsum(abs(mff4$Mkt), 12)
mff4$quietude <- 1/mff4$vol_like
mff4$quietude <- mff4$quietude/mean(mff4$quietude,
    na.rm = TRUE)
```

Based on the entire sample available to you, compute the 95% prediction interval on the Sharpe ratio of inverse volatility weighted returns for a further $n_2 = 12\text{mo.}$ of out-of-sample data.

4. The Sharpe ratio in a rotten world

It's such a fine line between stupid, and uh ... clever.

(David St. Hubbins, This is Spinal Tap*)*

That's Romper-room stuff.

(Arlen Jividen, Algebra Class*)*

In the previous chapter, we considered the distribution of the Sharpe ratio for the case where returns are 1. independent, 2. identically distributed, 3. drawn from a normal distribution 4. independently of previously observable state variables. In fact, probably each of these conditions are violated by real returns series. In this chapter, we will consider the effect of these assumptions, and try to correct[1] for them where necessary. For the most part we will find that the Sharpe ratio is robust to these assumptions, and Mertens' formula for the standard error does a reasonable job of correcting for non-normality of returns.

Here we are tempted to catalog the tests for detecting deviances from assumptions, but for a few issues:

1. For some of these tests, it is almost a foregone conclusion that they will reject the null of *i.i.d.* Gaussian on real returns. For example, normality tests, dozens of which exist each of varying power under different alternatives, often will reject the null for large samples from ostensibly very good Gaussian pseudo-random number generators!
2. Some of these tests provide little indication of the effect size, thus the magnitude of the problem (and whether it matters for inference on the signal-noise ratio) is hard to determine.
3. There is a huge body of literature on each of these topics, far too large for us to summarize.

Instead, we illustrate a few deviances for one dataset, the returns of the Market.

Example 4.0.1 (Is the Market *i.i.d.* normal?). Consider the monthly relative returns of the Market, introduced in Example 1.2.1. First, we apply the **R** function, `shapiro.test` to the monthly returns. As expected, the test rejects, reporting a p-value of 5.2186×10^{-24}.

[1] It should be noted that while much of the research focus on 'fixing' the Sharpe ratio has been on assumption of normality, likely this assumption has a small effect in the real world *for long time horizons*, likely smaller than the problem of omitted variable bias, which is difficult to model.

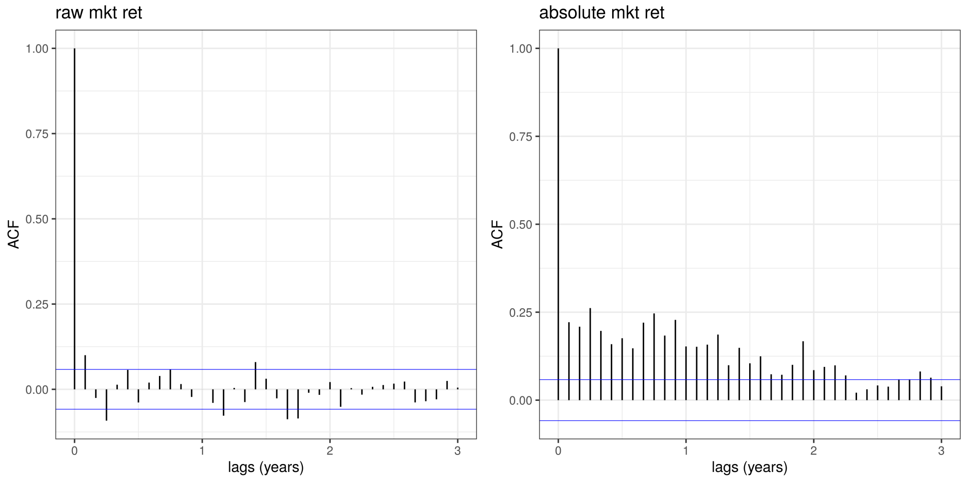

Figure 4.1.: Autocorrelation plots of the raw and absolute monthly relative returns of the Market over the period from Jan 1927 to Dec 2020 are shown.

We calculate the autocorrelation of raw monthly returns, and the autocorrelation of the absolute value of monthly returns. These are plotted for up to 36 lags in Figure 4.1. The first autocorrelation of the raw returns is computed to be approximately 0.1, which seems rather high. One suspects that such a high autocorrelation would have been 'arb'ed' out by market participants. While 'bid-ask bounce' (*cf.* Exercise 2.25) would contribute to an apparent *negative* autocorrelation of returns, it is unlikely to have much of any effect at this time scale. Rather, it turns out that the trade to capture this autocorrelation effect has a very small implied signal-noise ratio. *cf.* Exercise 2.24.

The first autocorrelation of absolute returns is computed to be 0.22. This indicates autoregressive conditional heteroskedasticity ('ARCH'). An interesting question is whether periods of higher volatility are paired with a concomitant increase in mean return, as one theory of heteroskedasticity is that it is caused by a difference in 'market time' and 'wallclock time'. (*cf.* Section 2.4.2.) Under this theory, price discovery occurs faster during periods of higher volatility, thus one should see mean return scale as the *square* of volatility[2]. This does *not* appear to occur for the Market. In Figure 4.2, the mean of daily log returns is plotted against the standard deviation of daily log returns for years from 1926 through 2020. The slope of the regression line is negative, and statistically significantly so at the 0.05 level, but there are many caveats with such a statement: one suspects that significance result is highly sensitive to a few 'outliers[3],' that the results might have been different if one had considered quarterly aggregation instead of annual, that one ought to regress against the log of volatility, and so on.

[2]Another way of viewing this is that the signal-noise ratio should be 'reannualized' to match market time, and thus should increase as the square root of volatility.

[3]It seems a shame to call them outliers, since they correspond to huge macroeconomic shifts *e.g.*, in the years 1931, 2008 and so on.

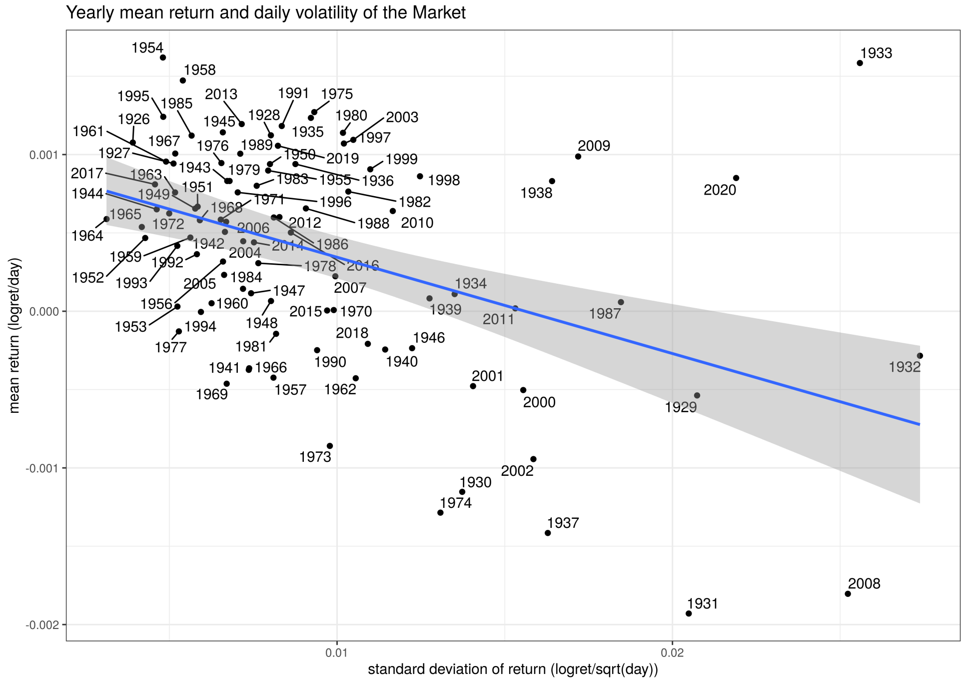

Figure 4.2.: The mean of daily log returns of the Market are plotted versus the standard deviation of daily log returns, for each year from 1926 through 2020.

Nevertheless, there is hardly much evidence for a *positive* relationship between return and volatility in the Market, and one strongly suspects there is some relation between the two. ⊣

4.1. The Sharpe ratio for (non-i.i.d.) Elliptical returns

We now consider the distribution of the Sharpe ratio for the case of general Elliptical returns, relaxing assumptions of independence and identically distributed. That is, consider the returns $x_1, x_2, \ldots, x_n$ as an n-vector, $\boldsymbol{x}$ drawn from an Elliptical distribution with mean $\boldsymbol{\mu}$, covariance Σ, and kurtosis factor κ (*cf.* Section 1.3.2.). We are subsuming the Gaussian case, which is simply $\kappa = 1$.

"Independence" corresponds to a diagonal Σ, while "identically distributed" corresponds to $\boldsymbol{\mu}$ equal to some constant times $\mathbf{1}$, and Σ equal to some constant times I. The sample Sharpe ratio has the form

$$\hat{\zeta} = \sqrt{\frac{n-1}{n}} \frac{\frac{1}{n}\mathbf{1}^{\top}\boldsymbol{x} - r_0}{\sqrt{\frac{\boldsymbol{x}^{\top}\boldsymbol{x}}{n} - \left(\frac{1}{n}\mathbf{1}^{\top}\boldsymbol{x}\right)^2}} = \sqrt{\frac{n-1}{n}} f\left(\boldsymbol{v}\right),$$

where $\boldsymbol{v}$ is the 2-vector defined as

$$\boldsymbol{v} =_{\mathrm{df}} \left[\begin{array}{c} \frac{1}{n}\mathbf{1}^{\top}\boldsymbol{x} \\ \frac{\boldsymbol{x}^{\top}\boldsymbol{x}}{n} - \frac{\left(\mathbf{1}^{\top}\boldsymbol{x}\right)^2}{n^2} \end{array} \right],$$

and $f\left(\boldsymbol{v}\right) =_{\mathrm{df}} \left(v_1 - r_0\right)/\sqrt{v_2}$.

We can easily find the expected value of $\boldsymbol{v}$:

$$\begin{aligned} \mathrm{E}\left[\boldsymbol{v}\right] &= \left[\begin{array}{c} \frac{1}{n}\mathbf{1}^{\top}\boldsymbol{\mu} \\ \frac{\mathrm{tr}\left(\Sigma+\boldsymbol{\mu}\boldsymbol{\mu}^{\top}\right)}{n} - \frac{\mathbf{1}^{\top}\left(\Sigma+\boldsymbol{\mu}\boldsymbol{\mu}^{\top}\right)\mathbf{1}}{n^2} \end{array} \right], \\ &= \frac{1}{n}\left[\begin{array}{c} \mathbf{1}^{\top}\boldsymbol{\mu} \\ \mathrm{tr}\left(\Sigma_c\right) + {\boldsymbol{\mu}_c}^{\top}\boldsymbol{\mu}_c \end{array} \right], \end{aligned} \tag{4.1}$$

where we define

$$\begin{aligned} \boldsymbol{\mu}_c &=_{\mathrm{df}} \boldsymbol{\mu} - \frac{\mathbf{1}\mathbf{1}^{\top}}{n}\boldsymbol{\mu}, \\ \Sigma_c &=_{\mathrm{df}} \Sigma - \frac{\mathbf{1}\mathbf{1}^{\top}}{n}\Sigma. \end{aligned} \tag{4.2}$$

We can view $\boldsymbol{\mu}_c$ and Σ_c as the (column) centered mean and covariance, respectively, where the mean values of each column have been subtracted out. (See Exercise 4.3.) Via Isserlis' theorem (and a lot of computation, *cf.* Exercise 4.4) the covariance matrix of $\boldsymbol{v}$ is [74]

$$\mathrm{Var}\left(\boldsymbol{v}\right) = \frac{1}{n^2}\left[\begin{array}{cc} \mathbf{1}^{\top}\Sigma\mathbf{1} & 2{\boldsymbol{\mu}_c}^{\top}\Sigma\mathbf{1} \\ 2{\boldsymbol{\mu}_c}^{\top}\Sigma\mathbf{1} & \left(\kappa-1\right)\mathrm{tr}\left(\Sigma_c\right)\mathrm{tr}\left(\Sigma_c\right) + 4{\boldsymbol{\mu}_c}^{\top}\Sigma\boldsymbol{\mu}_c + 2\kappa\,\mathrm{tr}\left({\Sigma_c}^2\right) \end{array} \right]. \tag{4.3}$$

By Taylor's theorem,

$$\hat{\zeta} \approx \sqrt{\frac{n-1}{n}}\left[f\left(\mathrm{E}\left[\boldsymbol{v}\right]\right) + \left(\nabla_{\boldsymbol{v}} f\left(\boldsymbol{v}\right)\big|_{\mathrm{E}\left[\boldsymbol{v}\right]}\right)^{\top}\left(\boldsymbol{v} - \mathrm{E}\left[\boldsymbol{v}\right]\right) + \ldots\right]$$

Taking expectations, we have

$$\begin{aligned} \mathrm{E}\left[\hat{\zeta}\right] &\approx \sqrt{\frac{n-1}{n}} f\left(\mathrm{E}\left[\boldsymbol{v}\right]\right), \\ \mathrm{Var}\left(\hat{\zeta}\right) &\approx \frac{n-1}{n}\left(\nabla_{\boldsymbol{v}} f\left(\boldsymbol{v}\right)\big|_{\mathrm{E}\left[\boldsymbol{v}\right]}\right)^{\top}\mathrm{Var}\left(\boldsymbol{v}\right)\left(\nabla_{\boldsymbol{v}} f\left(\boldsymbol{v}\right)\big|_{\mathrm{E}\left[\boldsymbol{v}\right]}\right). \end{aligned} \tag{4.4}$$

For now, we leave underdefined exactly in what sense the approximations hold[4]. The approximate expected Sharpe ratio can be expressed compactly as

$$\mathrm{E}\left[\hat{\zeta}\right] \approx \sqrt{\frac{n-1}{n}}\frac{\frac{1}{n}\mathbf{1}^{\top}\boldsymbol{\mu} - r_0}{\sqrt{\left(\mathrm{tr}\left(\Sigma_c\right) + {\boldsymbol{\mu}_c}^{\top}\boldsymbol{\mu}_c\right)/n}}. \tag{4.5}$$

[4]After all, we have not defined how $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ might change as $n \to \infty$, rendering a discussion of the asymptotics premature.

For the variance, compute the gradient of $f(\cdot)$ as

$$\nabla_{\boldsymbol{v}} f(\boldsymbol{v}) = \begin{bmatrix} \frac{1}{\sqrt{v_2}} \\ -\frac{f(\boldsymbol{v})}{2 v_2} \end{bmatrix}. \tag{4.6}$$

The approximate variance of the Sharpe ratio is thus

$$\begin{aligned} \operatorname{Var}\left(\hat{\zeta}\right) \approx \frac{n-1}{n^2} \frac{\mathbf{1}^\top \Sigma \mathbf{1}}{\left(\operatorname{tr}\left(\Sigma_c\right) + {\boldsymbol{\mu}_c}^\top \boldsymbol{\mu}_c\right)} & \left[1 - 2\left(\frac{\mathbf{1}^\top \boldsymbol{\mu} - n r_0}{\operatorname{tr}\left(\Sigma_c\right) + {\boldsymbol{\mu}_c}^\top \boldsymbol{\mu}_c}\right) \frac{\mathbf{1}^\top \Sigma \boldsymbol{\mu}_c}{\mathbf{1}^\top \Sigma \mathbf{1}} \right. \\ & \left. + \left(\frac{\mathbf{1}^\top \boldsymbol{\mu} - n r_0}{\operatorname{tr}\left(\Sigma_c\right) + {\boldsymbol{\mu}_c}^\top \boldsymbol{\mu}_c}\right)^2 \frac{\frac{\kappa-1}{4}\left(\operatorname{tr}\left(\Sigma_c\right)\right)^2 + {\boldsymbol{\mu}_c}^\top \Sigma \boldsymbol{\mu}_c + \frac{\kappa}{2} \operatorname{tr}\left({\Sigma_c}^2\right)}{\mathbf{1}^\top \Sigma \mathbf{1}}\right]. \end{aligned} \tag{4.7}$$

To get any meaningful simplification of this expression, we must assume some form for $\boldsymbol{\mu}$ and Σ, and approach the problem on a case by case basis. Moreover, the model given here is completely unfalsifiable by data since it posits $n(n+3)/2$ unknown parameters (plus κ) for n observations. We must assume some form for $\boldsymbol{\mu}$ and Σ to keep ourselves honest. Also if $\boldsymbol{\mu}$ and Σ can be completely arbitrary, it is not clear why one would want to perform inference on them, since they would potentially have no relation to future returns.

Keep it normal Despite our enthusiasm for elliptical distributions, one observes odd behavior from them when used to model returns over time of a single asset, instead of correlated contemporaneous returns of multiple assets. For example, consider a multivariate t-distribution with zero mean and covariance $\Sigma = \sigma^2 \mathsf{I}$. The marginals have zero correlation, but are not independent. If you compute the sample standard deviation over the elements, the standard error does not go to 0 as the number of elements increases. See Exercise 4.6. Perhaps real asset returns display this behavior, but it implies a dependence between returns at arbitrary time scales. So for the most part we will consider Equation 4.7 for the case of multivariate Gaussian only, with $\kappa = 1$.

4.1.1. Independent Gaussian returns

Let us simplify the model somewhat to assume independence of returns, which corresponds to a diagonal Σ matrix, say $\Sigma = \operatorname{Diag}\left(\boldsymbol{\sigma}^2\right)$. (*n.b.*, here $\boldsymbol{\sigma}^2$ is a vector of length n whose elements are variances, not volatilities.) Under this assumption, some further simplification is possible. Note that $\Sigma \mathbf{1} = \boldsymbol{\sigma}^2$, and thus $\mathbf{1}^\top \Sigma \mathbf{1} = \mathbf{1}^\top \boldsymbol{\sigma}^2$. Then

$$\begin{aligned} \operatorname{tr}\left(\Sigma_c\right) &= \frac{n-1}{n} \mathbf{1}^\top \boldsymbol{\sigma}^2, \\ \operatorname{tr}\left(\Sigma_c^2\right) &= \frac{n-1}{n} {\boldsymbol{\sigma}^2}^\top \boldsymbol{\sigma}^2 + \left(\frac{\mathbf{1}^\top \boldsymbol{\sigma}^2}{n}\right)^2. \end{aligned}$$

Then, asymptotically, Equation 4.5 simplifies to

$$\mathrm{E}\left[\hat{\zeta}\right] \approx \frac{\frac{1}{n}\mathbf{1}^{\top}\boldsymbol{\mu} - r_0}{\sqrt{\left(\mathbf{1}^{\top}\boldsymbol{\sigma}^2 + \boldsymbol{\mu}_c{}^{\top}\boldsymbol{\mu}_c\right)/n}}. \tag{4.8}$$

We can simply define

$$\zeta_i =_{\mathrm{df}} \frac{\frac{1}{n}\mathbf{1}^{\top}\boldsymbol{\mu} - r_0}{\sqrt{\mathbf{1}^{\top}\boldsymbol{\sigma}^2/n}}, \tag{4.9}$$

as the population quantity we wish to estimate[5], as it is the mean expected return divided by the square root of the mean variance, and thus represents a kind of 'long term average' signal-noise ratio. (In fact, however, the variance is inflated by variation in $\boldsymbol{\mu}$, via the $\boldsymbol{\mu}_c{}^{\top}\boldsymbol{\mu}_c$ term.) Define the geometric bias as

$$b =_{\mathrm{df}} \left(1 + \boldsymbol{\mu}_c{}^{\top}\boldsymbol{\mu}_c/\mathbf{1}^{\top}\boldsymbol{\sigma}^2\right)^{-1/2}. \tag{4.10}$$

We have $0 < b \le 1$, with equality only when $\boldsymbol{\mu} = \mu\mathbf{1}$ for some μ. We can then express Equation 4.8 compactly as $\mathrm{E}\left[\hat{\zeta}\right] \approx b\zeta_i$.

Equation 4.7 becomes, asymptotically,

$$\mathrm{Var}\left(\hat{\zeta}\right) \approx \frac{b^2}{n}\left[1 - 2\zeta_i b^2\sqrt{n}\frac{\boldsymbol{\sigma}^{2\top}\boldsymbol{\mu}_c}{\left(\mathbf{1}^{\top}\boldsymbol{\sigma}^2\right)^{3/2}} + \frac{n}{2}b^4\zeta_i^2\frac{2\boldsymbol{\mu}_c{}^{\top}\left(\boldsymbol{\sigma}^2 \odot \boldsymbol{\mu}_c\right) + \boldsymbol{\sigma}^{2\top}\boldsymbol{\sigma}^2 + \left(\frac{\mathbf{1}^{\top}\boldsymbol{\sigma}^2}{n}\right)^2}{\left(\mathbf{1}^{\top}\boldsymbol{\sigma}^2\right)^2}\right]. \tag{4.11}$$

See Exercise 4.5.

4.1.2. Homoskedastic i.i.d. Gaussian returns

First, we check the computation for the *i.i.d.* case where $\boldsymbol{\mu} = \mu\mathbf{1}$ and $\Sigma = \sigma^2\mathsf{I}$, mostly to check our work, since the answer is given to us already by Equation 3.24. In this case, because $\boldsymbol{\mu}$ is constant, we have $\boldsymbol{\mu}_c = \mathbf{0}$, simplifying the asymptotic expansion, and removing the geometric bias, *i.e.*, $b = 1$. Without geometric bias, the approximate expected value of the signal-noise ratio, from Equation 4.8, becomes

$$\mathrm{E}\left[\hat{\zeta}\right] \approx \frac{\mu - r_0}{\sigma} = \zeta. \tag{4.12}$$

Eliminating the bias term and terms involving $\boldsymbol{\mu}_c$ from Equation 4.11, and using $\zeta_i = \zeta$, we have

$$\begin{aligned}\mathrm{Var}\left(\hat{\zeta}\right) &\approx \frac{1}{n}\left[1 + \frac{n}{2}\zeta^2\frac{\boldsymbol{\sigma}^{2\top}\boldsymbol{\sigma}^2 + \left(\frac{\mathbf{1}^{\top}\boldsymbol{\sigma}^2}{n}\right)^2}{\left(\mathbf{1}^{\top}\boldsymbol{\sigma}^2\right)^2}\right],\\ &= \frac{1}{n}\left[1 + \frac{n}{2}\zeta^2\frac{n\sigma^4 + \sigma^4}{n^2\sigma^4}\right].\end{aligned} \tag{4.13}$$

[5] Though it is arguable that this is really the quantity of interest. In some cases, it is clearly of interest, in others the case is not as clear.

Asymptotically in n, this is the approximate standard error of Johnson and Welch, as given in Equation 3.24. [99, 77].

4.1.3. Heteroskedastic independent Gaussian returns

Now relax the homoskedastic independent case considered previously by assuming that the observed x are homoskedastic *i.i.d.* Gaussian returns 'polluted' by some variable l. Such a returns stream could arise in a number of ways:

1. As introduced in Section 2.4.2, one model of market heteroskedasticity is that price discovery unfolds along a 'market clock' instead of wall clock time. In this (somewhat optimistic) model, if the underlying returns stream has constant signal-noise ratio when measured in market time, the observed returns will have mean and variance equally scaled by l, the 'speed' of price discovery. (Though this is not often supported by data, *cf.* Example 2.4.3 and Example 4.0.1.)
   Under this model the expected value and variance vary from their long term mean values in lockstep. It must be noted that this is a somewhat optimistic model of market heteroskedasticity; the pessimistic alternative is that expected returns are unaffected (or even adversely impacted) by the variable l, which only scales variance.
2. Similarly, if returns of the asset were homoskedastic *i.i.d.*, but were measured over irregular time points, say weekly returns jumbled together with daily returns, similar population dynamics would arise.

The optimistic and pessimistic model mentioned above can be couched in a more general model where the expected return is scaled by l^{λ} for some constant λ. The optimistic model is captured by $\lambda = 1$, while $\lambda = 0$ corresponds to the pessimistic model. One could also consider, say, $\lambda = \frac{1}{2}$ as a compromise value. If underlying returns were homoskedastic *i.i.d.*, but one observed *levered* returns, with leverage chosen *e.g.*, by a fund manager, then expected return and *volatility* would scale together, resulting in $\lambda = \frac{1}{2}$.

So assume the existence of a vector $\boldsymbol{l}$ with strictly positive elements. Define M_k as the empirical raw k^{th} moment of $\boldsymbol{l}$:

$$M_k =_{\text{df}} \frac{1}{n} \sum_{1 \le i \le n} {l_i}^{k}. \tag{4.14}$$

Without loss of generality, we can assume that $\boldsymbol{l}$ has been rescaled such that $M_1 = 1$.

Now suppose λ is given, and $\boldsymbol{l}$ scales variance directly, and expectation via λ, *i.e.*, $\Sigma = \sigma^2 \operatorname{Diag}(\boldsymbol{l})$, and $\boldsymbol{\mu} = \mu \boldsymbol{l}^{\lambda}$, where here a vector to a power is interpreted as the Hadamard (elementwise) power. Since we are assuming independence of returns, we can rely on the simplified form of expectation and standard error. Note that

$$\boldsymbol{\mu}_c = \mu \boldsymbol{l}^{\lambda} - \mu M_{\lambda} \mathbf{1}, \quad \text{and } \boldsymbol{\sigma}^2 = \sigma^2 \boldsymbol{l}.$$

Then Equation 4.9 becomes

$$\zeta_i = \frac{\mu M_{\lambda} - r_0}{\sigma}, \tag{4.15}$$

and the geometric bias term of the sample Sharpe ratio is

$$b = \left(1 + \frac{\mu^2 \left(M_{2\lambda} - M_{\lambda}^2\right)}{\sigma^2}\right)^{-1/2}. \tag{4.16}$$

Note that when $\lambda = 0$, then $M_{2\lambda} - M_{\lambda}^2 = 0$, and there is no bias: $b = 1$. When $\lambda = 1$, the term $\left(M_{2\lambda} - M_{\lambda}^2\right)$ is effectively the squared coefficient of variation of $\boldsymbol{l}$. It balances that of the underlying returns in the geometric bias term. Again, it is not unambiguously the case that ζ_i is the population quantity of interest, since variation within $\boldsymbol{\mu}$ should probably be counted as volatility of returns, especially in the case where $\boldsymbol{l}$ is stochastic, rather than the case of jumbled periodicities. Ignoring this point, we can proceed as previously.

To compute the standard error of the Sharpe ratio, we have to compute a few more terms. We have

$$\begin{aligned}
\mathbf{1}^{\top}\boldsymbol{\sigma}^2 &= n\sigma^2 M_{1+\lambda},\\
{\boldsymbol{\sigma}^2}^{\top}\boldsymbol{\mu}_c &= n\mu\sigma^2\left(M_{1+\lambda} - M_1 M_{\lambda}\right) = n\mu\sigma^2\left(M_{1+\lambda} - M_{\lambda}\right),\\
{\boldsymbol{\sigma}^2}^{\top}\boldsymbol{\sigma}^2 &= n\sigma^4 M_2,\\
{\boldsymbol{\mu}_c}^{\top}\left(\boldsymbol{\sigma}^2 \odot \boldsymbol{\mu}_c\right) &= n\mu^2\sigma^2 M_{1+2\lambda}.
\end{aligned} \tag{4.17}$$

Plugging these into Equation 4.11 gives the approximate variance of the Sharpe ratio as

$$\operatorname{Var}\left(\hat{\zeta}\right) \approx \frac{b^2}{n}\left[1 - 2\zeta_i b^2 \frac{\mu\left(M_{1+\lambda} - M_{\lambda}\right)}{\sigma M_{1+\lambda}^{3/2}} + \frac{b^4}{2}\zeta_i^2 \frac{2\frac{\mu^2}{\sigma^2}M_{1+2\lambda} + M_2 + \mathcal{O}\left(n^{-1}\right)}{M_{1+\lambda}^2}\right]. \tag{4.18}$$

Further simplification is possible for various values of λ. Note, however, that in some situations it appears to be the case that ζ (and even the debiased version, *viz.* $\sqrt{b}\zeta$) has a *lower* standard error than in the homoskedastic *i.i.d.* case (*cf.* Equation 4.13). This can occur if the errors in estimating the mean return and volatility are positively correlated. (Equivalently, ${\boldsymbol{\sigma}^2}^{\top}\boldsymbol{\mu}_c > 0$ and sufficiently large.) Thus estimation error in the two components of the Sharpe ratio counteract each other, resulting in a (ever so slightly) decreased standard error. In practice this will have very little effect.

Sweeping out small terms (for most cases μ^2/σ^2 will be tiny), for a few special cases of λ we have

$$\begin{aligned}
(\lambda = 0)\quad \operatorname{Var}\left(\hat{\zeta}\right) &\approx \frac{b^2}{n}\left[1 + \frac{b^4}{2}\zeta_i^2\frac{M_2}{M_1^2}\right] = \frac{1}{n}\left[1 + \frac{1}{2}\zeta_i^2\frac{M_2}{M_1^2}\right],\\
(\lambda = 1/2)\quad \operatorname{Var}\left(\hat{\zeta}\right) &\approx \frac{b^2}{n}\left[1 - 2\zeta_i b^2\frac{\mu\left(M_{3/2} - M_{1/2}\right)}{\sigma M_{3/2}^{3/2}} + \frac{b^4}{2}\zeta_i^2\frac{M_2^{-1/2}}{M_{3/2}^2}\right],\\
(\lambda = 1)\quad \operatorname{Var}\left(\hat{\zeta}\right) &\approx \frac{b^2}{n}\left[1 - 2\zeta_i b^2\frac{\mu\left(M_2 - 1\right)}{\sigma M_2^{3/2}} + \frac{b^4}{2}\zeta_i^2 M_2^{-1}\right].
\end{aligned} \tag{4.19}$$

It turns out that when variation in l is modest and the signal-noise ratio is realistic, the geometric bias introduced into the Sharpe ratio by heteroskedasticity has little effect. Moreover the standard error of the Sharpe ratio is not much changed either, and can even be slightly decreased. These findings are illustrated in the following examples which use the (rescaled) VIX index as $\boldsymbol{l}$. In summary,

Heteroskedasticity and Sharpe ratio

Modest heteroskedasticity causes only mild bias in the Sharpe ratio and very little effect on standard error.

Example 4.1.1 (VIX and heteroskedastic bias). Consider the (rescaled to unit mean) VIX index introduced in Example 1.2.4, with data from 1990-01-02 through 2020-12-31. For values of k between 0 and 2, empirical estimates of M_k range (non-monotonically) from around 0.98 to 1.2.

Consider the case where $\mu = 0.0007\text{day}^{-1}$, and $\sigma = 0.013\text{day}^{-1/2}$. For values of λ tested between 0 and 2, the geometric bias term, b, ranged from around 0.9976 to 1, with b monotonically decreasing in λ. ⊣

Example 4.1.2 (VIX and Sharpe ratio standard error). Continuing Example 4.1.1, consider the (rescaled) VIX index of Example 1.2.4, with data from 1990-01-02 through 2020-12-31. Suppose $\mu = 0.0007\text{day}^{-1}$, and $\sigma = 0.013\text{day}^{-1/2}$. Take $r_0 = 0\text{day}^{-1}$, and suppose n is large.

For values of λ tested between 0 and 2, n/b^2 times the approximate variance from Equation 4.18 was computed, and found to range from around 0.9993 to 1.002, monotonically decreasing in λ.

The equivalent value for the homoskedastic *i.i.d.* case, via Equation 4.13, is around 1.0014. For values of λ greater than around 0.22, the heteroskedastic case has a (slightly) lower standard error. ⊣

Example 4.1.3 (Weekly and daily returns). Suppose you have observed 260 weeks of hypothetical weekly returns and 260 days of daily returns for a strategy, with the returns independent. Suppose the returns are Gaussian with $\mu = 0.001\text{day}^{-1}$, and $\sigma = 0.018\text{day}^{-1/2}$. Because of how expectation and variance scale, we have $\lambda = 1$. To get $M_1 = 1$, we define $\boldsymbol{l}$ to be 260 values of $5/3$ and 260 values of $1/3$. This means that if we compute the Sharpe ratio naïvely, jumbling together daily and weekly returns, we are effectively estimating μ and σ (and thus ζ_i) at the scale of 3 day.

For these values of μ and σ, the geometric bias is quite modest: $b \approx 0.9979$. The standard error of the (biased) Sharpe ratio is around 0.0437. This is the standard error around estimation of signal-noise ratio at the 3 day scale. If we convert the 3 day Sharpe ratio to daily scale, the standard error of the (still slightly biased) daily Sharpe ratio is $0.0252\text{day}^{-1/2}$. If we had, instead, observed 1560 day of *daily* returns, the standard error, via Equation 4.13, is approximately $0.0253\text{day}^{-1/2}$, effectively the same. We have lost little, if anything, in terms of bias or variance by pooling weekly and daily returns together.

To check these results, we perform 10000 Monte Carlo simulations of this situation. The mean Sharpe ratio over these realizations is $0.0554\text{day}^{-1/2}$, while the population value is $0.0556\text{day}^{-1/2}$. The standard deviation of the Sharpe ratio over these 10000 realizations is around $0.0251\text{day}^{-1/2}$, which is consistent with the approximate theoretical value of $0.0252\text{day}^{-1/2}$. See also Exercise 4.11. ⊣

Example 4.1.4 (Daily and Quarterly returns, the Market). We consider the Market returns, introduced in Example 1.2.1. We observe 172 quarterly returns from the period Mar 1927 to Dec 1969, and 12361 daily returns from 1970-01-02 to 2018-12-31. Observing only these, we wish to compute the Sharpe ratio. As this is a 'jumbled periodicity' problem, we have $\lambda = 1$. We assume 252 days per year. To To get $M_1 = 1$, we define $\boldsymbol{l}$ to be 172 values of 63/1.8509 and then 12361 values of 1/1.8509. If we compute the Sharpe ratio naïvely on the quarterly and daily returns, we are effectively estimating μ and σ (and thus ζ_i) at the scale of 1.8509 day. Recast in annual units we compute the Sharpe ratio as $0.5219\text{yr}^{-1/2}$. In Example 3.3.4, we estimated the Sharpe ratio to be $0.5278\text{yr}^{-1/2}$ based on quarterly values.

Plugging in the long term sample values for μ and σ, the geometric bias is $b \approx 0.9788$. Dividing the Sharpe ratio by this bias gives us a geometrically nearly unbiased Sharpe ratio of $0.5332\text{yr}^{-1/2}$. The standard error of the (biased) Sharpe ratio is around 0.0087, at the 1.8509 day scale. Annualized the standard error is $0.1387\text{yr}^{-1/2}$. ⊣

See also Exercise 4.9 through Exercise 4.12.

Remark (Weighted estimation?). For the case where the $\boldsymbol{l}$ are known (*e.g.*, jumbled weekly and daily returns) or can be roughly estimated, one is tempted to perform some kind of weighted estimation to correct for the heteroskedasticity. It is not clear, however, that by so doing, one reduces the standard error of the Sharpe ratio. Indeed, in some cases the approximate standard error in the heteroskedastic case seems lower than that in the homoskedastic case. Perhaps a better approach would be to correct for the approximate geometric bias, though this should be very close to 1 for most applications. On the other hand, arguably the apparent reduction in standard error is due to our insistence that the population parameter of interest does not count variation in $\boldsymbol{\mu}$ towards volatility of returns.

4.1.4. Homoskedastic autocorrelated Gaussian returns

Now consider the case of constant mean and errors drawn from an AR(1) process with autocorrelation ρ. [24] Thus we have again $\boldsymbol{\mu} = \mu\mathbf{1}$, and $\Sigma_{i,j} = \sigma^2\rho^{|i-j|}$. We will assume that $|\rho|$ is bounded away from one, say smaller than one half. Because $\boldsymbol{\mu}$ is constant, we have $\boldsymbol{\mu}_c = \mathbf{0}$, and $\mathbf{1}^\top\boldsymbol{\mu} = n\mu$. We will need to compute the quantity $\mathbf{1}^\top\mathbf{\Sigma}\mathbf{1}$. For large n this is approximately

$$\mathbf{1}^\top\mathbf{\Sigma}\mathbf{1} \approx n\sigma^2\left(\frac{1+\rho}{1-\rho} + \frac{1}{n}\frac{2\rho}{(1-\rho)^2}\right) = n\sigma^2\left(\frac{1+\rho}{1-\rho} + \mathcal{O}\left(n^{-1}\right)\right), \tag{4.20}$$

cf. Exercise 4.7. Similarly,

$$\operatorname{tr}\left(\boldsymbol{\Sigma}_c\right) = n\sigma^2 - \frac{\mathbf{1}^{\top}\boldsymbol{\Sigma}\mathbf{1}}{n} \approx n\sigma^2\left(1 - \mathcal{O}\left(n^{-1}\right)\right),$$
$$\operatorname{tr}\left(\boldsymbol{\Sigma}_c^2\right) \approx n\sigma^4\left(\frac{1+\rho^2}{1-\rho^2} - \mathcal{O}\left(n^{-1}\right)\right).$$

The approximation of Equation 4.5 becomes

$$\operatorname{E}\left[\hat{\zeta}\right] \approx \sqrt{\frac{n-1}{n}}\frac{\mu - r_0}{\sqrt{\sigma^2\left(1 - \mathcal{O}\left(n^{-1}\right)\right)}} \to \frac{\mu - r_0}{\sigma}. \tag{4.21}$$

Thus the Sharpe ratio is asymptotically unbiased in n. The standard error, however, is somewhat affected by the autocorrelation. Picking up from Equation 4.7,

$$\begin{aligned}
\operatorname{Var}\left(\hat{\zeta}\right) &\approx \frac{n-1}{n^2}\frac{\mathbf{1}^{\top}\boldsymbol{\Sigma}\mathbf{1}}{\operatorname{tr}\left(\boldsymbol{\Sigma}_c\right)}\left[1 + \left(\frac{\mathbf{1}^{\top}\boldsymbol{\mu} - nr_0}{\operatorname{tr}\left(\boldsymbol{\Sigma}_c\right)}\right)^2 \frac{\frac{1}{2}\operatorname{tr}\left({\boldsymbol{\Sigma}_c}^2\right)}{\mathbf{1}^{\top}\boldsymbol{\Sigma}\mathbf{1}}\right],\\
&\approx \frac{1}{n}\frac{1}{n\sigma^2}\left(n\sigma^2\frac{1+\rho}{1-\rho} + \frac{1}{2}\left(\frac{\mu - r_0}{\sigma^2}\right)^2 n\sigma^4\left(\frac{1+\rho^2}{1-\rho^2} + \mathcal{O}\left(n^{-1}\right)\right)\right),\\
&\approx \frac{1}{n}\left(\frac{1+\rho}{1-\rho} + \frac{1}{2}\left(\frac{\mu - r_0}{\sigma}\right)^2\left(\frac{1+\rho^2}{1-\rho^2} + \mathcal{O}\left(n^{-1}\right)\right)\right),\\
&\approx \frac{1}{n}\frac{1+\rho}{1-\rho}\left(1 + \frac{1+\rho^2}{\left(1+\rho\right)^2}\frac{1}{2}\left(\frac{\mu - r_0}{\sigma}\right)^2\right).
\end{aligned} \tag{4.22}$$

Compare this with the standard error for the homoskedastic case (Equation 4.13, equivalently Equation 3.24). For small signal-noise ratio and small ρ, the standard error of the Sharpe ratio is approximately $n^{-1/2}\sqrt{\left(1+\rho\right)/\left(1-\rho\right)}$. This corresponds to the standard error of the t statistic for AR(1) variates under the null $\mu = 0$ as described by van Belle. [162, sec. 8.7] The 'small angle' approximation for this correction is $1 + 2\rho$, which is reasonably accurate for $|\rho| < 0.1$. In summary:

Autocorrelation and Sharpe ratio

A small autocorrelation of ρ inflates the standard error of the Sharpe ratio by about $200\rho\%$.

Since we expect ρ to be very small for real world returns[6], autocorrelation should have little to no effect on the bias or standard error of the Sharpe ratio.

Example 4.1.5 (AR(1) returns and Sharpe ratio standard error). A Monte Carlo study confirms the standard error of the Sharpe ratio given in Equation 4.22. For values of

[6] While persistent significant autocorrelation should be 'arb'ed out' by market participants, significant autocorrelation at long time scales may be observed in real markets, *cf.* Example 4.0.1.

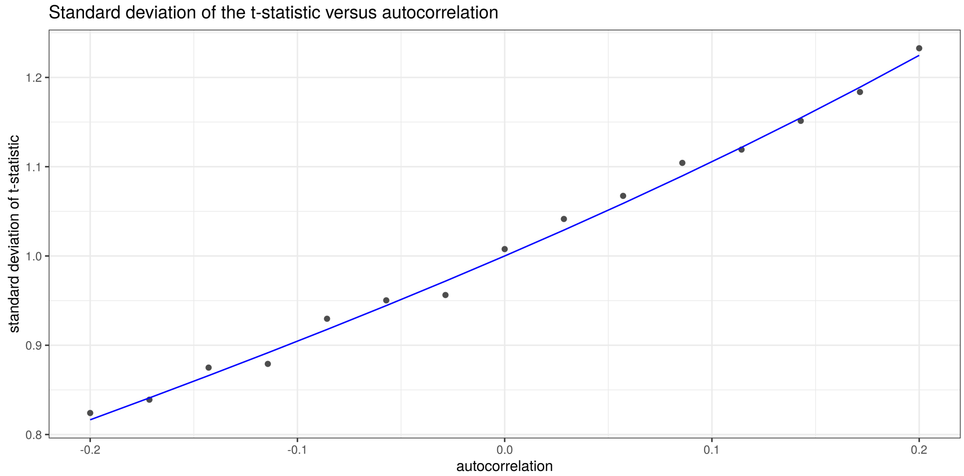

Figure 4.3.: The empirical standard deviation of the t-statistic (*i.e.*, the rescaled Sharpe ratio) is plotted versus the autocorrelation, ρ. Each point represents 8,000 replications of approximately 3 years of daily data, with each series generated by an AR(1) process with normal innovations and $\mu = 0$. The line is $y = \sqrt{(1 + \rho)/(1 - \rho)}$, as given by Equation 4.22, *not* a fit of the data.

ρ ranging from -0.2 to 0.2, 8000 realizations of approximately 4 years of daily data generated by an AR(1) process with $\mu = 0$ were generated. The Sharpe ratio of each was computed, and the empirical standard deviation over realizations was computed. This is multiplied by $\sqrt{n}$ and then plotted versus ρ in Figure 4.3. The fit value of $\sqrt{(1 + \rho)/(1 - \rho)}$ is also shown. See also Exercise 2.29. ⊣

Example 4.1.6 (Autocorrelated Market returns and Sharpe ratio standard error). Consider the returns of the Market, introduced in Example 1.2.1. Based on 1128 monthly returns from Jan 1927 to Dec 2020, we estimate the autocorrelation of returns to be 0.1001. Feeding the sample estimates of ρ, μ and σ into Equation 4.22, we estimate the standard error of the Sharpe ratio to be $0.0174\text{yr}^{-1/2}$. If the autocorrelation is assumed to be zero, we would estimate the standard error as $0.0143\text{yr}^{-1/2}$. The inflation caused by autocorrelation is approximately 22%, which is close to twice the estimated autocorrelation. ⊣

Caution (Returns are autocorrelated). By the way that they are constructed, returns of an asset are typically slightly *negatively* autocorrelated. That is, since returns are defined as $x_{t+1} = \log p_{t+1}/p_t$, where p_t are the mark prices, if there is any amount of 'error' in the marks, it will create negatively autocorrelated returns. This is discussed in the context of the 'bounce effect', *cf.* Exercise 2.25. In general, if the 'true' returns are *i.i.d.* with variance σ_m^2, and the log mark prices have a noise term with variance σ_b^2, then the autocorrelation of observed returns is

$$\rho = \frac{-\sigma_b^2}{2\sigma_b^2 + \sigma_m^2}$$

Given that σ_b^2/σ_m^2 is likely to be on the order of 0.01 or smaller, the bias introduced by this bounce effect should be very small indeed.

4.2. Asymptotic Distribution of Sharpe ratio

Here we derive the asymptotic distribution of Sharpe ratio, following Jobson and Korkie *inter alia.* [76, 99, 111, 89, 96, 174] Consider the case of p possibly correlated returns streams, with each observation denoted by the p-vector $\boldsymbol{x}$. Let $\boldsymbol{\mu}$ be the p-vector of population means, and let $\boldsymbol{\alpha}_2$ be the p-vector of the uncentered second moments. Let $\boldsymbol{\zeta}$ be the vector of signal-noise ratio of the assets. Let r_0 be the 'risk free rate'. We have

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

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

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

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

Under the multivariate central limit theorem [166]

$$\sqrt{n}\left(\left[\hat{\boldsymbol{\mu}}^\top, \hat{\boldsymbol{\alpha}}_2{}^\top\right]^\top - \left[\boldsymbol{\mu}^\top, {\boldsymbol{\alpha}_2}^\top\right]^\top\right) \rightsquigarrow \mathcal{N}\left(0, \Omega\right). \tag{4.23}$$

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

$$\sqrt{n}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}\right) \rightsquigarrow \mathcal{N}\left(0, \left(\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, {\boldsymbol{\alpha}_2}^\top]^\top}\right) \Omega \left(\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, {\boldsymbol{\alpha}_2}^\top]^\top}\right)^\top\right). \tag{4.24}$$

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

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, {\boldsymbol{\alpha}_2}^\top]^\top} &= \left[\ \mathrm{Diag}\left(\tfrac{\boldsymbol{\alpha}_2 - \boldsymbol{\mu} r_0}{(\boldsymbol{\alpha}_2 - \boldsymbol{\mu}^2)^{3/2}}\right) \quad \mathrm{Diag}\left(\tfrac{r_0 - \boldsymbol{\mu}}{2(\boldsymbol{\alpha}_2 - \boldsymbol{\mu}^2)^{3/2}}\right)\ \right], \\ &= \left[\ \mathrm{Diag}\left(\tfrac{\boldsymbol{\sigma} + \boldsymbol{\mu}\boldsymbol{\zeta}}{\boldsymbol{\sigma}^2}\right) \quad \mathrm{Diag}\left(\tfrac{-\boldsymbol{\zeta}}{2\boldsymbol{\sigma}^2}\right)\ \right]. \end{aligned} \tag{4.25}$$

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, where we define the vector $\boldsymbol{\sigma} =_{\mathrm{df}} \left(\boldsymbol{\alpha}_2 - \boldsymbol{\mu}^2\right)^{1/2}$, with powers taken elementwise.

In practice, the population values, $\boldsymbol{\mu}$, $\boldsymbol{\alpha}_2$, 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, taken from estimating the covariance of the samples of $\boldsymbol{x}$ and $\boldsymbol{x}^2$ stacked, using Equation 4.25 we have the approximation

$$\hat{\boldsymbol{\zeta}} \approx \mathcal{N}\left(\boldsymbol{\zeta}, \frac{1}{n}\left[\ \mathrm{Diag}\left(\tfrac{\hat{\boldsymbol{\sigma}} + \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\zeta}}}{\hat{\boldsymbol{\sigma}}^2}\right) \quad \mathrm{Diag}\left(\tfrac{-\hat{\boldsymbol{\zeta}}}{2\hat{\boldsymbol{\sigma}}^2}\right)\ \right] \hat{\Omega} \left[\begin{array}{c} \mathrm{Diag}\left(\tfrac{\hat{\boldsymbol{\sigma}} + \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\zeta}}}{\hat{\boldsymbol{\sigma}}^2}\right) \\ \mathrm{Diag}\left(\tfrac{-\hat{\boldsymbol{\zeta}}}{2\hat{\boldsymbol{\sigma}}^2}\right) \end{array}\right]\right), \tag{4.26}$$

where we have plugged in sample estimates. [99, 111]

Example 4.2.1 (Elliptically distributed returns). Consider the case where $\boldsymbol{x}$ is drawn from an elliptical distribution with mean $\boldsymbol{\mu}$, covariance Σ, and kurtosis factor κ. When returns are Gaussian, $\kappa = 1$. Then we have

$$\Omega = \left[\begin{array}{cc} \Sigma & 2\Sigma\,\mathrm{Diag}\left(\boldsymbol{\mu}\right) \\ 2\,\mathrm{Diag}\left(\boldsymbol{\mu}\right)\Sigma & (\kappa - 1)\,\mathrm{diag}\left(\Sigma\right)\left(\mathrm{diag}\left(\Sigma\right)\right)^\top + 2\kappa\Sigma \odot \Sigma + 4\,\mathrm{Diag}\left(\boldsymbol{\mu}\right)\Sigma\,\mathrm{Diag}\left(\boldsymbol{\mu}\right) \end{array}\right]. \tag{4.27}$$

cf. Exercise 4.2.

Let R be the correlation matrix of the returns, defined as

$$\mathsf{R} =_{\mathrm{df}} \mathrm{Diag}\left(\boldsymbol{\sigma}^{-1}\right)\Sigma\,\mathrm{Diag}\left(\boldsymbol{\sigma}^{-1}\right), \tag{4.28}$$

where $\boldsymbol{\sigma}$ is the (positive) square root of the diagonal of Σ. Then Equation 4.26 becomes

$$\hat{\boldsymbol{\zeta}} \approx \mathcal{N}\left(\boldsymbol{\zeta}, \frac{1}{n}\left(\mathsf{R} + \frac{\kappa - 1}{4}\boldsymbol{\zeta}\boldsymbol{\zeta}^\top + \frac{\kappa}{2}\mathrm{Diag}\left(\boldsymbol{\zeta}\right)\left(\mathsf{R} \odot \mathsf{R}\right)\mathrm{Diag}\left(\boldsymbol{\zeta}\right)\right)\right). \tag{4.29}$$

cf. Exercise 4.36, and Exercise 4.37. Note how in the case of scalar Gaussian returns, this reduces to Equation 3.25.

⊣

4.2.1. Frequentist analysis

Equation 4.26 can be used to perform hypothesis tests and construct confidence intervals with approximately nominal coverage. To be explicit, we outline those procedures here.

test of single contrast on multiple signal-noise ratios Suppose $\boldsymbol{v}$ is a fixed p-vector, and c a fixed scalar. Then to test the null hypothesis:

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

we first compute $\hat{\boldsymbol{\zeta}}$, then estimate $\hat{\Omega}$, either by the empirical covariance of the stacked vector of samples of $\boldsymbol{x}$ and $\boldsymbol{x}^2$, or some other method. Then reject the null if the statistic

$$z = \frac{\boldsymbol{v}^\top \hat{\boldsymbol{\zeta}} - c}{\sqrt{n\boldsymbol{v}^\top \left(\left[\begin{array}{cc} \operatorname{Diag}\left(\frac{\hat{\boldsymbol{\sigma}}+\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\zeta}}}{\hat{\boldsymbol{\sigma}}^2}\right) & \operatorname{Diag}\left(\frac{-\hat{\boldsymbol{\zeta}}}{2\hat{\boldsymbol{\sigma}}^2}\right) \end{array} \right] \hat{\Omega} \left[\begin{array}{c} \operatorname{Diag}\left(\frac{\hat{\boldsymbol{\sigma}}+\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\zeta}}}{\hat{\boldsymbol{\sigma}}^2}\right) \\ \operatorname{Diag}\left(\frac{-\hat{\boldsymbol{\zeta}}}{2\hat{\boldsymbol{\sigma}}^2}\right) \end{array} \right] \right) \boldsymbol{v}}}$$

is less than z_α, the α quantile of the standard normal distribution. Alternatively, if one assumes returns are drawn from an elliptical distribution, instead use the statistic

$$z = \frac{\boldsymbol{v}^\top \hat{\boldsymbol{\zeta}} - c}{\sqrt{n\boldsymbol{v}^\top \left(\hat{\mathsf{R}} + \frac{\hat{\kappa}-1}{4}\hat{\boldsymbol{\zeta}}\hat{\boldsymbol{\zeta}}^\top + \frac{\hat{\kappa}}{2}\operatorname{Diag}\left(\hat{\boldsymbol{\zeta}}\right)\left(\hat{\mathsf{R}} \odot \hat{\mathsf{R}}\right)\operatorname{Diag}\left(\hat{\boldsymbol{\zeta}}\right)\right)\boldsymbol{v}},}$$

where $\hat{\kappa}$ and $\hat{\mathsf{R}}$ are sample estimates of the kurtosis factor and correlation matrix respectively.

test of multiple signal-noise ratios Suppose $\boldsymbol{\zeta}_0$ is some fixed p-vector. Then to test the null hypothesis:

$$H_0 : \boldsymbol{\zeta} = \boldsymbol{\zeta}_0 \quad \text{versus} \quad H_1 : \boldsymbol{\zeta} \neq \boldsymbol{\zeta}_0,$$

we first compute $\hat{\boldsymbol{\zeta}}$, then estimate $\hat{\Omega}$ as described above. Then reject the null if the statistic

$$c^2 = n\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right)^\top \left(\left[\begin{array}{cc} \operatorname{Diag}\left(\frac{\hat{\boldsymbol{\sigma}}+\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\zeta}}}{\hat{\boldsymbol{\sigma}}^2}\right) & \operatorname{Diag}\left(\frac{-\hat{\boldsymbol{\zeta}}}{2\hat{\boldsymbol{\sigma}}^2}\right) \end{array} \right] \hat{\Omega} \left[\begin{array}{c} \operatorname{Diag}\left(\frac{\hat{\boldsymbol{\sigma}}+\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\zeta}}}{\hat{\boldsymbol{\sigma}}^2}\right) \\ \operatorname{Diag}\left(\frac{-\hat{\boldsymbol{\zeta}}}{2\hat{\boldsymbol{\sigma}}^2}\right) \end{array} \right] \right)^{-1} \left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right) \tag{4.30}$$

is greater than ${\chi^2}_{1-\alpha}(p)$, the $1-\alpha$ quantile of the chi-square distribution with p degrees of freedom. Alternatively, if one assumes returns are drawn from an elliptical distribution, instead use the statistic

$$c^2 = n\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right)^\top \left(\hat{\mathsf{R}} + \frac{\hat{\kappa}-1}{4}\hat{\boldsymbol{\zeta}}\hat{\boldsymbol{\zeta}}^\top + \frac{\hat{\kappa}}{2}\operatorname{Diag}\left(\hat{\boldsymbol{\zeta}}\right)\left(\hat{\mathsf{R}} \odot \hat{\mathsf{R}}\right)\operatorname{Diag}\left(\hat{\boldsymbol{\zeta}}\right)\right)^{-1}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right), \tag{4.31}$$

where $\hat{\kappa}$ and $\hat{\mathsf{R}}$ are sample estimates of the kurtosis factor and correlation matrix respectively.

Similarly to construct $1-\alpha$ confidence ellipsoids on $\boldsymbol{\zeta}$, one takes

$$\left\{\boldsymbol{\zeta}_0 \;\middle|\; \left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right)^\top \left(\begin{bmatrix} \operatorname{Diag}\left(\frac{\hat{\boldsymbol{\sigma}}+\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\zeta}}}{\hat{\boldsymbol{\sigma}}^2}\right) \\ \operatorname{Diag}\left(\frac{-\hat{\boldsymbol{\zeta}}}{2\hat{\boldsymbol{\sigma}}^2}\right)\end{bmatrix}^\top \hat{\Omega} \begin{bmatrix} \operatorname{Diag}\left(\frac{\hat{\boldsymbol{\sigma}}+\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\zeta}}}{\hat{\boldsymbol{\sigma}}^2}\right) \\ \operatorname{Diag}\left(\frac{-\hat{\boldsymbol{\zeta}}}{2\hat{\boldsymbol{\sigma}}^2}\right)\end{bmatrix}\right)^{-1} \left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right) \le \frac{{\chi^2}_{1-\alpha}(p)}{n}\right\}.$$

Example 4.2.2 (Asymptotic Sharpe ratio of SMB and HML). Consider the returns of two of the Fama French factors, SMB and HML, introduced in Example 1.2.1. Based on monthly returns from Jan 1927 to Dec 2020, we compute

$$\left[\boldsymbol{\mu}^\top, {\boldsymbol{\alpha}_2}^\top\right]^\top = \begin{bmatrix} 0.21 \\ 0.32 \\ 10.19 \\ 12.40 \end{bmatrix}, \qquad \frac{1}{n}\hat{\Omega} = \begin{bmatrix} 0.01 & 0.00 & 0.06 & 0.03 \\ 0.00 & 0.01 & 0.03 & 0.09 \\ 0.06 & 0.03 & 1.95 & 0.82 \\ 0.03 & 0.09 & 0.82 & 2.87 \end{bmatrix}.$$

Here the units of $\boldsymbol{\mu}$ are in %mo.$^{-1}$, and those of $\boldsymbol{\alpha}_2$ are %^{2}mo.$^{-1}$. We plug in the sample estimates to get the estimate of the derivative:

$$\frac{\mathrm{d}\hat{\boldsymbol{\zeta}}}{\mathrm{d}\left[\hat{\boldsymbol{\mu}}^\top, {\hat{\boldsymbol{\alpha}}_2}^\top\right]^\top} \approx \begin{bmatrix} 0.315 & 0.000 & -0.003 & 0.000 \\ 0.000 & 0.288 & 0.000 & -0.004 \end{bmatrix}.$$

The Sharpe ratios are computed as

$$\begin{bmatrix} SMB & HML \\ \hline 0.065 & 0.092 \end{bmatrix} \mathrm{mo.}^{-1/2}.$$

The estimated standard-error variance-covariance of the Sharpe ratio of the two returns is

$$\begin{bmatrix} & SMB & HML \\ \hline SMB & 0.00080 & 0.00006 \\ HML & 0.00006 & 0.00076 \end{bmatrix}.$$

Thus, for example, taking the SMB marginal, we suppose the observed Sharpe ratio is nearly normally distributed around the true value with standard deviation approximately 0.0283.

To test the null hypothesis that $+1SMB - 1HML = 0.01$ versus the alternative $+1SMB - 1HML \le 0.01$, we estimate the test statistic as $z = -0.9648$, which is just a bit bigger than the critical value of $z_{0.05} = -1.6449$, and we fail to reject the null hypothesis.

To test the null hypothesis that both Sharpe ratios are equal to zero, we compute the test statistic $c^2 = 15.43$ which is bigger than ${\chi^2}_{0.95}(2) = 5.99$, and we reject the null hypothesis. To instead test the null hypothesis that both Sharpe ratios are equal to $0.1\mathrm{mo.}^{-1/2}$, we compute $c^2 = 1.55$, and fail to reject at the 0.05 level.

One is tempted to plot confidence ellipsoids around for the signal-noise ratios. However, in this case the estimated standard errors of the two factors are nearly the same, and their estimated correlation is low. Thus the confidence ellipsoids strongly resemble a circle centered at $(0.0653, 0.0919)$. We leave it as an exercise for the reader to draw such a circle. ⊣

Example 4.2.3 (Correlation of errors, Sharpe ratio of Fama-French factors). The Sharpe ratio of all four Fama French factors monthly returns from Example 1.2.1 were computed, using monthly returns from Jan 1927 to Dec 2020. The estimated standard error variance-covariance matrix of the computed Sharpe ratios, converted to a correlation matrix is computed as:

$$\left[\begin{array}{cccc} Mkt & SMB & HML & UMD \\ \hline 1.0000 & 0.3054 & 0.1510 & -0.2429 \\ 0.3054 & 1.0000 & 0.0756 & -0.1639 \\ 0.1510 & 0.0756 & 1.0000 & -0.3506 \\ -0.2429 & -0.1639 & -0.3506 & 1.0000 \end{array}\right].$$

cf. Exercise 4.13. ⊣

We defer a Bayesian analysis to the more general case of functions of signal-noise ratios, given in Section 4.3.2.

4.2.2. Scalar case

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

$$\begin{aligned} \Omega &= \left[\begin{array}{cc} \alpha_2 - \mu^2 & \alpha_3 - \mu\alpha_2 \\ \alpha_3 - \mu\alpha_2 & \alpha_4 - {\alpha_2}^2 \end{array}\right], \\ &= \left[\begin{array}{cc} \sigma^2 & \mu_3 + 2\mu\sigma^2 \\ \mu_3 + 2\mu\sigma^2 & \mu_4 + 4\mu_3\mu + 4\sigma^2\mu^2 - \sigma^4 \end{array}\right], \\ &= \left[\begin{array}{cc} \sigma^2 & \sigma^2\left(\sigma\gamma_1 + 2\mu\right) \\ \sigma^2\left(\sigma\gamma_1 + 2\mu\right) & \sigma^4\left(\frac{\mu_4}{\sigma^4} - 3 + 2\right) + 4\sigma^3\mu\frac{\mu_3}{\sigma^3} + 4\sigma^2\mu^2 \end{array}\right], \\ &= \sigma^2\left[\begin{array}{cc} 1 & \sigma\gamma_1 + 2\mu \\ \sigma\gamma_1 + 2\mu & \sigma^2\left(\gamma_2 + 2\right) + 4\sigma\mu\gamma_1 + 4\mu^2 \end{array}\right]. \end{aligned} \tag{4.32}$$

where α_i is the uncentered i^{th} moment of x, μ_i is the *centered* i^{th} moment of x, γ_1 is the skew, and γ_2 is the excess kurtosis of x. In this case the derivative of the scalar signal-noise ratio with respect to the first and second moments is

$$\frac{\mathrm{d}\zeta}{\mathrm{d}[\mu, \alpha_2]^{\top}} = \left[\begin{array}{cc} \frac{\sigma + \mu\zeta}{\sigma^2} & -\frac{\zeta}{2\sigma^2} \end{array}\right]. \tag{4.33}$$

After much algebraic simplification (Exercise 4.26), the asymptotic variance of Sharpe ratio is given by Mertens' formula, Equation 3.28:

$$\hat{\zeta} \approx \mathcal{N}\left(\zeta, \frac{1}{n}\left(1 - \zeta\gamma_1 + \frac{\gamma_2 + 2}{4}\zeta^2\right)\right). \tag{4.34}$$

Note that Mertens' equation applies even though our definition of Sharpe ratio includes a risk-free rate, r_0. It should be stressed that the signal-noise ratio, skew and excess kurtosis appearing in Equation 4.34 are *population* values, which are unknown in practice. Typically the standard error is estimated using the plug-in method where sample estimates of these quantities are used in their place. Estimation error in those sample estimates has an unknown affect on the standard error computation, but it is likely to be very small. (See Section 4.6.) Note also that the skew and excess kurtosis are zero for the normal distribution, in which case Mertens' formula reduces to the 'usual' standard error estimate given by Johnson and Welch, Equation 3.24.

Example 4.2.4 (Mertens versus vanilla standard error). Consider the returns of the Market, introduced in Example 1.2.1. Based on monthly returns from Jan 1927 to Dec 2020, the Sharpe ratio is computed as $0.61\,\mathrm{yr}^{-1/2}$. The standard error under the classical vanilla approximation, Equation 3.24, is estimated to be $0.104\,\mathrm{yr}^{-1/2}$, while under Equation 3.28, using the sample skew and excess kurtosis, it is estimated as $0.106\,\mathrm{yr}^{-1/2}$. There is no appreciable difference in these standard error estimates.

⊣

4.2.3. Asymptotic Bias and Variance of the Sharpe ratio

Equation 4.34 gives the *asymptotic* distribution of the scalar Sharpe ratio. In particular, it claims that the Sharpe ratio is asymptotically unbiased. For small n, however, this approximation may be too coarse. One can derive approximations of the bias of the Sharpe ratio involving the (typically unknown) higher order moments of returns.

While the sample mean is unbiased, the presence of the inverse square root of the sample variance in the Sharpe ratio introduces some bias. Consider the Taylor expansion of $x^{-1/2}$:

$$\frac{1}{\sqrt{x+\epsilon}} \approx \frac{1}{\sqrt{x}} - \frac{1}{2}\frac{1}{x^{3/2}}\epsilon + \frac{3}{8}\frac{1}{x^{5/2}}\epsilon^2 + \ldots$$

Now let $\hat{\sigma}^2 = \sigma^2\left(1+z\right)$ for some random variable z. Then

$$\frac{1}{\sqrt{\hat{\sigma}^2}} = \frac{1}{\sigma\sqrt{1+z}} \approx \frac{1}{\sigma}\left(1 - \frac{1}{2}z + \frac{3}{8}z^2 + \ldots\right).$$

Letting $\hat{\alpha}_i$ be the sample i^{th} moment, *e.g.*, $\hat{\alpha}_2 = \frac{1}{n}\sum_{1 \le i \le n} {x_i}^2$, we can express z as

$$z = \frac{n\left(\hat{\alpha}_2 - \hat{\mu}^2\right)}{\sigma^2\left(n-1\right)} - 1.$$

Then

$$\hat{\zeta} = \frac{\hat{\mu}}{\hat{\sigma}} \approx \frac{\hat{\mu}}{\sigma}\left(1 - \frac{1}{2}\frac{n\left(\hat{\alpha}_2 - \hat{\mu}^2\right)}{\sigma^2\left(n-1\right)} + \frac{1}{2} + \frac{3}{8}\left(\frac{n\left(\hat{\alpha}_2 - \hat{\mu}^2\right)}{\sigma^2\left(n-1\right)} - 1\right)^2 + \ldots\right)$$

Taking expectations, and omitting higher order terms for simplicity, we have

$$\mathrm{E}\left[\hat{\zeta}\right] = \zeta + \mathrm{E}\left[-\frac{1}{2}\frac{\hat{\mu}n\left(\hat{\alpha}_2 - \hat{\mu}^2\right)}{\sigma^3\left(n-1\right)} + \frac{\hat{\mu}}{2\sigma} + \ldots\right]$$

Using facts about expectations of products of raw sample moments (*cf.* Exercise 4.33), the bias of the Sharpe ratio can be expressed as

$$\begin{aligned}\mathrm{E}\left[\hat{\zeta}\right] - \zeta &= -\frac{1}{2\sigma^3}\frac{n}{n-1}\left(\frac{n-1}{n^2}\mu_3 + \frac{n-1}{n}\mu\sigma^2\right) + \frac{1}{2}\zeta + \ldots,\\ &= -\frac{1}{2}\left(\frac{1}{n}\frac{\mu_3}{\sigma^3} + \zeta\right) + \frac{1}{2}\zeta + \ldots,\\ &= -\frac{1}{2}\frac{\gamma_1}{n} + \ldots,\end{aligned} \tag{4.35}$$

where $\gamma_1 = \mu_3/\sigma^3$ is the skew.

Higher order formulæ, due to Bao, can be found by taking more terms in the Taylor expansion. [12] These are

$$\mathrm{E}\left[\hat{\zeta}\right] - \zeta \approx -\frac{\gamma_1}{2n} + \frac{3\zeta}{8n}\left(2+\gamma_2\right), \tag{4.36}$$

$$\begin{aligned}\mathrm{E}\left[\hat{\zeta}\right] - \zeta \approx &-\frac{\gamma_1}{2n} + \frac{3\zeta}{8n}\left(2+\gamma_2\right) + \frac{3}{8n^2}\left(\gamma_3 - \gamma_1 - \frac{5}{2}\gamma_1\gamma_2\right)\\ &+ \frac{\zeta}{32n^2}\left(49 - 10\gamma_4 - 15\gamma_2 - 40{\gamma_1}^2 + \frac{105{\gamma_2}^2}{4}\right),\end{aligned} \tag{4.37}$$

where γ_i are the standardized higher order cumulants. In particular, γ_1 is the skew and γ_2 is the excess kurtosis of returns. The formulægiven here involve the population cumulants, which typically are not be known. Rather when the bias is estimated, sample estimates of the mean, variance, and higher order cumulants are used instead.

Example 4.2.5 (Higher order bias, Market). Consider the returns of the Market, introduced in Example 1.2.1. Based on monthly returns from Jan 1927 to Dec 2020, the Sharpe ratio is computed as $0.61\,\mathrm{yr}^{-1/2}$. Using sample estimates for the cumulants, we estimate the bias of the Sharpe ratio to be $0.0019\,\mathrm{yr}^{-1/2}$ via Equation 4.36; and $0.0019\,\mathrm{yr}^{-1/2}$ via Equation 4.37, which are both very small. For example, the latter is around 0.307% of the computed Sharpe ratio, and around 1.81% of the vanilla standard error estimate.

If we restrict our attention to the period of 96 months from Jan 2013 to Dec 2020, then we compute a Sharpe ratio of $1.1\,\mathrm{yr}^{-1/2}$, and compute a higher order estimate of the bias of $0.0198\,\mathrm{yr}^{-1/2}$, which is around 1.8% of the computed Sharpe ratio, and around 5.45% of the standard error. ⊣

The same approach was used to find a higher order formula for the standard error,

$$\begin{aligned}\operatorname{Var}\left(\hat{\zeta}\right) = \frac{1}{n}\left(1 + \frac{2+\gamma_2}{4}\zeta^2 - \gamma_1\zeta\right) + \frac{\zeta^2}{32n^2}\left(76 + 12\gamma_2 - 12\gamma_4 - 48{\gamma_1}^2 + 39{\gamma_2}^2\right)\\ + \frac{5\zeta}{4n^2}\left(\gamma_3 - 3\gamma_2\gamma_1\right) + \frac{1}{4n^2}\left(8 + 7{\gamma_1}^2\right) + \ldots\end{aligned} \tag{4.38}$$

The n^{-1} term is Mertens' approximation of the variance. Again, the formula involves typically unknown population cumulants which have to be estimated in practice. It is not clear under which conditions using Equation 4.38 should be preferred to using Merten's standard error, as the higher order cumulants (γ_4 is the *sixth* standardized cumulant) must be estimated from the data. Moreover, since the additional terms are multiplied by n^{-2}, they should only make a difference when n is relatively small, which is exactly when estimating higher order cumulants would be difficult.

Example 4.2.6 (Bao, Mertens, vanilla standard errors). Continuing Example 4.2.4, we look at monthly returns of the Market factor from Jan 1927 to Dec 2020. As above, the standard error under Equation 3.24, is around $0.1039\,\mathrm{yr}^{-1/2}$, while by Merten's approximation, it is estimated as $0.1057\,\mathrm{yr}^{-1/2}$. Via the higher order formula, Equation 4.38, it is estimated as $0.106\,\mathrm{yr}^{-1/2}$. Again, there is no appreciable difference in the standard error estimates.

If we restrict our attention to the period of 96 months from Jan 2013 to Dec 2020, then the vanilla standard error is estimated as $0.362\,\mathrm{yr}^{-1/2}$; Merten's formula gives $0.391\,\mathrm{yr}^{-1/2}$; Bao's formula gives $0.398\,\mathrm{yr}^{-1/2}$.

$\dashv$

Confidence and Prediction Intervals Via any of the variance formulations, Bao, Mertens or vanilla (respectively, Equation 4.38, Equation 3.28, and Equation 3.24), one can estimate the standard error of the Sharpe ratio. These require one to plug in estimates of the signal-noise ratio, and perhaps higher order moments. As outlined in Section 3.5.9, a confidence interval can be inflated to a prediction interval via a simple formula, namely the factor given in Equation 3.40.

4.3. Asymptotic distribution of functions of multiple Sharpe ratios

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

$$\sqrt{n}\left(\boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right) - \boldsymbol{g}\left(\boldsymbol{\zeta}\right)\right) \rightsquigarrow \mathcal{N}\left(0, \left(\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}}\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, {\boldsymbol{\alpha}_2}^\top]^\top}\right)\Omega\left(\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}}\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, {\boldsymbol{\alpha}_2}^\top]^\top}\right)^\top\right) \tag{4.39}$$

For example, if one wanted to test the hypothesis that the signal-noise ratios of the p assets are equal, one would let $\boldsymbol{g}\left(\cdot\right)$ be the function which constructs the $p-1$

differences:

$$\boldsymbol{g}\left(\boldsymbol{\zeta}\right) = \left[\zeta_1 - \zeta_2, \ldots, \zeta_{p-1} - \zeta_p\right]^{\top}. \tag{4.40}$$

We will, however, consider the case of general $\boldsymbol{g}\left(\cdot\right)$.

4.3.1. Frequentist analysis

We assume $\boldsymbol{g}\left(\cdot\right)$ is defined such that it equals zero under the null. Asymptotically, under the null hypothesis that $\boldsymbol{g}\left(\boldsymbol{\zeta}\right) = 0$,

$$n\boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right)^{\top}\left(\left(\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}}\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^{\top}, \boldsymbol{\alpha}_2{}^{\top}]^{\top}}\right)\Omega\left(\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}}\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^{\top}, \boldsymbol{\alpha}_2{}^{\top}]^{\top}}\right)^{\top}\right)^{-1}\boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right) \sim \chi^2\left(\nu\right), \tag{4.41}$$

where ν is the rank of the matrix $\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}}$ at $\boldsymbol{\zeta}$.

To test the null hypothesis:

$$H_0 : \boldsymbol{g}\left(\boldsymbol{\zeta}\right) = \boldsymbol{0} \quad \text{versus} \quad H_1 : \boldsymbol{g}\left(\boldsymbol{\zeta}\right) \neq \boldsymbol{0},$$

first compute the test statistic

$$x^2 = n\boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right)^{\top}\left(\left(\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}}\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^{\top}, \boldsymbol{\alpha}_2{}^{\top}]^{\top}}\right)\hat{\Omega}\left(\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}}\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^{\top}, \boldsymbol{\alpha}_2{}^{\top}]^{\top}}\right)^{\top}\right)^{-1}\boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right). \tag{4.42}$$

Then reject if $x^2 \geq \chi^2{}_{1-\alpha}\left(\nu\right)$ the $1-\alpha$ quantile of the chi-square distribution with degrees of freedom ν equal to the rank of the matrix $\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}}$ at $\boldsymbol{\zeta}$.

This test was proposed by Wright *et al.* [174] The same test statistic was proposed by Leung and Wong for the $\boldsymbol{g}\left(\cdot\right)$ given in Equation 4.40. [96] However, Leung and Wong propose that one reject if

$$\frac{\left(n-p+1\right)}{\left(n-1\right)\left(p-1\right)}x^2 \geq f_{1-\alpha}\left(p-1, n-1\right), \tag{4.43}$$

where $f_{1-\alpha}\left(p-1, n-1\right)$ is the $1-\alpha$ quantile of the F distribution with $p-1$ and $n-1$ degrees of freedom. Write *et al.* find that the chi-square test gives closer to nominal coverage under the null than the F-test when returns are fat-tailed, even for large n. See also Exercise 4.15.

Again we compute the estimated covariance $\hat{\Omega}$ as described in Section 4.2, either by assuming elliptical returns and using Equation 4.27, or by computing the sample covariance of the vector of returns 'stacked' with elementwise squared returns, $\left[\boldsymbol{x}^{\top}, \boldsymbol{x}^{2\top}\right]^{\top}$. Ledoit and Wolf propose computing $\hat{\Omega}$ using HAC estimators or bootstrapping on the sample of stacked vectors $\left[\boldsymbol{x}^{\top}, \boldsymbol{x}^{2\top}\right]^{\top}$. [89]

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

$$\sqrt{n}g\left(\hat{\boldsymbol{\zeta}}\right)\left(\left(\frac{\mathrm{d}g}{\mathrm{d}\boldsymbol{\zeta}}\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^{\top},\boldsymbol{\alpha}_2{}^{\top}]^{\top}}\right)\Omega\left(\frac{\mathrm{d}g}{\mathrm{d}\boldsymbol{\zeta}}\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^{\top},\boldsymbol{\alpha}_2{}^{\top}]^{\top}}\right)^{\top}\right)^{-\frac{1}{2}} \sim t\left(n-1\right). \tag{4.44}$$

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 $\boldsymbol{g}\left(\boldsymbol{\zeta}\right)$.

Following the rejection of the null of equal signal-noise ratios, one may be interested in testing which of the assets have different signal-noise ratios. This is via a *post hoc* test, *cf.* Section 5.2.1.

Example 4.3.1 (Equality of signal-noise ratios Fama-French factors). Consider the four Fama French factors from Example 1.2.1. Based on monthly returns from Jan 1927 to Dec 2020, the Sharpe ratios were computed as

$$\hat{\boldsymbol{\zeta}} = \begin{bmatrix} Mkt & SMB & HML & UMD \\ \hline 0.1773 & 0.0653 & 0.0919 & 0.1377 \end{bmatrix} \text{mo.}^{-1/2},$$

and thus the *differences* in Sharpe ratios were computed as

$$\boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right) = \begin{bmatrix} 0.1120 & -0.0266 & -0.0458 \end{bmatrix} \text{mo.}^{-1/2}.$$

The covariance matrix of the error was estimated as

$$\begin{aligned}\hat{\Sigma}_{1/2} &=_{\mathrm{df}} \left.\left(\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}}\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^{\top},\boldsymbol{\alpha}_2{}^{\top}]^{\top}}\right)\hat{\Omega}\left(\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}}\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^{\top},\boldsymbol{\alpha}_2{}^{\top}]^{\top}}\right)^{\top}\right|_{\boldsymbol{\mu}=\hat{\boldsymbol{\mu}},\boldsymbol{\alpha}_2=\hat{\boldsymbol{\alpha}}_2}, \\ &\approx \begin{bmatrix} 1.3570 & -0.6802 & 0.1928 \\ -0.6802 & 1.6200 & -0.9948 \\ 0.1928 & -0.9948 & 3.2058 \end{bmatrix} \text{mo.}^{-1}.\end{aligned}$$

The test statistic is then computed as 11.9. Under the null hypothesis of equal signal-noise ratios, this is asymptotically distributed as a $\chi^2\left(3\right)$. This corresponds to a p-value of 0.0076, and a Frequentist might narrowly reject the null hypothesis of equal signal-noise ratios.

Using the F-test, the test statistic is essentially divided by 3, resulting in a putative $F\left(3, 1125\right)$ statistic under the null. The test statistic is computed as 3.97, with a p-value of 0.0079, little substantive difference.

Using the HAC estimator of covariance, the chi-squared test statistic is computed as 9.02, corresponding to a p-value under the null of 0.029. A Frequentist would then narrowly *fail* to reject the null of equal signal-noise ratios. See also Example 5.2.2.

$\dashv$

Example 4.3.2 (Equality of signal-noise ratios, HML and SMB). Continuing Example 4.2.2, consider the returns of the Fama French factors, SMB and HML. Based on monthly returns from Jan 1927 to Dec 2020, we compute the difference in Sharpe ratios of SMB and HML as $-0.027\text{mo.}^{-1/2}$. Under the null hypothesis of equal signal-noise ratios, the asymptotically t statistic of Equation 4.44 is computed as -0.7, which is distributed as a $t\left(1127\right)$ under the null, corresponding to a p-value of 0.76. There is little evidence to suggest either of SMB or HML has a higher signal-noise ratio than the other. $\dashv$

Example 4.3.3 (Equality of two signal-noise ratios, Normally distributed returns). Consider the case where of two assets where $\boldsymbol{x}$ takes a bivariate normal distribution. Let ρ be the correlation between the two returns. Suppose the signal-noise ratios of the two assets are, respectively, $\zeta\left(1+\epsilon\right)$ and ζ. Suppose we are testing for equality of signal-noise ratio, so let $\boldsymbol{g}$ be the difference function. As noted in Example 4.2.1, $\hat{\boldsymbol{\zeta}} \approx \mathcal{N}\left(\boldsymbol{\zeta}, \frac{1}{n}\left(\mathsf{R} + \frac{1}{2}\operatorname{Diag}\left(\boldsymbol{\zeta}\right)\left(\mathsf{R}\odot\mathsf{R}\right)\operatorname{Diag}\left(\boldsymbol{\zeta}\right)\right)\right)$. Then Equation 4.39 becomes

$$\left[\hat{\zeta}_1 - \hat{\zeta}_2\right] \rightsquigarrow \mathcal{N}\left(\epsilon\zeta, \frac{2}{n}\left(1-\rho\right) + \frac{\zeta^2}{2n}\left(1 + \left(1+\epsilon\right)^2 - 2\rho^2\left(1+\epsilon\right)\right)\right). \tag{4.45}$$

So for high correlation, the differences in Sharpe ratios will be very small. See also Exercise 4.28.

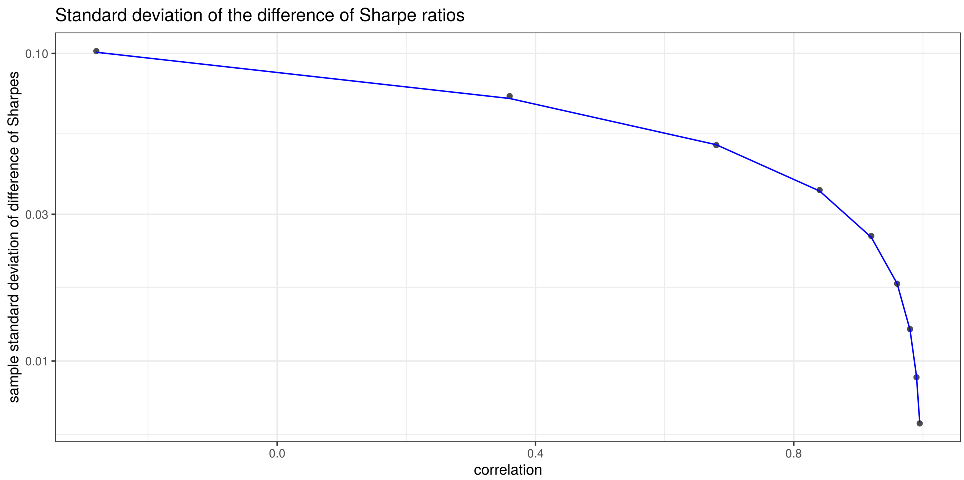

Figure 4.4.: The empirical standard deviation for the difference of Sharpe ratios of two correlated normally distributed returns is shown for different values of the correlation, ρ. Each point represents 10,000 simulations of 252 days of normally distributed returns. The signal-noise ratios of the two strategies are $\zeta = 1\text{yr}^{-1/2}$ and $\zeta\left(1+\epsilon\right)$ with $\epsilon = 0.02$, assuming 252 days per year. The plotted line is the theoretical value from Equation 4.45, *not* a fit of the data.

To confirm the findings of Equation 4.45, for different values of ρ, 10000 simulations of 252 days of correlated returns were generated, with $\zeta = 1\mathrm{yr}^{-1/2}$ and $\epsilon = 0.02$, assuming 252 days per year. The standard deviation of the differences in Sharpe ratios was computed for each value of ρ, and plotted versus ρ in Figure 4.4. The relationship of Equation 4.45 is given as a line, and fits the experimental data rather well.

$\dashv$

Caution. The test for equality of signal-noise ratios, as described by Leung and Wong, and Wright *et al.* is designed to deal with correlation of returns. However, as demonstrated in Example 4.3.3 it can exhibit very high power when the returns series are highly positively correlated, deeming very small differences in Sharpe ratios as 'significant.' While a statistical test with high power sounds like a practitioner's dream, it is likely to be completely confounded by omitted variable bias. Unless you are sure that your returns are not affected by an omitted variable (and are you really?), take great care in interpreting the results of this test when returns are highly correlated.

4.3.2. Bayesian analysis

The approximation of Equation 4.39 can be used in a Bayesian framework as well. [56, 114] To stay within the usual Gaussian framework, consider the asymptotic distribution of $\boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right)$ as being approximately multivariate normal. That is, if one observes the sample statistic $\hat{\boldsymbol{\zeta}}$ (either for the in-sample observations, or some future observations, say), based on a sample of size n, then simply take the following as true, even though it is only an approximation:

$$\boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right) \sim \mathcal{N}\left(\boldsymbol{g}\left(\boldsymbol{\zeta}\right), \frac{1}{n}\mathsf{B}\right). \tag{4.46}$$

The parameter $\boldsymbol{g}\left(\boldsymbol{\zeta}\right)$ is of primary interest, and B is a nuisance. Let us suppose there are some prior beliefs on $\boldsymbol{g}\left(\boldsymbol{\zeta}\right)$ and B. We start with a Normal-Inverse Wishart model,

$$\begin{aligned}\mathsf{B} &\propto \mathcal{IW}\left(\mathsf{B}_0, m_0\right),\\ \boldsymbol{g}\left(\boldsymbol{\zeta}\right)|\mathsf{B} &\propto \mathcal{N}\left(\boldsymbol{a}_0, \frac{1}{l_0}\mathsf{B}\right).\end{aligned} \tag{4.47}$$

By '$\mathcal{IW}\left(\mathsf{B}_0, m_0\right)$', we mean an inverse Wishart distribution with matrix parameter B_0 and degrees of freedom m_0. Thus $\mathsf{B} \propto \mathcal{IW}\left(\mathsf{B}_0, m_0\right)$ means that the prior probability on B is proportional to

$$p\left(\mathsf{B}\right) \propto \left|\mathsf{B}\right|^{-(m_0+p+1)/2} \exp\left(-\frac{1}{2}\operatorname{tr}\left(\mathsf{B}_0\mathsf{B}^{-1}\right)\right), \tag{4.48}$$

where p is the number of elements in the vector $\boldsymbol{g}\left(\boldsymbol{\zeta}\right)$. Another way of stating this is that $\mathsf{X} \propto \mathcal{IW}\left(\Psi, m\right)$ means that X^{-1} takes a Wishart distribution with scale matrix Ψ^{-1} and m degrees of freedom. (See Section 1.3.4.)

Via Equation 1.14, the prior marginal distribution of $\boldsymbol{g}(\boldsymbol{\zeta})$ is that of a *multivariate t*, written as

$$\boldsymbol{g}(\boldsymbol{\zeta}) \sim \mathcal{T}\left(m_0 - d + 1, \frac{\mathsf{B}_0}{l_0(m_0 - d + 1)}, \boldsymbol{a}_0\right).$$

(See Section 1.3.5.)

The normal prior on $\boldsymbol{g}(\boldsymbol{\zeta})$ means that

$$p(\boldsymbol{g}(\boldsymbol{\zeta}))\,|\mathsf{B} \propto \left|\frac{\mathsf{B}}{l_0}\right|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{g}(\boldsymbol{\zeta}) - \boldsymbol{a}_0)^\top (\mathsf{B}/l_0)^{-1} (\boldsymbol{g}(\boldsymbol{\zeta}) - \boldsymbol{a}_0)\right). \tag{4.49}$$

Having observed n observations, and computing $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\alpha}}_2$, and thus $\hat{\boldsymbol{\zeta}}$, as well as estimating the overall covariance $\hat{\Omega}$, consider the approximation of Equation 4.39 as being exact, and furthermore suppose that the estimated covariance from that equation takes a Wishart distribution with n degrees of freedom. Letting

$$\hat{\Sigma}_{1/2} =_{\mathrm{df}} \left(\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}} \frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, \boldsymbol{\alpha}_2{}^\top]^\top}\right) \hat{\Omega} \left(\frac{\mathrm{d}\boldsymbol{g}}{\mathrm{d}\boldsymbol{\zeta}} \frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, \boldsymbol{\alpha}_2{}^\top]^\top}\right)^\top \Bigg|_{\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}, \boldsymbol{\alpha}_2=\hat{\boldsymbol{\alpha}}_2}, \tag{4.50}$$

we are assuming that $n\hat{\Sigma}_{1/2}$ is Wishart with parameter B and n degrees of freedom. Thus the joint likelihood of $\boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right)$ and $\hat{\Sigma}_{1/2}$ is

$$\begin{aligned} p\left(\boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right), \hat{\Sigma}_{1/2}\right) \propto \frac{\left|\hat{\Sigma}_{1/2}\right|^{\frac{n-p-2}{2}}}{|\mathsf{B}|^{n/2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left(n\hat{\Sigma}_{1/2}\mathsf{B}^{-1}\right)\right) \times \\ \exp\left(-\frac{1}{2}\left(\boldsymbol{g}(\boldsymbol{\zeta}) - \boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right)\right)^\top \left(\frac{\mathsf{B}}{n}\right)^{-1} \left(\boldsymbol{g}(\boldsymbol{\zeta}) - \boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right)\right)\right). \end{aligned} \tag{4.51}$$

The posterior distribution is then (*cf.* Exercise 4.30)

$$\begin{aligned} \mathsf{B} &\propto \mathcal{IW}(\mathsf{B}_1, m_1), \\ \boldsymbol{g}(\boldsymbol{\zeta})\,|\mathsf{B} &\propto \mathcal{N}\left(\boldsymbol{a}_1, \frac{1}{l_1}\mathsf{B}\right), \end{aligned} \tag{4.52}$$

where

$$\begin{aligned} l_1 &= l_0 + n, \\ \boldsymbol{a}_1 &= \frac{l_0\boldsymbol{a}_0 + n\boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right)}{l_1}, \\ m_1 &= m_0 + n, \\ \mathsf{B}_1 &= \mathsf{B}_0 + n\hat{\Sigma}_{1/2} + \frac{l_0 n}{l_1}\left(\boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right) - \boldsymbol{a}_0\right)\left(\boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right) - \boldsymbol{a}_0\right)^\top. \end{aligned} \tag{4.53}$$

The posterior marginal distribution of $\boldsymbol{g}(\boldsymbol{\zeta})$ is that of a multivariate t, written as

$$\boldsymbol{g}(\boldsymbol{\zeta}) \sim \mathcal{T}\left(m_1 - d + 1, \frac{\mathsf{B}_1}{l_1 (m_1 - d + 1)}, \boldsymbol{a}_1\right).$$

Example 4.3.4 (Difference of signal-noise ratios Fama-French factors, Bayesian analysis). We continue the analysis of Example 4.3.1, but as a Bayesian. We analyze the difference of signal-noise ratios, Market minus SMB, SMB minus HML, and HML minus UMD respectively. We take the Bayesian prior to be

$$\begin{aligned}
l_0 &= 60\text{mo.},\\
\boldsymbol{a}_0 &= [0, 0, 0]^\top \text{mo.}^{-1/2},\\
m_0 &= 20,\\
\mathsf{B}_0 &= \begin{bmatrix} 20.00 & 0.00 & 0.00 \\ 0.00 & 20.00 & 0.00 \\ 0.00 & 0.00 & 20.00 \end{bmatrix}.
\end{aligned} \tag{4.54}$$

Having observed the Fama French monthly returns data from from Jan 1927 to Dec 2020, we estimate, as in Example 4.3.1,

$$\begin{aligned}
\boldsymbol{g}\left(\hat{\boldsymbol{\zeta}}\right) &= [0.112, -0.0266, -0.0458]^\top \text{mo.}^{-1/2},\\
\hat{\mathsf{\Sigma}}_{1/2} &= \begin{bmatrix} 1.3570 & -0.6802 & 0.1928 \\ -0.6802 & 1.6200 & -0.9948 \\ 0.1928 & -0.9948 & 3.2058 \end{bmatrix} \text{mo.}^{-1}.
\end{aligned}$$

Combining this with the prior above, we have the posterior parameters

$$\begin{aligned}
l_1 &= 1188\text{mo.},\\
\boldsymbol{a}_1 &= [0.1063, -0.0252, -0.0435]^\top \text{mo.}^{-1/2},\\
m_1 &= 1148,\\
\mathsf{B}_1 &= \begin{bmatrix} 1551.37 & -767.43 & 217.17 \\ -767.43 & 1847.43 & -1122.11 \\ 217.17 & -1122.11 & 3636.31 \end{bmatrix}.
\end{aligned} \tag{4.55}$$

For sufficiently large m_1, the belief in B is highly concentrated around the mean value of the inverse Wishart, in this case $\mathsf{B}_1 / (m_1 - 3 - 1)$. We compute this to be

$$\frac{1}{m_1 - 3 - 1}\mathsf{B}_1 = \begin{bmatrix} 1.3561 & -0.6708 & 0.1898 \\ -0.6708 & 1.6149 & -0.9809 \\ 0.1898 & -0.9809 & 3.1786 \end{bmatrix} \text{mo.}^{-1}.$$

As expected, this is not substantively different from $\hat{\mathsf{\Sigma}}_{1/2}$, as the number of pseudo-observations in the prior is small compared to the number of actual observations.

To explore the implications of this posterior, consider the Market minus SMB element. Our posterior belief is that the difference in signal-noise ratios of Market minus SMB is approximately normally distributed with mean $0.1063\text{mo.}^{-1/2}$ and standard deviation around $0.0338\text{mo.}^{-1/2}$. In Example 4.3.1, we estimated the difference to be $0.112\text{mo.}^{-1/2}$, with estimated standard error around $0.0347\text{mo.}^{-1/2}$. We draw 1000 draws from the posterior, finding the minimal value of the difference in signal-noise ratios drawn is $0.0183\text{mo.}^{-1/2}$.

⊣

4.4. † The ex-factor Sharpe ratio for (non-i.i.d.) Gaussian and Elliptical returns

We will now consider the sensitivity of the ex-factor Sharpe ratio of Equation 2.10 to *i.i.d.* assumptions, but still assuming Gaussian returns. This is rather more complicated than the vanilla Sharpe ratio. Here we abuse the 'augmented form' of vectors, which we will use later to analyze the distribution of portfolios. [121]

Recall the setup leading to Equation 2.10: we have a scalar return x_i, which is aligned with a l-vector, $\boldsymbol{f}_i$, which typically contains one element that is a constant 1. We will consider $\boldsymbol{f}$ to be random, which matches typical usage. Define

$$\tilde{\boldsymbol{x}}_i =_{\text{df}} \left[x_i, {\boldsymbol{f}_i}^\top\right]^\top. \tag{4.56}$$

Define the uncentered second moment of $\tilde{\boldsymbol{x}}$ as

$$\Theta =_{\text{df}} \operatorname{E}\left[\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^\top\right]. \tag{4.57}$$

Given a sample of size n of returns and corresponding factors, we can define $\tilde{\mathsf{X}}$ as the $n \times 1 + l$ matrix whose rows are $\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, \ldots, \tilde{\boldsymbol{x}}_n$. The vanilla sample estimate[7] of Θ is

$$\hat{\Theta} =_{\text{df}} \frac{1}{n}\sum_i \tilde{\boldsymbol{x}}_i{\tilde{\boldsymbol{x}}_i}^\top = \frac{1}{n}\tilde{\mathsf{X}}^\top\tilde{\mathsf{X}}. \tag{4.58}$$

With some work we can extract the ex-factor Sharpe ratio from some transformations of $\hat{\Theta}$. Note that since $\hat{\Theta}$ is a simple average, it is unbiased.

Now note that

$$\Theta = \left[\begin{array}{cc} \sigma^2 + \beta^\top\Gamma_f\beta & \beta^\top\Gamma_f \\ \Gamma_f\beta & \Gamma_f \end{array}\right], \tag{4.59}$$

where $\Gamma_f =_{\text{df}} \operatorname{E}\left[\boldsymbol{f}\boldsymbol{f}^\top\right]$. Simple matrix multiplication (Exercise 4.21) confirms that the inverse of Θ is

$$\Theta^{-1} = \left[\begin{array}{cc} \sigma^{-2} & -\beta^\top\sigma^{-2} \\ -\beta\sigma^{-2} & {\Gamma_f}^{-1} + \sigma^{-2}\beta\beta^\top \end{array}\right], \tag{4.60}$$

[7] *i.e.*, the one with n in the denominator, instead of, say $n - l$. Typical estimators of the regression coefficient may use a differrent denominator, an immaterial difference for sufficiently large n.

and the Cholesky factor of that inverse is

$$\mathsf{\Theta}^{-1/2} = \left[\begin{array}{cc} \sigma^{-1} & 0 \\ -\beta\sigma^{-1} & \mathsf{\Gamma}_f^{\;-1/2} \end{array}\right]. \tag{4.61}$$

The ex-factor signal-noise ratio (*cf.* Equation 2.10) can thus be expressed as

$$\zeta_g = \frac{\boldsymbol{\beta}^\top \boldsymbol{v} - r_0}{\sigma} = -\operatorname{tr}\left(\boldsymbol{e}_1 \left[r_0, \boldsymbol{v}^\top\right] \mathsf{\Theta}^{-1/2}\right). \tag{4.62}$$

The sample ex-factor Sharpe ratio takes the same form in the sample analogue:

$$-\operatorname{tr}\left(\boldsymbol{e}_1 \left[r_0, \boldsymbol{v}^\top\right] \hat{\mathsf{\Theta}}^{-1/2}\right) \to \hat{\zeta}_g.$$

For fixed r_0, $\boldsymbol{v}$, define the function $g\left(\cdot\right)$ by

$$g\left(\hat{\mathsf{\Theta}}\right) =_{\mathrm{df}} -\operatorname{tr}\left(\boldsymbol{e}_1 \left[r_0, \boldsymbol{v}^\top\right] \hat{\mathsf{\Theta}}^{-1/2}\right). \tag{4.63}$$

Example 4.4.1 (ex-factor Sharpe ratio on Fama-French factors). Consider the four Fama French factors from Example 1.2.1. We will model UMD as a linear combination of Market, SMB, HML and an intercept term. Given the 1128 months of data from Jan 1927 to Dec 2020, we compute

$$\hat{\mathsf{\Theta}} = \left[\begin{array}{ccccc} \mathit{UMD} & \mathit{Mkt} & \mathit{SMB} & \mathit{HML} & \mathit{intercept} \\ \hline 22.54 & -8.00 & -2.24 & -6.61 & 0.65 \\ -8.00 & 29.54 & 5.71 & 4.80 & 0.95 \\ -2.24 & 5.71 & 10.19 & 1.50 & 0.21 \\ -6.61 & 4.80 & 1.50 & 12.40 & 0.32 \\ 0.65 & 0.95 & 0.21 & 0.32 & 1.00 \end{array}\right] \mathrm{bps\,mo.}^{1}$$

The Cholesky factor of the inverse is

$$\hat{\mathsf{\Theta}}^{-1/2} = \left[\begin{array}{ccccc} \mathit{UMD} & \mathit{Mkt} & \mathit{SMB} & \mathit{HML} & \mathit{intercept} \\ \hline 0.24 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.05 & 0.20 & 0.00 & 0.00 & 0.00 \\ 0.01 & -0.10 & 0.32 & 0.00 & 0.00 \\ 0.11 & -0.06 & -0.04 & 0.29 & 0.00 \\ -0.25 & -0.15 & -0.05 & -0.09 & 1.00 \end{array}\right] \%^{-1}\,\mathrm{mo.}^{1/2}$$

Now consider the idiosyncratic return of UMD. That is, we isolate the intercept portion of the attribution by taking $\boldsymbol{v} = \left[0, 0, 0, 1\right]^\top\%$, and we assume $r_0 = 0\%$. Via Equation 4.62 we compute,

$$\zeta_g = 0.2467\mathrm{mo.}^{-1/2}$$

Via the sample estimate of Ω, and using the delta method, we estimate the standard error of this to be $0.0331\mathrm{mo.}^{-1/2}$. In Example 3.5.9 under the assumption that the factor returns were deterministic, we computed a slightly smaller estimate of the standard error, $0.0306\mathrm{mo.}^{-1/2}$.

⊣

Now we will consider the effect of the distribution of $\hat{\zeta}_g$ on assumptions of independence, homoskedasticity, *etc.*, but keeping normality of returns. To tame the computations, some simplifying assumptions are made. We assume that

$$\tilde{\mathsf{X}} \sim \mathcal{N}\left(\mathsf{M}, \mathsf{\Sigma}_f \otimes \mathsf{H}\right), \tag{4.64}$$

for some $n \times 1 + l$ matrix M, symmetric positive definite $n \times n$ matrix H and symmetric positive semidefinite[8] $(1+l) \times (1+l)$ matrix $\mathsf{\Sigma}_f$. Recall (Section **??**) that this means

$$\operatorname{vec}\left(\tilde{\mathsf{X}}\right) \sim \mathcal{N}\left(\operatorname{vec}\left(\mathsf{M}\right), \mathsf{\Sigma}_f \otimes \mathsf{H}\right).$$

For this characterization to be comparable to the vanilla case, we require $\mathsf{M}^{\top}\boldsymbol{1_n} = n\boldsymbol{\mu}$ and $\operatorname{tr}\left(\mathsf{H}\right) = n$. Note that the homoskedastic *i.i.d.* case corresponds to $\mathsf{M} = \boldsymbol{1_n}\boldsymbol{\mu}^{\top}$ and $\mathsf{H} = \mathsf{I}$.

From these we can compute the expected value and covariance of $\hat{\mathsf{\Theta}}$. This will require heavy use of Isserlis' theorem. [74] Tedious computation (Exercise 4.20) yields

$$\begin{aligned} \mathrm{E}\left[\hat{\mathsf{\Theta}}\right] &= \frac{1}{n}\left(\mathsf{M}^{\top}\mathsf{M} + \operatorname{tr}\left(\mathsf{H}\right)\mathsf{\Sigma}_f\right), \\ \operatorname{Var}\left(\operatorname{vec}\left(\hat{\mathsf{\Theta}}\right)\right) &= \frac{1}{n^2}\left(\mathsf{I} + \mathsf{K}\right)\left\{\mathsf{M}^{\top}\mathsf{H}\mathsf{M} \otimes \mathsf{\Sigma}_f + \mathsf{\Sigma}_f \otimes \mathsf{M}^{\top}\mathsf{H}\mathsf{M}\right. \\ &\qquad \left. + \operatorname{tr}\left(\mathsf{H}^2\right)\mathsf{\Sigma}_f \otimes \mathsf{\Sigma}_f\right\}. \end{aligned} \tag{4.65}$$

cf. Lemma 9 of Magnus and Neudecker. [105] We will again insist, perhaps pigheadedly, that the sample estimate only be unbiased in the absence of variation of the mean value. That is, we take $\mathsf{\Theta}$ to be the expected value of the expected second moment under uniform sampling of the $\tilde{\boldsymbol{x}}_i$, or

$$\mathsf{\Theta} = \mathrm{E}_i\left[\mathrm{E}\left[\tilde{\boldsymbol{x}}_i\tilde{\boldsymbol{x}}_i^{\top}\right]\right],$$

where we sample the $\tilde{\boldsymbol{x}}_i$ uniformly. This implies that

$$\mathsf{\Theta} = \mathsf{\Sigma}_f + \boldsymbol{\mu}\boldsymbol{\mu}^{\top}, \tag{4.66}$$

and thus $\hat{\mathsf{\Theta}}$ is only unbiased when $\mathsf{M} = \boldsymbol{1_n}\boldsymbol{\mu}^{\top}$, *cf.* Exercise 4.22.

Now we proceed as previously, using Taylor's theorem to claim the following approximations:

$$\begin{aligned} \mathrm{E}\left[\hat{\zeta}_g\right] &\approx g\left(\mathrm{E}\left[\hat{\mathsf{\Theta}}\right]\right), \\ \operatorname{Var}\left(\hat{\zeta}_g\right) &\approx \left(\left.\frac{\mathrm{d}g\left(\mathsf{\Theta}\right)}{\mathrm{d}\operatorname{vech}\left(\mathsf{\Theta}\right)}\right|_{\mathsf{\Theta}=\mathrm{E}\left[\hat{\mathsf{\Theta}}\right]}\right)\operatorname{Var}\left(\operatorname{vech}\left(\hat{\mathsf{\Theta}}\right)\right)\left(\left.\frac{\mathrm{d}g\left(\mathsf{\Theta}\right)}{\mathrm{d}\operatorname{vech}\left(\mathsf{\Theta}\right)}\right|_{\mathsf{\Theta}=\mathrm{E}\left[\hat{\mathsf{\Theta}}\right]}\right)^{\top}, \end{aligned} \tag{4.67}$$

[8] We can allow elements of $\boldsymbol{f}$ to be deterministic by taking corresponding rows and columns of $\mathsf{\Sigma}_f$ to be all zeros.

where, via Lemma 1.4.6, the derivative can be computed as

$$\frac{\mathrm{d}g\left(\Theta\right)}{\mathrm{d}\,\mathrm{vech}\left(\Theta\right)} = \left(\mathsf{K}\,\mathrm{vec}\left(\boldsymbol{e}_1\left[r_0, \boldsymbol{v}^\top\right]\right)\right)^\top \left(\mathsf{L}\left(\mathsf{I}+\mathsf{K}\right)\left(\left(\left(\Theta^{-1/2}\right)^\top\right)^{-1} \otimes \Theta\right)\mathsf{L}^\top\right)^{-1}, \tag{4.68}$$

where K is the commutation matrix of Definition 1.1.3.

Before proceeding, we check analytically whether this formulation reduces to Equation 4.13 for the trivial case.

Example 4.4.2 (Homoskedastic, independent returns and Sharpe ratio standard error). Consider the case where returns are homoskedastic and independent (and thus $\mathsf{M} = \mathbf{1}\boldsymbol{\mu}^\top$ and $\mathsf{H} = \mathsf{I}$), and where there is a single 'factor' identically equal to one. In this case

$$\Theta = \left[\begin{array}{cc} \sigma^2+\mu^2 & \mu \\ \mu & 1 \end{array}\right], \quad \text{and} \quad \Theta^{-1/2} = \left[\begin{array}{cc} \sigma^{-1} & 0 \\ -\zeta & 1 \end{array}\right].$$

The variance of $\hat{\Theta}$ from Equation 4.65 becomes

$$\mathrm{Var}\left(\mathrm{vech}\left(\hat{\Theta}\right)\right) = \frac{1}{n}\left[\begin{array}{ccc} \sigma^4\left(4\zeta^2+2\right) & 2\sigma^3\zeta & 0 \\ 2\sigma^3\zeta & \sigma^2 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

The derivative of Equation 4.68 becomes

$$\frac{\mathrm{d}g\left(\Theta\right)}{\mathrm{d}\,\mathrm{vech}\left(\Theta\right)} = \left[\begin{array}{ccc} -\frac{\zeta}{2\sigma^2} & \frac{1}{\sigma}\left(1+\zeta^2\right) & -\frac{\zeta^3}{2}-\zeta \end{array}\right].$$

The approximate variance of Equation 4.67 becomes

$$\mathrm{Var}\left(\hat{\zeta}_g\right) \approx \frac{1}{n}\left(1+\frac{\zeta^2}{2}\right),$$

i.e., the 'classical' approximate standard error of Equation 3.24. Note that this computation is unchanged in the case where $r_0 \neq 0$ as long as one defines $\zeta = \left(\mu - r_0\right)/\sigma$.

This computation is presented as a sympy notebook in Example B.0.1. ⊣

Independent Elliptical Returns We now derive similar results but instead of assuming normal returns, as in Equation 4.64, we assume returns follow an Elliptical distribution. As noted in Section 4.1, we should avoid assuming returns follow an elliptical distribution across time, as this can cause even uncorrelated returns to be dependent at arbitrary time separation, and can lead to statistics which do not converge to the population values.

However, it is attractive to be able to describe the vector $\tilde{\boldsymbol{x}}_i$ as following an Elliptical distribution (or distributions), with parameters varying over time. So assume that $\tilde{\boldsymbol{x}}_i$ follows an Elliptical distribution with kurtosis factor κ_i, mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$. Note that in the cases where $\tilde{\boldsymbol{x}}_i$ contains a deterministic 1, the corresponding row and column of Σ_i will be all zero. Define $\Theta_i = \boldsymbol{\mu}_i{\boldsymbol{\mu}_i}^{\top} + \Sigma_i$. Then Equation 4.65 is replaced by[121]

$$\begin{aligned} \mathrm{E}\left[\hat{\Theta}\right] &= \frac{1}{n}\sum_i \Theta_i, \\ \mathrm{Var}\left(\mathrm{vec}\left(\hat{\Theta}\right)\right) &= \frac{1}{n^2}\sum_i \left\{(\kappa_i - 1)\left[\mathrm{vec}\left(\Sigma_i\right)\mathrm{vec}\left(\Sigma_i\right)^{\top} + (\mathsf{I}+\mathsf{K})\,\Sigma_i\otimes\Sigma_i\right]\right. \\ &\qquad \left. + (\mathsf{I}+\mathsf{K})\left[\Theta_i\otimes\Theta_i - \boldsymbol{\mu}_i{\boldsymbol{\mu}_i}^{\top}\otimes\boldsymbol{\mu}_i{\boldsymbol{\mu}_i}^{\top}\right]\right\}. \end{aligned} \tag{4.69}$$

4.4.1. Market term, constant expectation

The most general form of the expectation and variance-covariance for the ex-factor Sharpe ratio presented above yields no simple closed form solution. As a simplifying assumption, in this section we assume that the factor returns consist of a constant one and the returns of a ‘Market’, which has mean and volatility of μ_m and σ_m respectively. Define $\zeta_m = \mu_m/\sigma_m$. Suppose that the asset has non-zero beta against the market, and define ζ_g as the idiosyncratic mean return divided by the residual volatility. We further require that M be constant over time, as otherwise the variance-covariance is too complicated.

$$\begin{aligned} \mathsf{M} &= \boldsymbol{1_n}\boldsymbol{\mu}^{\top}, \\ \boldsymbol{\mu}^{\top} &= \left[\mu + \beta\mu_m, 1, \mu_m\right], \\ \Sigma_f &= \begin{bmatrix} \beta^2\sigma_m^2 + \sigma^2 & 0 & \beta\sigma_m^2 \\ 0 & 0 & 0 \\ \beta\sigma_m^2 & 0 & \sigma_m^2 \end{bmatrix}, \\ \zeta_m &=_{\mathrm{df}} \frac{\mu_m}{\sigma_m}, \qquad \text{and} \qquad \zeta_g =_{\mathrm{df}} \frac{\mu - r_0}{\sigma}. \end{aligned}$$

With $\Theta = \Sigma_f + \boldsymbol{\mu}\boldsymbol{\mu}^{\top}$, we have

$$\Theta^{-1/2} = \begin{bmatrix} \frac{1}{\sigma} & 0 & 0 \\ -\zeta_g & \sqrt{\zeta_m^2 + 1} & 0 \\ -\frac{\beta}{\sigma} & -\frac{\zeta_m}{\sigma_m\sqrt{\zeta_m^2+1}} & \frac{1}{\sigma_m\sqrt{\zeta_m^2+1}} \end{bmatrix}.$$

Without variation in M over time, $\hat{\Theta}$ is unbiased for Θ. Via Equation 4.67, $\hat{\zeta}_g$ is approximately unbiased for ζ_g.

Via Equation 4.67, and a lot of algebra, the variance of $\hat{\zeta}_g$ is

$$\mathrm{Var}\left(\hat{\zeta}_g\right) \approx \frac{1}{n}\left[\frac{\mathrm{tr}\left(\mathsf{H}^2\right)}{n}\left(\frac{\zeta_g^2}{2} + \zeta_m^2\right) + \frac{\boldsymbol{1}^{\top}\mathsf{H}\boldsymbol{1}}{n}\right]. \tag{4.70}$$

4.4.2. Homoskedastic i.i.d. Elliptical returns

For the completely 'vanilla' case of homoskedastic, *i.i.d.* Elliptically distributed returns, plugging the covariance of Equation 4.69 into Equation 4.67 one arrives at the following equation:

$$\operatorname{Var}\left(\hat{\zeta}_g\right) \approx \frac{1}{n}\left[\frac{3\left(\kappa-1\right)}{4}\zeta_g^2 + \frac{\zeta_g^2}{2} + \kappa\zeta_m^2 + 1\right]. \tag{4.71}$$

Note that $3\left(\kappa-1\right)$ is the excess kurtosis of the marginal returns. One can derive this equation in the Gaussian case ($\kappa = 1$) by taking H in Equation 4.70 to be the identity matrix. See Example B.0.2.

This nicely generalizes the approximate standard error of Equation 3.24 in the Gaussian case, and Equation 3.28 for unskewed returns. Note that one expects that ζ_m^2 is very small or nearly zero[9], so the addition of the Market term and the regression against it does little to increase the variance in estimated Sharpe ratio.

Example 4.4.3 (Standard error of ex-factor Sharpe ratio with Market)*.* To verify Equation 4.71 in the Gaussian case, for selected values of ζ_m ranging from $0\text{yr}^{-1/2}$ to $5\text{yr}^{-1/2}$, 20,000 simulations were performed. In each simulation, 1260 days of daily returns of an asset with fixed ex-factor signal-noise ratio of $0.79\text{yr}^{-1/2}$ and 'beta' of 1 against the Market term were drawn; the ex-factor Sharpe ratio was computed, and the standard deviation over the 20000 simulations was computed. The experiment was repeated for returns drawn from an Elliptical distribution, a multivariate t distribution with 6 degrees of freedom, corresponding to $\kappa = 2$. The empirical standard errors are plotted versus ζ_m in Figure 4.5, along with the theoretical value from Equation 4.71.

See also Exercise 4.25.

$\dashv$

4.4.3. Homoskedastic autocorrelated Gaussian returns

The homoskedastic, simple autoregressive case of Section 4.1.4 corresponds to $\mathsf{H}_{i,j} = \rho^{|i-j|}$. Again, the ex-factor Sharpe ratio is approximately unbiased. The variance of $\hat{\zeta}_g$ becomes

$$\begin{aligned}\operatorname{Var}\left(\hat{\zeta}_g\right) &\approx \frac{1}{n}\left[\left(1 + 2\frac{n-1}{n}\rho^2 + \ldots\right)\left(\frac{\zeta_g^2}{2} + \zeta_m^2\right) + \left(1 + 2\frac{n-1}{n}\rho + \ldots\right)\right],\\ &\approx \frac{1}{n}\left[\left(\frac{\zeta_g^2}{2} + \zeta_m^2\right) + \frac{1+\rho}{1-\rho}\right],\end{aligned} \tag{4.72}$$

where we assume that ρ is small enough that ρ^n is negligible. The variance here is equivalent to Equation 4.22 for the case of the unattributed model under the small ρ case, and we see again that the Market term introduces a ζ_m^2/n term to the variance. Again, we see, as in the unattributed case, that a small autocorrelation of ρ inflates the standard error of the Sharpe ratio by about $200\rho\%$.

[9]If ζ_m^2 were large, presumably investors would merely invest in the Market instrument instead.

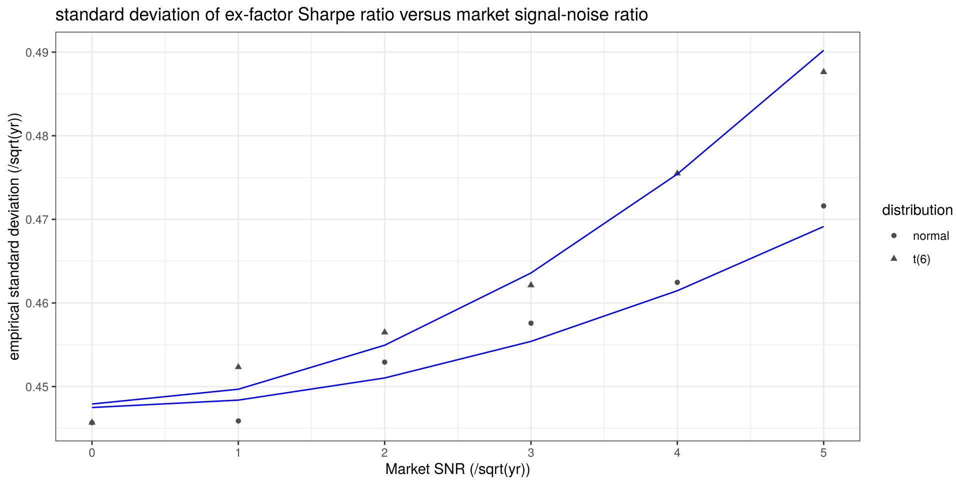

Figure 4.5.: The empirical standard error of the ex-factor Sharpe ratio is plotted versus the Market signal-noise ratio for the case of an asset with a beta of 1 against the Market, and a ex-factor signal-noise ratio of $0.79\text{yr}^{-1/2}$. Standard error estimated from 20,000 simulations. Returns are drawn from a normal distribution, and from a multivariate t distribution with 6 degrees of freedom. The theoretical values from Equation 4.71 are plotted as a line for each value of κ.

Example 4.4.4 (Standard error of ex-factor Sharpe ratio with Market, with autocorrelation)*.* To verify Equation 4.72, for selected values of ρ ranging from -0.15 to 0.15, 50000 simulations were performed. In each simulation, 1260 days of daily returns of an asset with fixed ex-factor signal-noise ratio of $0.79\text{yr}^{-1/2}$ and 'beta' of 1 against the Market term were drawn, where the Market had signal-noise ratio of $1\text{yr}^{-1/2}$. The ex-factor Sharpe ratio was computed, and the standard deviation was computed and plotted in Figure 4.6. We see good agreement of theoretical and empirical standard errors.

⊣

4.5. † Asymptotic Distribution of ex-factor Sharpe ratio

Continuing on the work from the previous section, here we consider the asymptotic distribution of the ex-factor Sharpe ratio, independent of assumptions on the type of distribution the returns take.

Note that since $\hat{\Theta}$ is a simple average, under mild conditions we can rely on the central limit theorem to claim that

$$\sqrt{n}\left(\operatorname{vech}\left(\hat{\Theta}\right) - \operatorname{vech}\left(\Theta\right)\right) \rightsquigarrow \mathcal{N}\left(0, \Omega\right), \tag{4.73}$$

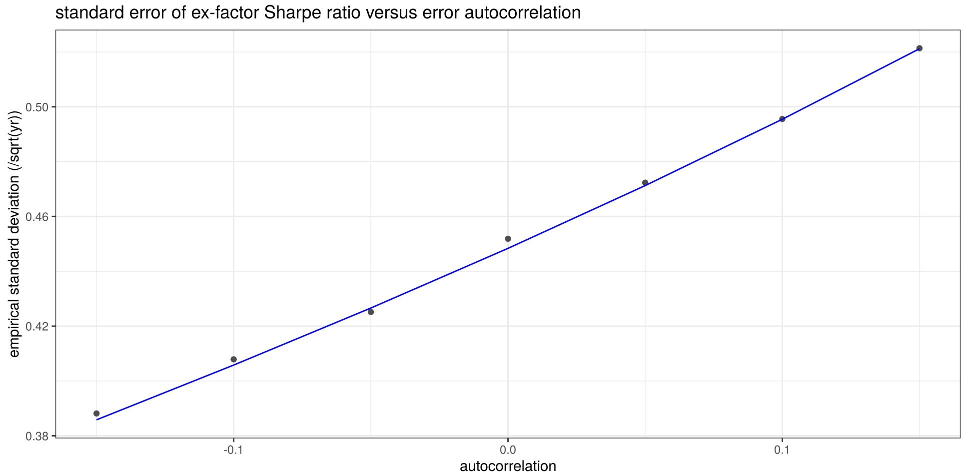

Figure 4.6.: The empirical standard error of the ex-factor Sharpe ratio is plotted versus ρ, the autocorrelation of errors. The asset has a beta of 1 against the Market, and a ex-factor signal-noise ratio of $0.79\text{yr}^{-1/2}$. The Market has a signal-noise ratio of $1\text{yr}^{-1/2}$. Empirical standard errors are estimated over 50,000 simulations. The theoretical value from Equation 4.72 is plotted as a line.

where

$$\Omega = \operatorname{Var}\left(\operatorname{vech}\left(\hat{\Theta}\right)\right).$$

In general this variance-covariance matrix, Ω, is unknown, but can be consistently estimated from the data. As described in Section 4.2, Ω can be estimated via the sample covariance of $\operatorname{vech}\left(\tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i{}^{\top}\right)$; or $\hat{\Omega}$ can be constructed using Equation 4.69 and sample estimates of $\boldsymbol{\mu}$, Σ and κ.

By the multivariate delta method, we now claim

$$\sqrt{n}\left(g\left(\hat{\Theta}\right) - g\left(\Theta\right)\right) \rightsquigarrow \mathcal{N}\left(0, \left(\left.\frac{\mathrm{d}g\left(\Theta\right)}{\mathrm{d}\operatorname{vech}\left(\Theta\right)}\right|_{\Theta=\mathrm{E}\left[\hat{\Theta}\right]}\right) \Omega \left(\left.\frac{\mathrm{d}g\left(\Theta\right)}{\mathrm{d}\operatorname{vech}\left(\Theta\right)}\right|_{\Theta=\mathrm{E}\left[\hat{\Theta}\right]}\right)^{\top}\right), \quad (4.74)$$

where the derivative is as given in Equation 4.68.

4.5.1. Testing the ex-factor Sharpe ratio

From the variance computation, we can compute a Wald statistic to perform hypothesis testing. To be concrete, for scalar-valued function $g(\cdot)$, to test the hypothesis

$$H_0 : g(\Theta) = g_0 \quad \text{versus} \quad H_1 : g(\Theta) \neq g_0,$$

one computes the Wald statistic

$$W = \frac{\sqrt{n}\left(g\left(\hat{\Theta}\right) - g_0\right)}{\sqrt{\left(\left.\frac{\mathrm{d}g(\Theta)}{\mathrm{d}\,\mathrm{vech}(\Theta)}\right|_{\Theta=\mathrm{E}[\hat{\Theta}]}\right)\hat{\Omega}\left(\left.\frac{\mathrm{d}g(\Theta)}{\mathrm{d}\,\mathrm{vech}(\Theta)}\right|_{\Theta=\mathrm{E}[\hat{\Theta}]}\right)^{\top}}},$$

where $\hat{\Omega}$ is some estimate of $\mathrm{Var}\left(\mathrm{vech}\left(\hat{\Theta}\right)\right)$, and rejects at the α level when W is greater than $z_{1-\alpha}$, the 1-α quantile of the standard normal distribution.

Example 4.5.1 (ex-factor Sharpe ratio on Fama-French factors (continued)). Continuing Example 4.4.1, Consider the four Fama French factors from Example 1.2.1. We model UMD as a linear combination of Market, SMB, HML and an intercept term. Now consider the idiosyncratic return of UMD. That is, we isolate the intercept portion of the attribution by taking $\boldsymbol{v} = [0, 0, 0, 1]^{\top}\%$, and we assume $r_0 = 0\%$. Via Equation 4.62 we compute,

$$\zeta_g = 0.2467\text{mo.}^{-1/2}$$

Via the sample estimate of Ω, and using the delta method, we estimate the standard error of this to be $0.0331\text{mo.}^{-1/2}$. The resulting Wald statistic is then $W = 7.4469$, and one would reject the null hypothesis that the idiosyncratic part is zero (or negative). $\dashv$

Example 4.5.2 (Technology ex-factor signal-noise ratio (continued)). Consider the ex-factor signal-noise ratio of monthly Technology industry returns against the Market, as described in Example 3.5.7. Taking $\boldsymbol{v} = [0, 1]^{\top}\%$ and $r_0 = 0\%$, we compute, via Equation 4.62

$$\zeta_g = 0.0327\text{mo.}^{-1/2}$$

First we estimate Ω by the sample estimate of the covariance of stacked vectors $\left[\tilde{\boldsymbol{x}}^{\top}, \tilde{\boldsymbol{x}}^{2\top}\right]^{\top}$. With this estimate, using the delta method, we estimated the standard error of ζ_g to be $0.0304\text{mo.}^{-1/2}$. The resulting Wald statistic is then $W = 1.0745$, and we fail to reject the null hypothesis that the idiosyncratic part is zero (or negative).

If instead we use Equation 4.69 and sample estimates of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, with $\kappa = 1$ (assuming Gaussian returns), the results change very little. We estimate the standard error of ζ_g to be $0.0306\text{mo.}^{-1/2}$, and the Wald statistic to be $W = 1.0705$. If we feed in $\kappa = 2.8969$, which is the average kurtosis of Technology and Market, divided by three, then the standard error estimate becomes $0.0314\text{mo.}^{-1/2}$, and the Wald statistic 1.0415.

Using the delta method with Equation 4.69 should give the same results as using Equation 4.71 directly. Plugging $n = 1104$, $\zeta_g = 0.0327\text{mo.}^{-1/2}$, $\zeta_m = 0.1727\text{mo.}^{-1/2}$, and $\kappa = 2.8969$ into Equation 4.71, we get a standard error estimate of $0.0314\text{mo.}^{-1/2}$, which is indeed the same.

Recall that in Example 3.5.10, we estimated the standard error of ζ_g to be $0.0301\text{mo.}^{-1/2}$. In that exercise we assumed that errors were Gaussian, that the returns of the Market were deterministic, and then adapted the approximate standard error of the t-distribution. This is nearly the same as the values we compute here via the delta method using three different estimators of Ω. In this case the extra computation has not changed our conclusions. For smaller n, however, one might see greater difference between the estimated standard errors. ⊣

4.5.2. ex-factor Sharpe ratio Prediction Intervals

As in Section 3.5.9, we can construct approximate prediction intervals on functions of $\hat{\Theta}$. So suppose that for $i = 1, 2$, $\hat{\Theta}_i$ is computed on n_i independent observations drawn from the same population. Starting from Equation 4.73, we claim that

$$\left(\operatorname{vech}\left(\hat{\Theta}_1\right) - \operatorname{vech}\left(\hat{\Theta}_2\right)\right) \rightsquigarrow \mathcal{N}\left(0, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\Omega\right), \tag{4.75}$$

where convergence is jointly in n_1 and n_2.

Then by the multivariate delta method, we now have

$$\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-\frac{1}{2}} \left(g\left(\hat{\Theta}_1\right) - g\left(\hat{\Theta}_2\right)\right) \rightsquigarrow \mathcal{N}\left(0, \left(\left.\frac{\mathrm{d}g\left(\Theta\right)}{\mathrm{d}\operatorname{vech}\left(\Theta\right)}\right|_{\Theta=\mathrm{E}\left[\hat{\Theta}\right]}\right)\Omega\left(\left.\frac{\mathrm{d}g\left(\Theta\right)}{\mathrm{d}\operatorname{vech}\left(\Theta\right)}\right|_{\Theta=\mathrm{E}\left[\hat{\Theta}\right]}\right)^{\top}\right), \tag{4.76}$$

where the derivative is as given in Equation 4.68. This is the same inflation factor given by Equation 3.40.

4.6. Sharpe ratio and Non-normality, an empirical study

The results of this chapter largely exonerate the sample Sharpe ratio under non-normality: moderate autocorrelation and heteroskedasticity do not seem to cause great bias in the Sharpe ratio, nor do they greatly inflate its standard error. The formula for the standard error of ζ which applies under normality, Equation 3.24, is commonly used for real world returns, even though Mertens' formula, Equation 3.28, is more accurate. However, Mertens' formula relies on higher order moments which must be estimated, and thus might be *less* accurate because of this additional estimation burden. Leaving behind the theoretical findings of the rest of this chapter, I present an empirical study. Roughly, the goals of this study are to compare these two estimates of the standard error and present practical guidelines and limitations for their use.

In this study, I spawn returns according to one of a number of distributions, of a given sample size and signal-noise ratio, then I use each of three different methods to construct one-sided lower confidence intervals of a given nominal type I rate, α. I then record whether the signal-noise ratio falls within the computed confidence interval. If the test maintains nominal coverage we should see the signal-noise ratio fall outside the confidence interval at the type I rate, that is, the empirical type I rate should match the nominal. For each distribution, sample size, and signal-noise ratio, I repeat this process 10^6 times, recording the empirical average type I rate. Each simulated returns series is tested with each combination of method and α.

The probability distributions used are given in Table 4.1, along with their population skews and excess kurtoses. The distributions are:

1. The normal distribution, with zero skew and excess kurtosis.
2. Student's t distribution, considered twice, with degrees of freedom $\nu = 10$ and $\nu = 4$. These are both unskewed. The $\nu = 10$ distribution has moderate kurtosis, while in the $\nu = 4$ case the kurtosis is infinite.
3. The daily returns of the S & P 500 from 1970-01-05 through 2015-12-31, shifted and rescaled to be zero mean and unit standard deviation. As shown in Table 4.1, this distribution has mild negative skew and modest kurtosis. I also sample from a 'symmetrized' S & P 500 distribution, where the absolute value of the daily S & P returns are given a random sign, which removes the skew, but leaves most of the kurtosis. When sampling from these distributions, all autocorrelation is removed.
4. Tukey's h-distribution, a special case of the 'g and h distribution' when $g = 0$, for varying values of h. If z is a standard unit normal, then $y = \mu + ze^{hz^2/2}$ takes a Tukey h distribution. I consider the cases $h \in \{0.1, 0.2, 0.3\}$, which yield unskewed, but progressively more kurtotic distributions.
5. 'Trio,' a discrete distribution with three values, a low probability, high-value payout of v which occurs with probability p, and equally probable payouts of l and h, each with probability $(1 - p)/2$. Given p and v (with $v^2 \leq (1 - p)/p$), the values of l and h are chosen so the process has zero mean and unit variance. The distribution is then shifted and scaled to achieve the target signal-noise ratio and variance.
6. Draws from the 'Lambert W $\times$ Gaussian' distribution, with different values of the skew parameter, δ. [58, 57, 59] If z takes a standard zero mean, unit variance normal distribution, then $y = \mu + ze^{\delta z}$ takes a Lambert W $\times$ Gaussian distribution with parameter δ. I consider the cases $\delta \in \{0.4, 0.2, -0.2, -0.4, -0.6\}$ with progressively more negative skews, and with kurtosis monotonic in $|\delta|$.

Again, during the simulations the returns distributions are shifted and rescaled to have a fixed signal-noise ratio. The signal-noise ratio in the simulations varies from $0\text{yr}^{-1/2}$ through $2\text{yr}^{-1/2}$. We simulate the case of observing between 1 through 48 months of *daily* returns, where a month always consists of exactly 21 days of returns.

I consider three methods of computing confidence intervals:

1. Approximating the standard error by 'plugging in' to the form of Equation 3.24, and assuming approximate normality of the Sharpe ratio. I refer to this as 'vanilla' for the vanilla standard error, Equation 3.24

nickname	family	param	skew	kurtosis
normal	Gaussian		0	0
t10	Student's t	$\nu = 10$	0	1
t4	Student's t	$\nu = 4$	0	∞
sp500	SP500		-1.01	25.4
symsp500	Symmetric SP500		0	25.3
trio2	Trio	$p = 0.02, v = 2.5$	0.179	-1.43
trio1	Trio	$p = 0.01, v = 5$	1.14	3.83
trio05	Trio	$p = 0.005, v = 10$	4.92	47.3
tukey1	Tukey's h	$h = 0.1$	0	5.51
tukey2	Tukey's h	$h = 0.2$	0	36.2
tukey3	Tukey's h	$h = 0.3$	0	∞
lambertp4	Lambert $\times$ Gaussian	$\delta = 0.4$	2.74	17.8
lambertp2	Lambert $\times$ Gaussian	$\delta = 0.2$	1.24	5.69
lambertn2	Lambert $\times$ Gaussian	$\delta = -0.2$	-1.24	5.69
lambertn4	Lambert $\times$ Gaussian	$\delta = -0.4$	-2.74	17.8
lambertn6	Lambert $\times$ Gaussian	$\delta = -0.6$	-4.9	58.1

Table 4.1.: The distributions used in the empirical are listed, with theoretical skew and *excess* kurtosis of each. The S&P daily returns are sampled from 1970-01-05 through 2015-12-31.

2. Doing the same, but using Mertens' correction, Equation 3.28, called 'mertens'.
3. Assuming returns are drawn from a normal and using the quantile function of the non-central t-distribution to compute confidence intervals. This is referred to as the 't' method.

In all, I considered 5 values of signal-noise ratio, 7 values of the number of months, and 16 different returns distributions. This translates into 560 sets of simulations, each of which consist of 10^6 replications of generating the returns stream and then computing confidence intervals for 4 different values of the type I rate for each of the 3 methods considered. The result is a collection of 6720 empirical type I rejection rates.

It is not immediately clear how to evaluate and compare different methods of constructing confidence intervals. I compute the 'conservatism' of each set of simulations, defined as the nominal type I rate divided by the empirical rate. This should be exactly equal to one; larger values indicate the confidence intervals are conservative; values smaller than 1 indicate the test is anti-conservative. Having computed the conservatism, I boxplot them in Figure 4.7 versus distribution skew, with facets for the number of months observed. In Figure 4.8 the conservatism is boxplotted against the population excess kurtosis.

One clear trend visible in Figure 4.7 is that all three methods seem to be anti-conservative for negative skew, and conservative for large positive skew. This is consistent with what we know of the true standard error of the Sharpe ratio via Equation 3.28, *i.e.*, that it is smaller when the population skew and the signal-noise ratio have the same size, and thus the we see a lower type I rate in practice. This is 'solved' to some degree by using the Mertens' method for estimating standard errors, which sees near-nominal coverage for large absolute skew. Conversely, when the true skew is near or exactly zero, and thus Equation 3.24 gives the standard error of the Sharpe ratio to a good approximation, Mertens' method suffers from its ability to mis-estimate the skew.

A similar pattern is seen in Figure 4.8, with Mertens' method giving better coverage for large excess kurtosis, though not universally, and perhaps worse coverage when the true population excess kurtosis is zero. Again, this is based on the observation that if what you assume away when simplifying Mertens' form to the vanilla form happens to be true, you come out ahead, otherwise you lose control of your type I rate. Note that the skew of the S&P 500 returns is apparently sufficiently close enough to zero that Mertens' method has worse coverage for small sample sizes and perhaps no better for large sample sizes.

Mertens Correction

Mertens correction is appropriate when population skew or kurtosis are known to be large, otherwise the usual standard error is acceptable.

Note that this 'rule' is merely a realization of the fact that if you can assume away something true (*e.g.*, that the population skew is truly zero), you gain inferential power, and, conversely, if what you assume is false you lose power. We will witness this rule again.

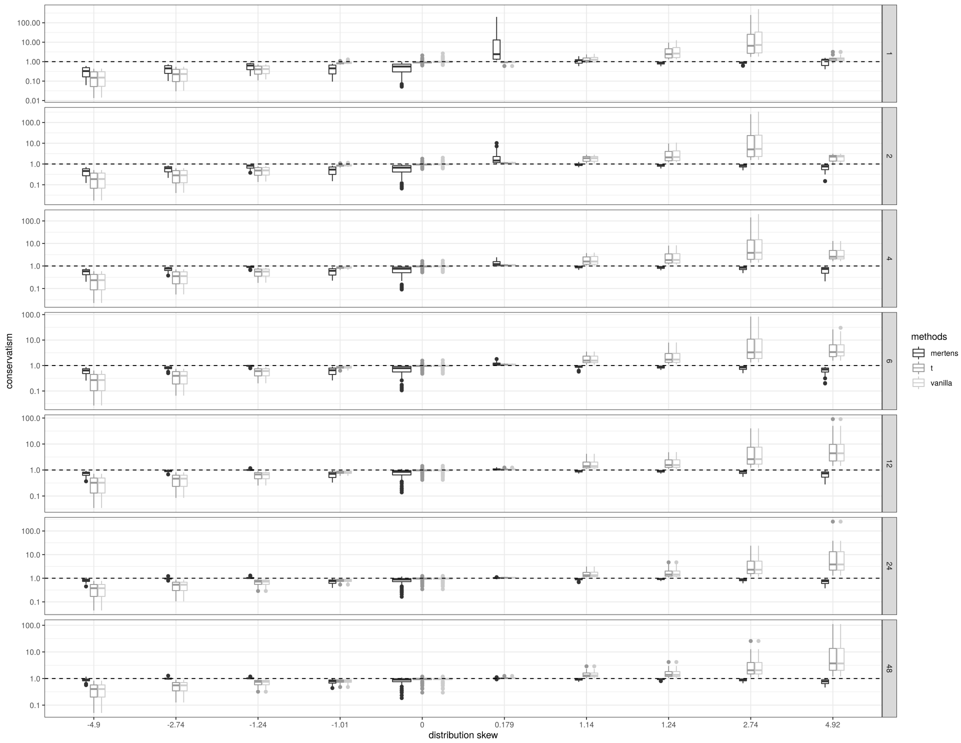

Figure 4.7.: The geometric conservatism for each confidence interval is plotted versus the skew of the returns distribution. Conservatism is defined as the nominal type I rate divided by the empirical rate. Each box is over multiple values of signal-noise ratio, type I rate, and possibly returns distributions. Facets represent the number of months of daily returns simulated. The x axis is not to scale. The boxes are plotted, from left to right, for the methods 'vanilla', 'mertens', and 't.' For non-zero skews Mertens' method does a better job of maintaining near nominal coverage than the other two methods, while it seems to perform worse for zero skew.

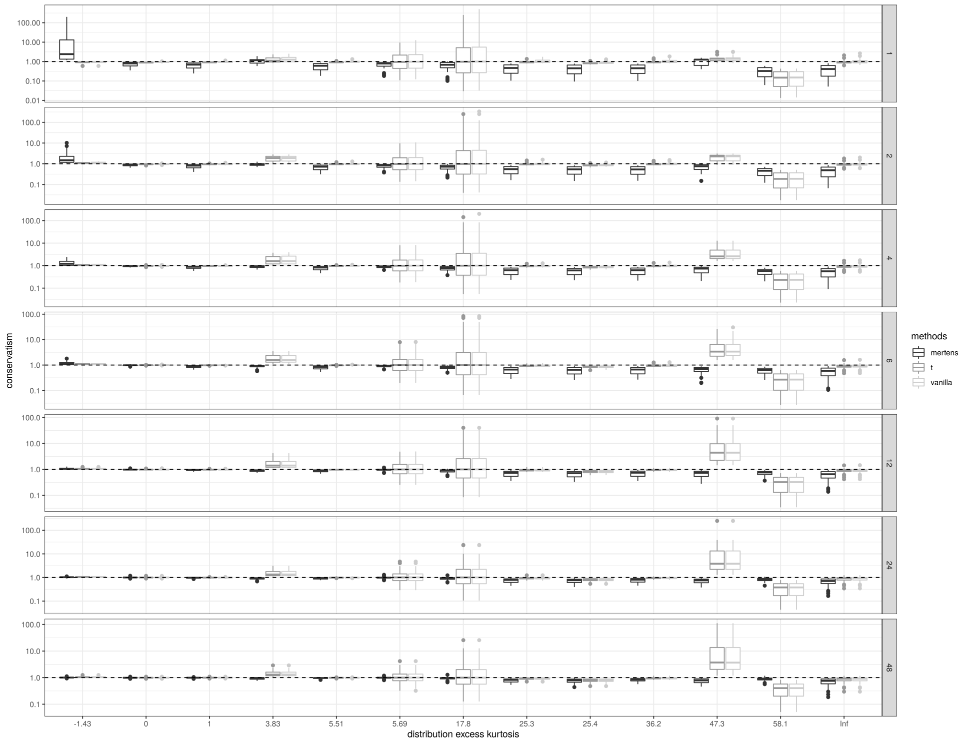

Figure 4.8.: The geometric conservatism for each confidence interval is plotted versus the excess kurtosis of the returns distribution. Conservatism is defined as the nominal type I rate divided by the empirical rate. Each box is over multiple values of signal-noise ratio, type I rate, and possibly returns distributions. Facets represent the number of months of daily returns simulated. The x axis is not to scale. The boxes are plotted, from left to right, for the methods 'vanilla', 'mertens', and 't.'

Exercises

Ex. 4.1 The Market, Heteroskedasticity by quarter Repeat the analysis of Figure 4.2, but aggregating mean and standard deviation of daily log returns over quarters.

1. Does the negative relationship still hold? Is it statistically significant?
2. Compute the autocorrelation of quarterly aggregated volatilities.
3. Since volatility seems autocorrelated, and current quarter volatility seems to be correlated with negative returns, look for a market timing model based on vol. Lag the quarterly volatilities by one quarter, and scatter quarterly mean returns against last quarter's volatility. Is there still a negative relationship? Is it significant?

** **Ex. 4.2 Expectation of products of Elliptical Random Variables**
Let $\boldsymbol{x}$ follow an Elliptical distribution with parameters $\boldsymbol{\mu}$, Σ, and kurtosis factor κ (*cf.* Section 1.3.2.). In the Gaussian case, we have $\kappa = 1$.
An extension of Isserlis' theorem to Elliptical distributions gives us

$$\begin{aligned} \mathrm{E}\left[\left(x_i - \mu_i\right)\left(x_j - \mu_j\right)\left(x_k - \mu_k\right)\right] &= 0,\\ \mathrm{E}\left[\left(x_i - \mu_i\right)\left(x_j - \mu_j\right)\left(x_k - \mu_k\right)\left(x_l - \mu_l\right)\right] &= \kappa\left(\Sigma_{i,j}\Sigma_{k,l} + \Sigma_{i,k}\Sigma_{j,l} + \Sigma_{i,l}\Sigma_{j,k}\right), \end{aligned}$$

where there may be replication in the indices i, j, k, l. [163]

1. Show that
$$\mathrm{E}\left[x_i x_j x_k\right] = \mu_i\mu_j\mu_k + \mu_i\Sigma_{j,k} + \mu_j\Sigma_{i,k} + \mu_k\Sigma_{i,j}.$$
2. Show that
$$\begin{aligned} \mathrm{E}\left[x_i x_j x_k x_l\right] = \kappa&\left(\Sigma_{i,j}\Sigma_{k,l} + \Sigma_{i,k}\Sigma_{j,l} + \Sigma_{i,l}\Sigma_{j,k}\right)\\ &+ \mu_i\mu_j\Sigma_{k,l} + \mu_i\mu_k\Sigma_{j,l} + \mu_i\mu_l\Sigma_{j,k}\\ &+ \mu_j\mu_k\Sigma_{i,l} + \mu_j\mu_l\Sigma_{i,k} + \mu_k\mu_l\Sigma_{i,j}\\ &+ \mu_i\mu_j\mu_k\mu_l \end{aligned}$$
3. Show that
$$\mathrm{E}\left[\left(\boldsymbol{x}^{\top}\boldsymbol{x}\right)\right] = \boldsymbol{\mu}^{\top}\boldsymbol{\mu} + \operatorname{tr}\left(\Sigma\right) = \operatorname{tr}\left(\boldsymbol{\mu}\boldsymbol{\mu}^{\top} + \Sigma\right).$$
4. Show that
$$\mathrm{E}\left[\boldsymbol{x}^{\top}\boldsymbol{x}\boldsymbol{x}^{\top}\mathbf{1}\right] = \boldsymbol{\mu}^{\top}\boldsymbol{\mu}\boldsymbol{\mu}^{\top}\mathbf{1} + 2\boldsymbol{\mu}^{\top}\Sigma\mathbf{1} + \operatorname{tr}\left(\Sigma\right)\boldsymbol{\mu}^{\top}\mathbf{1}.$$
5. Show that
$$\mathrm{E}\left[\left(\boldsymbol{x}^{\top}\mathbf{1}\right)^3\right] = \left(\boldsymbol{\mu}^{\top}\mathbf{1}\right)^3 + 3\boldsymbol{\mu}^{\top}\mathbf{1}\mathbf{1}^{\top}\Sigma\mathbf{1}.$$

6. Show that
$$\operatorname{Cov}\left(x_i, {x_j}^2\right) = 2\mu_j \Sigma_{i,j}.$$
7. Show that
$$\begin{aligned}\operatorname{Cov}\left(x_i x_j, x_j x_k\right) = (\kappa - 1)\, \Sigma_{i,j}\Sigma_{k,l} + \kappa\left(\Sigma_{i,k}\Sigma_{j,l} + \Sigma_{i,l}\Sigma_{j,k}\right)\\ + \mu_i\mu_k\Sigma_{j,l} + \mu_i\mu_l\Sigma_{j,k} + \mu_j\mu_k\Sigma_{i,l} + \mu_j\mu_l\Sigma_{i,k}.\end{aligned}$$
Use this to prove that
$$\operatorname{Cov}\left({x_i}^2, {x_j}^2\right) = (\kappa - 1)\, \Sigma_{i,i}\Sigma_{j,j} + 2\kappa\Sigma_{i,j}\Sigma_{i,j} + 4\mu_i\mu_j\Sigma_{i,j}.$$
8. Let $\boldsymbol{y} = \operatorname{vec}\left(\boldsymbol{x}\boldsymbol{x}^\top\right)$. Show that
$$\operatorname{E}\left[\boldsymbol{y}\right] = \boldsymbol{\mu} \otimes \boldsymbol{\mu} + \operatorname{vec}\left(\Sigma\right).$$
Show that
$$\begin{aligned}\operatorname{Var}\left(\boldsymbol{y}\right) = (\kappa - 1)\left[\operatorname{vec}\left(\Sigma\right)\operatorname{vec}\left(\Sigma\right)^\top + \left(\mathsf{I} + \mathsf{K}\right)\Sigma \otimes \Sigma\right]\\ + \left(\mathsf{I} + \mathsf{K}\right)\left(\left(\boldsymbol{\mu}\boldsymbol{\mu}^\top + \Sigma\right) \otimes \left(\boldsymbol{\mu}\boldsymbol{\mu}^\top + \Sigma\right) - \left(\boldsymbol{\mu}\boldsymbol{\mu}^\top\right) \otimes \left(\boldsymbol{\mu}\boldsymbol{\mu}^\top\right)\right),\end{aligned}$$
where K is the commutation matrix of Definition 1.1.3.
9. Prove Equation 4.27: let $\boldsymbol{y} = \left[\boldsymbol{x}^\top, {\boldsymbol{x}^2}^\top\right]^\top$, and let Ω be the covariance matrix of $\boldsymbol{y}$. Show that Ω equals
$$\left[\begin{array}{cc}\Sigma & 2\Sigma\operatorname{Diag}\left(\mu\right)\\ 2\operatorname{Diag}\left(\mu\right)\Sigma & (\kappa - 1)\operatorname{diag}\left(\Sigma\right)\left(\operatorname{diag}\left(\Sigma\right)\right)^\top + 2\kappa\left(\Sigma \odot \Sigma\right) + 4\operatorname{Diag}\left(\mu\right)\Sigma\operatorname{Diag}\left(\mu\right)\end{array}\right].$$
where $\operatorname{Diag}\left(\mu\right)$ is the matrix with vector μ on its diagonal, and $\operatorname{diag}\left(\Sigma\right)$ is the column vector of the diagonal of Σ.

Ex. 4.3 **Sweeping column means** Let A be a matrix with n rows (or a column vector of length n). Show that $\frac{\mathbf{1}\mathbf{1}^\top}{n}\mathsf{A}$ is the row vector of means of each column of A. Show that each column of $\mathsf{A} - \frac{\mathbf{1}\mathbf{1}^\top}{n}\mathsf{A}$ is zero-mean.

** **Ex. 4.4** **More moments of products of Elliptical Random Variables**
Let $\boldsymbol{x}$ follow an Elliptical distribution with parameters $\boldsymbol{\mu}$, Σ, and kurtosis factor κ. Use results from Exercise 4.2 to prove the relationhip in Equation 4.3.

1. Show that
$$\operatorname{E}\left[\left(\boldsymbol{x}^\top\mathbf{1}\right)^4\right] = \left(\boldsymbol{\mu}^\top\mathbf{1}\right)^4 + 3\kappa\left(\mathbf{1}^\top\Sigma\mathbf{1}\right)^2 + 6\left(\boldsymbol{\mu}^\top\mathbf{1}\right)^2\mathbf{1}^\top\Sigma\mathbf{1}.$$
Use this to prove that
$$\operatorname{Var}\left(\left(\boldsymbol{x}^\top\mathbf{1}\right)^2\right) = (3\kappa - 1)\left(\mathbf{1}^\top\Sigma\mathbf{1}\right)^2 + 4\left(\boldsymbol{\mu}^\top\mathbf{1}\right)^2\mathbf{1}^\top\Sigma\mathbf{1}.$$

2. Show that
$$\mathrm{E}\left[\left(\boldsymbol{x}^{\top}\boldsymbol{x}\right)^2\right] = \left(\boldsymbol{\mu}^{\top}\boldsymbol{\mu}\right)^2 + \kappa\left(\mathrm{tr}\left(\Sigma\right)\right)^2 + 2\kappa\,\mathrm{tr}\left(\Sigma\Sigma\right) + 2\boldsymbol{\mu}^{\top}\boldsymbol{\mu}\,\mathrm{tr}\left(\Sigma\right) + 4\boldsymbol{\mu}^{\top}\Sigma\boldsymbol{\mu}.$$
Use this to prove that
$$\mathrm{Var}\left(\left(\boldsymbol{x}^{\top}\boldsymbol{x}\right)\right) = \left(\kappa - 1\right)\left(\mathrm{tr}\left(\Sigma\right)\right)^2 + 2\kappa\,\mathrm{tr}\left(\Sigma\Sigma\right) + 4\boldsymbol{\mu}^{\top}\Sigma\boldsymbol{\mu}.$$

3. Show that
$$\begin{aligned}\mathrm{E}\left[\left(\boldsymbol{x}^{\top}\mathbf{1}\right)^2\boldsymbol{x}^{\top}\boldsymbol{x}\right] &= \left(\kappa\mathbf{1}^{\top}\Sigma\mathbf{1} + \boldsymbol{\mu}^{\top}\boldsymbol{\mu}\right)\mathrm{tr}\left(\Sigma\right) + 2\kappa\mathbf{1}^{\top}\Sigma^2\mathbf{1}\\ &\quad + \left(\boldsymbol{\mu}^{\top}\boldsymbol{\mu} + \mathrm{tr}\left(\Sigma\right)\right)\left(\boldsymbol{\mu}^{\top}\mathbf{1}\right)^2 + 4\left(\boldsymbol{\mu}^{\top}\mathbf{1}\right)\boldsymbol{\mu}^{\top}\Sigma\mathbf{1}.\end{aligned}$$
Use this to prove that
$$\mathrm{Cov}\left(\boldsymbol{x}^{\top}\boldsymbol{x}, \left(\boldsymbol{x}^{\top}\mathbf{1}\right)^2\right) = \left(\kappa - 1\right)\mathbf{1}^{\top}\Sigma\mathbf{1}\,\mathrm{tr}\left(\Sigma\right) + 2\kappa\mathbf{1}^{\top}\Sigma^2\mathbf{1} + 4\left(\boldsymbol{\mu}^{\top}\mathbf{1}\right)\boldsymbol{\mu}^{\top}\Sigma\mathbf{1}.$$

** 4. Use the results of the previous three questions to show that
$$\mathrm{Var}\left(\boldsymbol{x}^{\top}\boldsymbol{x} - \frac{\left(\boldsymbol{x}^{\top}\mathbf{1}\right)^2}{n}\right) = \left(\kappa - 1\right)\mathrm{tr}\left(\Sigma_c\right)\mathrm{tr}\left(\Sigma_c\right) + 2\kappa\,\mathrm{tr}\left(\Sigma_c\Sigma_c\right) + 4\,\mathrm{tr}\left(\boldsymbol{\mu}_c{\boldsymbol{\mu}_c}^{\top}\Sigma\right),$$
where $\boldsymbol{\mu}_c =_{\mathrm{df}} \boldsymbol{\mu} - \frac{\mathbf{1}\mathbf{1}^{\top}}{n}\boldsymbol{\mu}$ and $\Sigma_c =_{\mathrm{df}} \Sigma - \frac{\mathbf{1}\mathbf{1}^{\top}}{n}\Sigma$ are the column centered mean and covariance. This is the bottom corner of Equation 4.3.

Ex. 4.5 Proofs

1. Derive Equation 4.11 from Equation 4.7 assuming $\kappa = 1$.

* 2. Derive Equation 4.11 from Equation 4.7 for the case of general κ.

Ex. 4.6 Elliptical marginals over time Let $\boldsymbol{x}$ be an n-vector which follows an Elliptical distribution with mean zero and covariance parameter $\Sigma = \sigma^2\mathsf{I}$, and kurtosis factor κ. Let $s^2 = \boldsymbol{x}^{\top}\boldsymbol{x}/n$

1. Show that $\mathrm{E}\left[s^2\right] = \sigma^2$.
2. Show that $\mathrm{Var}\left(s^2\right) = \sigma^4\left(\kappa - 1 + \frac{2\kappa}{n}\right)$.
3. Confirm that the standard error of s^2 does not go to zero as $n \to \infty$ by way of Monte Carlo simulations using draws from a multivariate t distribution with 5 degrees of freedom.

Ex. 4.7 Autocorrelation matrix computations Let Σ be the $n \times n$ correlation matrix associated with simple autocorrelation, from Section 4.1.4. That is, $\Sigma_{i,j} = \sigma^2\rho^{|i-j|}$.

1. Show that $\mathbf{1}^{\top}\mathbf{\Sigma}\mathbf{1} \approx \sigma^2 \left(n\frac{1+\rho}{1-\rho} + \frac{2\rho}{(1-\rho)^2}\right)$ for large n. (*Hint:* write $\mathbf{1}^{\top}\mathbf{\Sigma}\mathbf{1}/\sigma^2 = n + 2(n-1)\rho^1 + 2(n-2)\rho^2 + \ldots + 2\rho^{n-1}$. Then derive approximations for $n\sum_i \rho^i$ and $\sum_i i\rho^{i-1}$. For the latter, use calculus.)
2. Show that $\operatorname{tr}\left(\mathbf{\Sigma}^2\right) \approx \sigma^4 \left(n\frac{1+\rho^2}{1-\rho^2} + \frac{2\rho^2}{(1-\rho^2)^2}\right)$ for large n. (*Hint:* Use the previous approximation.)

Ex. 4.8 Autocorrelation and standard error A positive (negative) autocorrelation inflates (deflates) the standard error of the Sharpe ratio. Can you give an intuitive explanation for why this is so? How should an investor buy a strongly mean reverting asset? A strongly trend-following asset?

* **Ex. 4.9 Bounds on heteroskedasticity bias** Suppose you observe otherwise homoskedastic returns levered by a fund manager (*i.e.*, the $\lambda = 1$ case). The fund manager can choose leverage l between $\frac{1}{2}$ and 2. What is the largest possible value of the squared coefficient of variation of l? (Note: this corresponds to normalizing to unit mean, and then computing M_2.) Suppose the underlying homoskedastic returns have signal-noise ratio between 0 and $0.15\text{day}^{-1/2}$.

1. What is the smallest possible value of the geometric bias, b?
2. What is the largest possible value of the (approximate) variance of the Sharpe ratio, given in Equation 4.19?

Ex. 4.10 Homoskedastic varying mean returns In Section 4.1.3, we considered the case where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$ vary together. Here consider the homoskedastic case, *i.e.*, where $\mathsf{\Sigma} = \sigma^2\mathsf{I}$, with varying mean: $\boldsymbol{\mu} = \mu\boldsymbol{l}$, where $\frac{1}{n}\mathbf{1}^{\top}\boldsymbol{l} = 1$. Again, M_2 is the squared coefficient of variation of $\boldsymbol{l}$.

1. Assuming $0 \le \mu/\sigma \le 0.2$, what is the smallest possible value of b, expressed in terms of M_2?
2. Simplify Equation 4.11 for this case. Express the variance in terms of μ, σ, b and M_2. Is it much different than in the *i.i.d.* homoskedastic case?

Ex. 4.11 Returns at different frequencies Repeat the exercise of Example 4.1.3, but under different situations:

1. Assume you have 10 years of monthly data and 2 years of daily data (at a rate of 253 days per year). Assume $r_0 = 0$, $\mu = 0.0012\text{day}^{-1}$, $\sigma = 0.02\text{day}^{-1/2}$. What is the bias in the Sharpe ratio? Converting the Sharpe ratio back to daily units, what is the approximate standard error? If, instead, one had observed 12 years of daily data, what would the standard error be? Perform 1000 Monte Carlo simulations to test your work.
2. Assume you have 100 years of yearly data and 20 years of monthly data on normal returns. Assume $r_0 = 0$, $\mu = 0.014\text{mo.}^{-1}$, $\sigma = 0.05\text{mo.}^{-1/2}$. What is the bias in the Sharpe ratio? Converting the Sharpe ratio back to monthly units, what is the approximate standard error? If, instead, one had observed

120 years of monthly data, what would the standard error be? Perform 1000 Monte Carlo simulations to test your work.

Ex. 4.12 Any unskewed distribution Suppose returns are drawn *i.i.d.* from the product of a normal and an independent positive random variable. That is, let

$$x_i = z_i l_i + \mu,$$

where z_i are *i.i.d.* standard normal random variables independent of l_i which are *i.i.d.* and positive. Let M_k be the raw k^{th} population moment of l_i:

$$M_k =_{\text{df}} \operatorname{E}\left[l^k\right].$$

Under these assumptions, the analysis of Section 4.1.3 is relevant with $\lambda = 0$, except here we are using M_k to be the *population* moments whereas we had previously defined M_k to be the *empirical* moment for a given sample. For large n, the empirical moments will approach the population moments.

1. Find the mean and variance of x. What is the signal-noise ratio of x?
2. Find the skew and kurtosis of x, in terms of the M_k.
3. Using Equation 4.16, what is the expected geometric bias of $\hat{\zeta}$?
4. Using Equation 4.19, write the approximate standard error of $\hat{\zeta}$.
5. Draw $n = 10^4$ realizations of a log normal l_i with $M_1 = 1$ and $M_2 = e^1$. Multiply these by independent normals z_i and add $\mu = 0.001$ to get x_i, and compute the signal-noise ratio of the x series. Repeat this 100 times. Confirm the theoretical bias and standard errors of $\hat{\zeta}$.

* 6. Compute the centered moments of x with respect to M_k. Is it the case that *any* unskewed probability distribution admits a representation as the product of a normal and an independent positive random variable? (The point here is that the results of Section 4.1.3 are not universally applicable to all unskewed returns distributions.)

Ex. 4.13 Correlation of Fama-French Sharpe ratio, elliptical returns Repeat the analysis of Example 4.2.3, but assume the Fama French returns follow an elliptical distribution, and use Equation 4.29 to compute the variance-covariance of $\hat{\boldsymbol{\zeta}}$. Take the four Fama French factors monthly returns from Jan 1927 to Dec 2020.

1. Using those returns, estimate the correlation matrix R via the usual sample estimator.
2. Using those returns, estimate the kurtosis of each series separately. Take the median value and divide by three to get an estimate of κ.
3. Use Equation 4.29 to compute the variance-covariance of $\hat{\boldsymbol{\zeta}}$.
4. Turn that variance-covariance estimate into a correlation estimate and compare to the results of Example 4.2.3.

5. Is elliptical distribution a good assumption for the Fama French returns? Check by computing the skewness of the 'HML' returns, and bootstrapping for significance. Elliptical distributions have zero skewness.

Ex. 4.14 Equality of Industry signal-noise ratios Use the Leung-Wong test (Equation 4.41) on the Fama French 5 Industry returns data to test the hypothesis that the 5 industries have equal signal-noise ratios.

Ex. 4.15 Comparing tests of equality of signal-noise ratios Wright *et al.* find that the chi-square test (Inequality 4.42) has closer to nominal coverage than the F-test (Inequality 4.43) for fat-tailed distributions. Replicate their work. Set $n = 2000$, and let $p = 5, 10, 20$. Draw returns from multivariate normal, and from the Elliptical multivariate t distribution with $4, 6$, and 8 degrees of freedom. Let the returns have zero mean. Test the hypothesis that all assets have the same signal-noise ratio at the $\alpha = 0.05$ level. Use Equation 4.27 to compute $\hat{\Omega}$, assuming normally distributed returns. Compare these with table I of Wright *et al.*, who find, for example, that for $t\,(4)$ returns and $p = 20$, the F test has an empirical rejection rate of around 74.7%, while the chi-square test has a rejection rate of around 17.3%. [174]

1. Repeat the experiment using the chi-square statistic, but use Equation 4.27 assuming the true value of κ is known to compute $\hat{\Omega}$.
2. Repeat the experiment using the chi-square statistic, but use HAC estimators on the sample of stacked vectors $\left[\boldsymbol{x}^{\top}, {\boldsymbol{x}^{2}}^{\top}\right]^{\top}$ to compute $\hat{\Omega}$, as in Ledoit & Wolf. [89]

Ex. 4.16 Testing equality of signal-noise ratios, unpaired samples The method outlined for testing multiple signal-noise ratios outlined in Section 4.3 and Equation 4.41 apply to the case of a paired sample, *i.e.*, contemporaneous returns with possible correlation among them.

1. Formulate the test for equality of 2 signal-noise ratios with an unpaired sample by abusing Equation 4.41. (Simply compute Ω as a diagonal matrix.) Your test should be capable of testing against a one-sided alternative.
2. Use your test on the returns of the Market, dividing the sample into two periods using the cutoff date of 1970-01-01. *cf.* Exercise 3.9, where the same test is performed assuming Gaussian returns.
3. Formulate the test for equality of multiple signal-noise ratios with an unpaired sample.

Ex. 4.17 Testing equality of signal-noise ratios, mixed samples Suppose you observe two returns streams with some period of overlap. So for example, suppose you observe returns x_i for $i = 1, \ldots, n_1$ and returns y_i for $i = m, m+1, \ldots, m+n_2-1$ with $1 < m \leq n_1 < m+n_2-1$. Formulate the test for equality of 2 signal-noise ratios in this case.

Ex. 4.18 Overlapping Returns Suppose you observe overlapping returns. That is, x_i correspond to returns over a time period of length q, measured in years, observed at a rate of $1/r$ per year, evenly spaced through the year, with $q \geq r$. For example, suppose you observe rolling quarterly returns every month, in which case $q = \frac{1}{4}$ and $r = \frac{1}{12}$. *cf.* Valkanov and Britten-Jones and Neuberger. [161, 23, 19]

1. Show that the autocorrelation of returns is $\rho = 1 - \frac{r}{q}$.
2. Using Equation 4.22, show that the standard error of the Sharpe ratio is inflated by a factor of approximately $\sqrt{2\frac{q}{r} - 1}$.
3. Confirm the inflated standard error empirically by computing 5 years of rolling quarterly returns observed monthly of a zero mean random variable, and computing the Sharpe ratio, repeating the experiment thousands of times.

* **Ex. 4.19 Augmented form moments, Normal returns** Let $\tilde{\mathsf{X}}$ be an normally distributed $n \times l$ matrix, *i.e.*, $\operatorname{vec}\left(\tilde{\mathsf{X}}\right)$ is normal with mean $\operatorname{vec}\left(\mathsf{M}\right)$, covariance $\Sigma_f \otimes \mathsf{H}$. As in Equation 4.58, define

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

Using results from Exercise 4.2, confirm Equation 4.65:

1. Show that
$$\operatorname{E}\left[\hat{\Theta}\right] = \frac{1}{n}\left(\mathsf{M}^{\top}\mathsf{M} + \operatorname{tr}\left(\mathsf{H}\right)\Sigma_f\right).$$
2. Show that
$$\operatorname{Var}\left(\operatorname{vec}\left(\hat{\Theta}\right)\right) = \frac{1}{n^2}\left(\mathsf{I} + \mathsf{K}\right)\left\{\mathsf{M}^{\top}\mathsf{H}\mathsf{M} \otimes \Sigma_f + \Sigma_f \otimes \mathsf{M}^{\top}\mathsf{H}\mathsf{M} \right. \\ \left. + \operatorname{tr}\left(\mathsf{H}^2\right)\Sigma_f \otimes \Sigma_f\right\}.$$

* **Ex. 4.20 Augmented form moments, Elliptical returns** Establish the variance equation of Equation 4.69. [121] Let $\tilde{\boldsymbol{x}}$ take an Elliptical distribution with kurtosis factor κ, mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.

1. Using results from Exercise 4.2, prove
$$\operatorname{Var}\left(\operatorname{vec}\left(\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\top}\right)\right) = \left\{\left(\kappa_i - 1\right)\left[\operatorname{vec}\left(\Sigma_i\right)\operatorname{vec}\left(\Sigma_i\right)^{\top} + \left(\mathsf{I} + \mathsf{K}\right)\Sigma_i \otimes \Sigma_i\right] \right. \\ \left. + \left(\mathsf{I} + \mathsf{K}\right)\left[\Theta_i \otimes \Theta_i - \boldsymbol{\mu}_i{\boldsymbol{\mu}_i}^{\top} \otimes \boldsymbol{\mu}_i{\boldsymbol{\mu}_i}^{\top}\right]\right\}.$$

Ex. 4.21 The ex-factor Sharpe ratio distribution

1. Confirm that Equation 4.60 gives the inverse of the matrix Θ from Equation 4.59.

2. Confirm that the matrix in Equation 4.61 is the Cholesky factor of the matrix from Equation 4.60.
3. Confirm the identity of Equation 4.62.

Ex. 4.22 Moments under general returns
Suppose

$$\mathrm{E}\left[\tilde{\mathsf{X}}\right] = \mathsf{M},$$
$$\mathrm{Var}\left(\mathrm{vec}\left(\tilde{\mathsf{X}}\right)\right) = \Sigma_f \otimes \mathsf{H},$$

1. Let the i^{th} row of $\tilde{\mathsf{X}}$ be $\tilde{\boldsymbol{x}}_i^\top$. Show that

$$\mathrm{E}\left[\tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i^\top\right] = \mathsf{M}_{i,}{}^\top \mathsf{M}_{i,} + \mathsf{H}_{i,i}\Sigma_f,$$

where $\mathsf{M}_{i,}$ is the i^{th} row of M.

2. From $\mathbf{1}_{\boldsymbol{n}}{}^\top \mathsf{M} = n\boldsymbol{\mu}$, $\mathrm{tr}\left(\mathsf{H}\right) = n$, confirm Equation 4.66, *i.e.*,

$$\Theta = \Sigma_f + \boldsymbol{\mu}\boldsymbol{\mu}^\top.$$

3. Show that $\hat{\Theta}$ is an unbiased estimator of Θ only when $\mathsf{M} = \mathbf{1}\boldsymbol{\mu}^\top$.

Ex. 4.23 The ex-factor Sharpe ratio variance, Market returns

1. What happens to the variance of Equation 4.71 as the returns of the Market factor approach a non-zero constant (over time) value?
2. Justify *why* that should happen to the variance.

Ex. 4.24 ex-factor Sharpe ratio standard error, Market term Prove Equation 4.70.

Ex. 4.25 The ex-factor Sharpe ratio variance, multiple attribution
Consider what happens to the standard error of the ex-factor Sharpe ratio expressed in Equation 4.71 as *multiple* market instruments are added to the attribution model. Assume $\kappa = 1$. The math is probably too complicated to find an analytical solution, so explore it empirically. Repeat the empirical experiments of Example 4.4.3, but expand the number of 'Market' terms against which one performs attribution. Look at $1, 2, 4, 8, 16$, and 32 terms in a factor model. For each choice of number of factors keep the sum of squared signal-noise ratios constant at a value of, say, $1.0\mathrm{yr}^{-1}$. Fix the beta of the asset at 1 against each Market term, and keep the ex-factor signal-noise ratio and volatility as in Example 4.4.3. The Market terms should have *independent* returns. Plot the empirical standard error versus the number of factors.

1. Repeat the experiment, but make the Market returns highly positively correlated. How does the standard error change?

Ex. 4.26 **Mertens correction** Show that Merten's correction of Equation 4.34 follows from the Delta method form of Equation 4.26.

Ex. 4.27 **Mertens contradiction** It appears that if the skew of the returns distribution is sufficiently large and positive, the standard error of the Sharpe ratio under Mertens' formula, Equation 4.34, can become negative.

1. Show that this is not the case.
2. Let $\zeta = 0.1\text{day}^{-1/2}$. What is the smallest value that $\left(1 - \zeta\gamma_1 + \frac{\gamma_2+2}{4}\zeta^2\right)$ can attain for skew γ_1 and excess kurtosis γ_2? Do you know of a distribution that achieves this value?

Ex. 4.28 **Test of equality of Sharpe ratios, correlated returns** Replicate the Monte Carlo simulations of Example 4.3.3.

1. Confirm that the empirical mean of the difference in Sharpe ratios is consistent with the value suggested by Equation 4.45.
2. Generalize Equation 4.45 to the case of bivariate Elliptical returns. (*cf.* Example 4.2.1.)
3. The form of the standard deviation of the difference in Sharpe ratios given in Equation 4.45 is predicated on bivariate Gaussian returns. Modify your simulation so that returns marginally follow a scaled, shifted $t(4)$ distribution. Does the empirical standard deviation still follow the form given by Equation 4.45? Is it more accurately described by the formula you found for Elliptical returns?

Ex. 4.29 **Asymptotic prediction intervals on Sharpe ratio** In Section 3.5.9, Frequentist prediction intervals on the Sharpe ratio for the case of *i.i.d.* Gaussian returns. Here consider *i.i.d.* returns without an assumption of normality. That is, suppose you observe $\hat{\zeta}_1$ on n_1 observations of *i.i.d.* returns, then observe $\hat{\zeta}_2$ on n_2 observations from the same returns stream.

1. Find some interval that is a function of $\hat{\zeta}_1, n_1, n_2$ such that with probability $1 - \alpha$, $\hat{\zeta}_2$ falls within the interval, where the probability is under replication of the entire experiment. Start from Equation 4.39.

* 2. Now assume that the returns are Gaussian, but follow an AR(1) process (see Exercise 2.29), with autocorrelation ρ. Assume, however, that the two samples, of size n_1 and n_2, are independent. Construct a prediction interval on $\hat{\zeta}_2$.

Ex. 4.30 **Bayesian update** Multiply the prior probabilities of Equation 4.49 and Equation 4.48 by the joint likelihood of Equation 4.51 to arrive at the posterior updating rules of Equation 4.53.

Ex. 4.31 **Bayesian analysis, difference of signal-noise ratios** Perform the analysis of Example 4.3.4, but use an uninformative prior.

Ex. 4.32 **Functional Sharpe ratio** Suppose one has backtested the returns of a number of different trading strategies, say k of them. For each strategy one

observes some vector of 'features' about the strategy, call it $\boldsymbol{f}_i$, for $i = 1, 2, \ldots, k$. For example, one feature might be the length of a sliding window over which to perform some machine learning mumbo jumbo to construct a portfolio, or whether a certain signal is used in constructing the portfolio, *etc.*
Let ζ_i be the signal-noise ratio of the i^{th} strategy. A simple linear model posits that $\zeta_i = {\boldsymbol{f}_i}^{\top}\boldsymbol{\beta}$. Suppose you compute the Sharpe ratio of the backtested returns, along with the estimated variance-covariance of the same, call it $\hat{\Omega}$.

1. How would you estimate $\boldsymbol{\beta}$?
2. How would you estimate the asymptotic variance-covariance of your estimate?

* **Ex. 4.33** **Expectations of Sums** Suppose that x are *i.i.d.* draws from some distribution whose i^{th} (uncentered) moment is α_i. Let $\hat{\alpha}_i$ be the i^{th} sample moment, $\hat{\alpha}_i =_{\text{df}} \frac{1}{n}\sum_{1\le j\le n} {x_j}^i$. (Note that $\mu = \alpha_1$ and $\hat{\mu} = \hat{\alpha}_1$.) By definition $\mathrm{E}\left[\hat{\alpha}_i\right] = \alpha_i$, but consider the expectation of products of these sample sums:

1. Show that
$$\mathrm{E}\left[\hat{\alpha}_1^2\right] = \frac{1}{n}\alpha_2 + \frac{n-1}{n}{\alpha_1}^2.$$
(*Hint:* express $\sum_i x_i \sum_j x_j$ as $\sum_i {x_i}^2 + \sum_{i\ne j} x_i x_j$, then use independence of the x_i.)
Using this result, show that
$$\mathrm{E}\left[\hat{\alpha}_2 - \hat{\alpha}_1^2\right] = \frac{n-1}{n}\alpha_2 - \frac{n-1}{n}{\alpha_1}^2,$$
and thus $\mathrm{E}\left[\hat{\sigma}^2\right] = \sigma^2$.
2. Show that
$$\mathrm{E}\left[\hat{\alpha}_2\hat{\alpha}_1\right] = \frac{1}{n}\alpha_3 + \frac{n-1}{n}\alpha_1\alpha_2.$$
3. Assume that $n \ge 2$. Show that
$$\mathrm{E}\left[\hat{\alpha}_1^3\right] = \frac{1}{n^2}\alpha_3 + 3\frac{n-1}{n^2}\alpha_2\alpha_1 + \frac{(n-1)(n-2)}{n^2}{\alpha_1}^3.$$
4. Assume that $n \ge 2$. Using the previous two results show that
$$\begin{aligned}\mathrm{E}\left[\hat{\alpha}_2\hat{\alpha}_1 - \hat{\alpha}_1^3\right] &= \frac{n-1}{n^2}\left(\alpha_3 - 3\alpha_2\alpha_1 + {\alpha_1}^3\right) + \frac{n-1}{n}\left(\alpha_2\alpha_1 - {\alpha_3}^3\right),\\ &= \frac{n-1}{n^2}\mu_3 + \frac{n-1}{n}\mu\sigma^2,\end{aligned} \tag{4.77}$$
where $\mu_3 = \mathrm{E}\left[(x-\mu)^3\right]$. (*Hint:* expand the centered μ_3 in terms of the raw moments α_i.)

Ex. 4.34 **Bias and Variance of Sharpe ratio for Gaussian Returns**
Consider the case of Gaussian returns:

1. Use Equation 4.37 to show that the geometric bias of the Sharpe ratio for Gaussian returns is approximately $1 + \frac{3}{4n} + \frac{49}{32n^2}$. (*cf.* Approximation 3.14.)
2. Use Equation 4.38 to derive an expression for the standard error of the Sharpe ratio for Gaussian returns.

** **Ex. 4.35 Nearly Unbiased Estimation of signal-noise ratio?** Try to build a 'better' estimator of the signal-noise ratio. Consider an estimator of the form

$$a_0 + \frac{a_1 + (1 + a_2)\hat{\mu} + a_3\hat{\mu}^2}{\hat{\sigma}},$$

for constants a_i to be determined. Note that the usual Sharpe ratio corresponds to $a_0 = a_1 = a_2 = a_3 = 0$.

1. Use the Taylor's expansion $(1 + x)^{-1/2} \approx 1 - \frac{x}{2}$ to approximate this estimator in terms of the uncentered sample moments $\hat{\alpha}_i$.
2. From this approximation, derive the approximate expected value and variance of this estimator in terms of the moments and cumulants of the returns distribution, and the coefficients a_i.
3. Compute the expected squared error of the estimator, again as a function of the moments and cumulants, and a_i.
4. Take the partial derivative of the mean squared error with respect to each a_i and show that the squared error is minimized for $a_0 = a_1 = a_2 = a_3 = 0$, *i.e.*, the vanilla estimator.

Ex. 4.36 Asymptotic distribution of Sharpe ratio, Elliptical returns

1. From Equation 4.27 and Equation 4.26, derive Equation 4.29.

Ex. 4.37 Correlation of Sharpe ratios, Uncorrelated returns
We claim in Equation 4.29 that the Sharpe ratios of *uncorrelated* elliptical returns streams can be correlated for $\kappa \neq 1$.

1. Confirm this numerically, for two assets whose returns follow a multivariate t distribution.
2. How can uncorrelated returns have correlated Sharpe ratios?

Ex. 4.38 Markets Following U.S. Elections It has been noted that, since World War II, the U.S. stock market has experienced higher returns following midterm elections than following presidential elections. Consider this hypothesis: that the signal-noise ratio of the Market is higher post-midterms than post-presidential elections. For the purposes of this exercise, define 'post-midterm' as the 12 month period starting in November of even years which are not divisible by four (*e.g.*, 2018), and define 'post-presidential elections' as the 12 month period starting in November of even years which *are* divisible by four (*e.g.*, 2016).
Use the monthly returns of the Fama-French factor data from **aqfb.data**, using code as given in Example 1.2.1.

1. Compute the Sharpe ratio of the Market for the two periods, along with confidence intervals using Mertens' correction.
2. Perform the frequentist hypothesis test for equality of signal-noise ratio of the two periods.
3. Perform a *paired* test on the difference in signal-noise ratio of the two periods, pairing midterm periods with the following presidential periods. Do you results change if you pair midterm periods with the *preceding* presidential periods?

* **Ex. 4.39 Research Problem: Omitted Variable Bias** §
Omitted variable bias is like the weather: everyone talks about it, but nobody does anything about it. For this open research problem, prove that omitted variable bias is not too large, subject to some reasonable conditions on the omitted variable. For example, one might reasonably assume that the ommitted variable is only slowly varying over time (has a low autocorrelation); based on this and the autocorrelation of observed returns, one should be able to bound the size of the omitted variable bias.

** **Ex. 4.40 Research Problem: Concentration Inequalities** §
In this chapter we have attempted to show that the standard error of the Sharpe ratio is fairly robust to violations of the assumptions of *i.i.d.* normal returns. However, it is not clear that stable standard errors are sufficient for inference beyond the $\alpha = 0.05$ level.

For this open research problem, prove some kind of concentration inequality for the Sharpe ratio: assume that returns are drawn from some distribution with bounded kurtosis, say, then find a nice bound on the probability that the Sharpe ratio exceeds some large value.

5. Overoptimism

The first principle is that you must not fool yourself and you are the easiest person to fool.

(Richard Feynman)

Everybody dies
Frustrated and sad
And that is beautiful

(They Might Be Giants, Don't Let's Start*)*

Suppose you will enter a coin-flipping contest: whoever flips the most Heads wins. You have at your disposal a chest full of seemingly identical coins of unknown provenance. To prepare you select one coin, flip it some number of times, noting the proportion of Heads. You repeat the process with another and another, finally selecting, through some procedure, one "lucky coin" to take to the contest. You naïvely estimate the probability of landing Heads by the in-sample proportion.

The difference between the estimated quality of the lucky coin based on historical data and the actual probability of landing Heads is the "overfit bias" of this experiment. If all the coins in the box are identical and fair, unbeknownst to you, the overfit bias is likely to be very large, *i.e.*, you have seriously overestimated the probability of landing Heads. If, on the other hand, there is a large variation in the coins, perhaps some of them *almost always* land Heads, say, then the overfit bias is probably smaller.

Constructing and testing trading strategies is, perhaps uncharitably, analagous to selecting lucky coins, but with the following differences:

1. While you could flip a given coin more times to collect more data, often a quantitative strategist is stuck with a fixed amount of historical data, and can only collect more data at a rate of one day per day.
2. Presumably tests of different coins have errors independent of each other, while the quant is usually observing historical returns of multiple strategies which are dependent on each other at some (backtested) point in time.
3. Coins are just coins. They lack parameters. Often a quant is refining a model with free parameters, either by an optimization procedure, brute force, or sequential knob twiddling. Thus the simulated historical returns of the strategies are not only dependent, they are dependent in a way driven by these parameters. Moreover, the model can be too complex for the data, which can deterioriate actual performance.

4. In our thought experiment, the coins appear identical and are selected at random. Often trading strategies are passed along by word of mouth, or are discovered via the media (an article, book, white-paper, blog, *etc.*), and the quant does not know how much overfitting was involved in the original 'discovery' of the strategy, and there is likely to be very little historical 'out of sample' data.
5. A coin can be physically inspected for bias, while quantitative strategies often leak very little information that can be tested independently of returns.
6. Coin flips have an unambiguous outcome. Quantitative strategies are often backtested, and so are subject to the biases and imprecision of simulated fill prices, market impact, and market reaction. The 'ground truth' of backtested strategies is often uncertain.

Some of these issues can be approached statistically, and we will attempt to address them here, while others are beyond the scope of this text.

Remark (Overoptimism and Overfit). Though the definitions are a little nebulous, we will use *overoptimism* to refer to the case where one has a (positively) biased estimate of the "performance" of some model caused by selection among many competing models or fitting of parameters. For example, if one selects among many trading strategies based on the observed Sharpe ratios, then the Sharpe ratio of the best strategy will be biased upwards by this selection, and may not be a good estimate of its signal-noise ratio.

We will use the term *overfitting* to describe the case where by selecting a too complex model one causes a decrease in the performance. In this case there is still a bias in the estimated performance, as otherwise one presumably would pick a better model. An example might be using more and more factors to forecast the returns of some asset. The Sharpe ratio will increase with more factors, though likely the signal-noise ratio is decreasing.

We illustrate these with Figure 5.1, where we plot overoptimism and overfit as the signal-noise ratio, the Sharpe ratio, and the bias between them versus some undefined "effort," which could be number of strategies tested, or amount of complexity added to the trading strategy, *etc.*

In this chapter we will consider overoptimism, and will consider overfitting in the sequel.

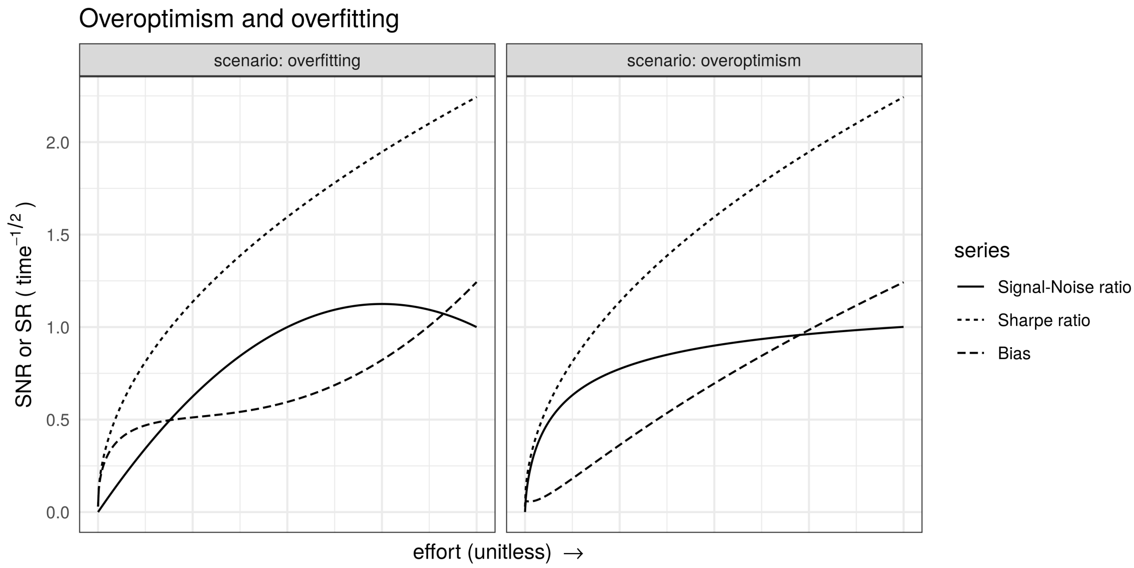

Figure 5.1.: Overoptimism and overfitting are illustrated as the signal-noise ratio, Sharpe ratio and their difference versus some "effort". In the case of overfitting, the signal-noise ratio peaks and increased effort contributes to declining out-of-sample performance. In the case of overoptimism, the signal-noise ratio continues to grow with increased effort, albeit slowly.

5.1. Overoptimism by selection

Here we consider perhaps the simplest model of strategy development, which we call *overoptimism by selection.* Suppose you observe returns from k different assets (or "strategies") each over n independent periods. The returns among assets may be correlated, but in the simplest realization we assume they are independent. We compute the Sharpe ratios of each asset then select the one with the highest Sharpe ratio, which presumably we will trade out-of-sample. The gap between the Sharpe ratio and signal-noise ratio of the chosen asset is the overoptimism.

Aronson presents a very accessible description of essentially this problem, calling our overoptimism, "data-mining bias." [7, chapter 6] He cites five factors affecting overoptimism, namely

1. the number of strategies tested, k,
2. the length of backtests, n,
3. the correlation of returns among strategies,
4. the presence of 'outliers',
5. the spread in the population expected returns.

All five factors are relevant in the context considered by Aronson, where strategies are apparently selected by maximum sample mean return. However, by using the Sharpe ratio to measure quality of a strategy we are somewhat unaffected by outliers, though they might be a sign of backtest biases. Also we have put a fair amount of work in the preceeding chapters to understand the error of the Sharpe ratio with respect to n, and can adjust for sample size as needed. The correlation of returns and the spread in population signal-noise ratios are the two relevant, unobservable factors that we must consider in our analyses.

Example 5.1.1 (Overoptimism by selection). We create a population of k strategies, with normally distributed signal-noise ratios:

$$\zeta_i \sim \mathcal{N}\left(0, 0.4032\mathrm{yr}^{-1}\right).$$

Then for each strategy we sample 504 days (2 years) of normally distributed returns with the given signal-noise ratio. The strategies' returns are generated independently. We compute the k Sharpe ratios, then select the strategy with the highest. We record its Sharpe ratio and signal-noise ratio. We also record the highest signal-noise ratio in the population. We let k vary from 1 to 10000. We repeat this experiment 20000 times, and compute empirical averages.

In Figure 5.2, we plot the Sharpe ratio, signal-noise ratio and overoptimism as a function of k. We see that the signal-noise ratio of the selected strategy appears to be growing in k, albeit very weakly. Thus the extra effort associated with testing more strategies appears to pay off in this case. In Figure 5.3, we again plot the signal-noise ratio of the selected strategy, but also the maximum signal-noise ratio of all k strategies. The difference between them is a kind of "regret", which we also plot. Note that the regret grows with k, which seems to be an inescapable curse of decision making with limited information. For $k = 10000$ the expected signal-noise ratio of the selected strategy is around 2.6 standard deviations above the mean, while the population maximum is around 3.8 standard deviations.

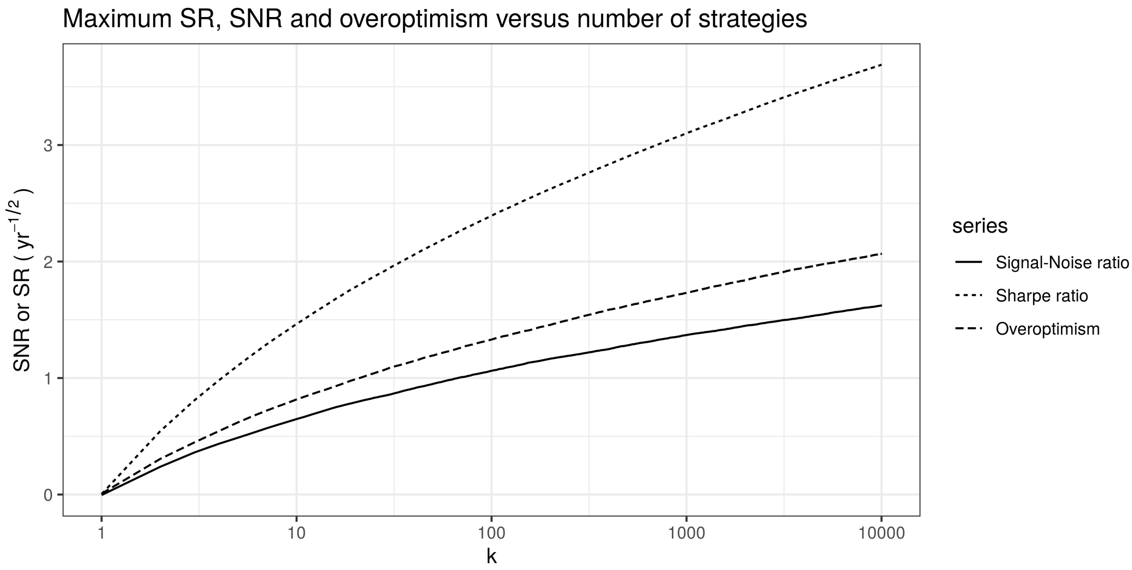

Figure 5.2.: The maximum Sharpe ratio and the associated signal-noise ratio of the selected strategy is plotted versus k for the case of overoptimism by selection with normally distributed signal-noise ratio. The x axis is in log scale.

⊣

Does fitting work? Aronson claims, citing White, that the selection procedure outlined here "works" in the sense that as $n \to \infty$, it selects the optimal model almost surely. [7, 167] This intuitively makes sense since the error of the Sharpe ratio around the signal-noise ratio goes to zero in large n. Under the (admittedly unrealistic) assumption of independent returns and non-zero spread in signal-noise ratios, the procedure also works in the other asymptotic direction. That is, as k increases the signal-noise ratio of the strategy selected for having maximal Sharpe ratio is increasing, albeit potentially very slowly in k.

Example 5.1.2 (Overoptimism by selection (again)). Continuing Example 5.1.1, since $\hat{\zeta}_i$ is approximately normally distributed around ζ_i, we can make the approximation

$$\left[\begin{array}{c}\zeta_i\\ \hat{\zeta}_i\end{array}\right] \sim \mathcal{N}\left(\left[\begin{array}{c}0\\0\end{array}\right], \left[\begin{array}{cc}\sigma_1^2 & \sigma_1^2\\ \sigma_1^2 & \sigma_1^2+\sigma_2^2\end{array}\right]\right),$$

where σ_1^2 is the variance in the generation of ζ_i, and σ_2^2 is approximated as n^{-1} via Equation 3.24. Then conditional on observing $\hat{\zeta}_i$ we have

$$\zeta_i \left| \hat{\zeta}_i \sim \mathcal{N}\left(\rho^2\hat{\zeta}_i, \left(1-\rho^2\right)\sigma_1^2\right),\right.$$

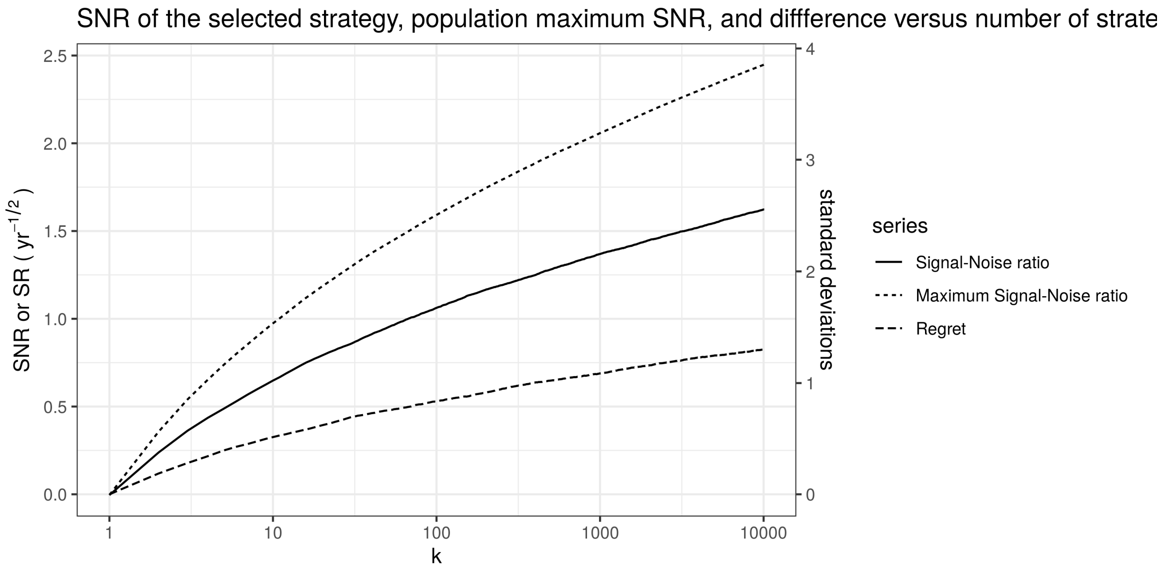

Figure 5.3.: The signal-noise ratio of the selected strategy and the population maximum signal-noise ratio, and their difference are plotted versus k for the case of overoptimism by selection. The x axis is in log scale. The right side expresses the signal-noise ratio in standard deviations above the mean, based on a standard devation of $0.04\text{day}^{-1/2}$.

where $\rho = \sigma_1/\sqrt{\sigma_1^2 + \sigma_2^2}$. That is, we expect the signal-noise ratio associated with the asset with maximal Sharpe ratio to be ρ^2 times that maximum Sharpe ratio. [168] If

$$\frac{\rho^2 \hat{\zeta}_i}{\sigma_1\sqrt{1-\rho^2}} = \frac{\rho\hat{\zeta}_i}{\sigma_2} \gg 2,$$

then we expect that ζ_i is likely to be bigger than zero. The ratio ρ/σ_2 is something like a 'signal-noise ratio of signal-noise ratios' in this example.

In our example we have $\sigma_1^2 = 0.4032\text{yr}^{-1}$, and $\sigma_2^2 \approx \frac{1}{n} = 0.5\text{yr}^{-1}$, and therefore $\rho = 0.6681$. In this case we have $\rho/\sigma_2 \approx 0.9449\text{yr}^{1/2}$, and we are unlikely to select a strategy with negative signal-noise ratio if $\hat{\zeta}_i \gg 2$. We plot the empirical expected value of the ζ associated with the strategy with maximal Sharpe ratio in Figure 5.4, along with ρ^2 times the empirical expected maximum Sharpe ratio. We find these lines to be in good agreement.

We note that in real use the spread of the ζ_i, which we have called σ_1, is unknown. However, one could estimate it from the sample. That is, observing many $\hat{\zeta}_i$, one could compute their sample variance, then subtract σ_2^2 to get an estimate for σ_1^2. It is not clear what one should do in the case this estimate is negative, however.

⊣

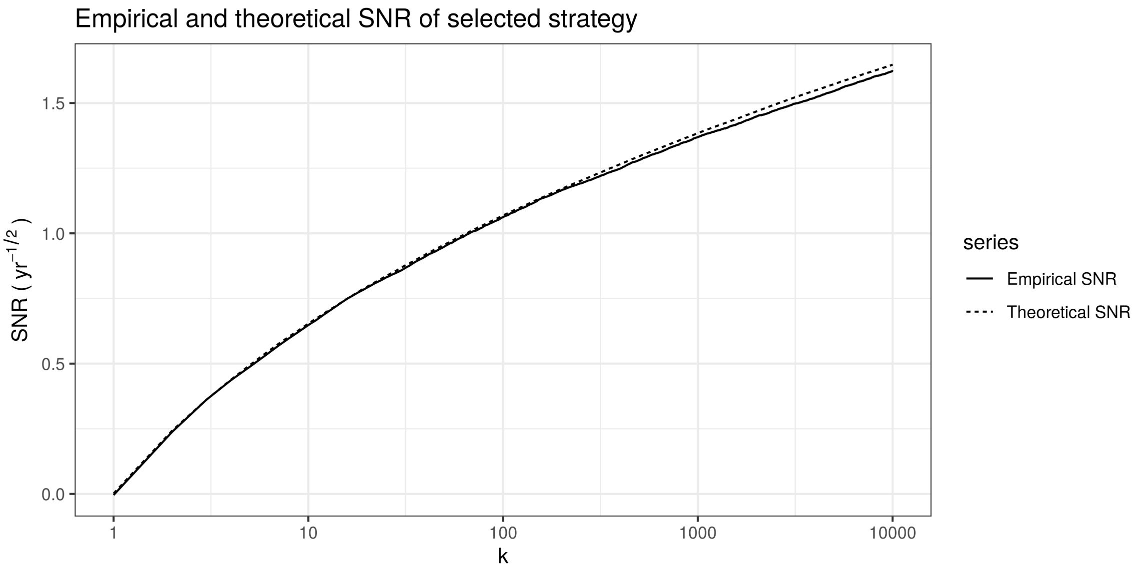

Figure 5.4.: The signal-noise ratio of the selected strategy, and its approximate expected value are plotted against k for the case of overoptimism by selection on normally distributed signal-noise ratio.

5.1.1. Multiple Hypothesis Testing

The most basic approach to overoptimism is to appeal to traditional corrections for *multiple hypothesis testing*, or MHT. [22, 167, 66, 100] So suppose we have observed Sharpe ratios $\hat{\zeta}_1, \hat{\zeta}_2, \ldots, \hat{\zeta}_k$. Consider the reindexing of the assets in order of their observed Sharpe ratio. That is, consider the permutation of $1, 2, \ldots, k$, denoted $(1), (2), \ldots, (k)$ such that

$$\hat{\zeta}_{(1)} \leq \hat{\zeta}_{(2)} \leq \ldots \hat{\zeta}_{(k-1)} \leq \hat{\zeta}_{(k)}.$$

We will use the same indexing on the signal-noise ratios, thus $\zeta_{(k)}$ will refer to the signal-noise ratio associated with the strategy that has maximum Sharpe ratio, and not the maximum of the ζ_i.

To test the null hypothesis

$$H_0 : \zeta_{(k)} = \zeta_0 \quad \text{versus} \quad H_1 : \zeta_{(k)} > \zeta_0,$$

one can instead perform a *Union Intersection Test.* In contrast to a Intersection Union test (*cf.* Section 3.5.4), here the null hypothesis is the intersection of hypotheses, and the alternative hypothesis (and the critical region, where one rejects) is a union. That is, to test the hypothesis above, one instead tests

$$H_0 : \forall i\, \zeta_i = \zeta_0 \quad \text{versus} \quad H_1 : \exists i : \zeta_i > \zeta_0. \tag{5.1}$$

To test this hypothesis via the *Bonferroni correction* at the α level, perform the k separate hypothesis tests for $\zeta_i = \zeta_0$, but at the α/k level, and reject H_0 if any of these reject. [166, 66, 22]

The Bonferroni correction is equivalent to the observation that one can attribute at most α/k probability of a type I error to each of the separate tests and arrive at a total type I rate of no more than α. A slightly more powerful correction is possible under the (generally suspect) assumption that returns of the strategies are independent. In this case, the $\hat{\zeta}_i$ are then also independent, as are the outcomes of the k separate hypothesis tests. In this case, one can use the *Šidák correction*, where one tests $\zeta_i = \zeta_0$ at the $1 - (1 - \alpha)^{1/k}$ level, rejecting the intersection H_0 if you reject for any ζ_i. The Šidák correction is equivalent to finding the probability that we commit a type I error for *none* of the separate sub-tests, and setting it to $1 - \alpha$, then using independence.

Note that in our case the null hypothesis posits the same value for each ζ_i. Moreover, because we observe n returns for each strategy, the distribution of each $\hat{\zeta}_i$ is identical under the null. Thus to use the Bonferroni or Šidák corrections we need not test every ζ_i, rather we need only test $\zeta_{(k)}$. Thus to test the null hypothesis above using the Bonferroni correction, reject if $\hat{\zeta}_{(k)} > SR_{1-\alpha/k}\left(\zeta_0, n\right)$, where $SR_p\left(\zeta, n\right)$ is the p^{th} quantile of the Sharpe ratio distribution with signal-noise ratio ζ on n samples; under the Šidák correction reject if $\hat{\zeta}_{(k)} > SR_{(1-\alpha)^{1/k}}\left(\zeta_0, n\right)$.

Another way to arrive at the Šidák correction is via the beta distribution: if $p_1, p_2, \ldots, p_k$ are independent random variables each uniform on $[0, 1]$, then the j^{th} largest of them, call it $p_{(j)}$ follows a beta distribution:

$$p_{(j)} \sim \mathcal{B}\left(j, k + 1 - j\right).$$

In particular, the smallest is distributed as $\mathcal{B}\left(1, k\right)$. Thus to test the intersection null hypothesis $\forall i\, \zeta_i = \zeta_0$ at the α level under the assumption of independence, one should compute the p-values of each of the hypotheses $\zeta_i = \zeta_0$, then compare the smallest to the α quantile of the beta distribution $\mathcal{B}\left(1, k\right)$. The smallest p-value will correspond to the largest Sharpe ratio, and the α quantile is $\beta_\alpha\left(1, k\right) = 1 - \left(1 - \alpha\right)^{1/k}$, which is the Šidák correction.

In practice the nominal gain in power from using the Šidák correction is negligible, as illustrated in Example 5.1.3. For large k, the Šidák correction offers little additional power for the assumption of independence and one typically sees the Bonferroni correction used instead.

Example 5.1.3 (Sharpe ratio, Bonferroni and Šidák). In Table 5.1, we display cutoffs computed by the Bonferroni and Šidák corrections for the case of testing the null hypothesis at the 0.05 level that all signal-noise ratios are equal to zero, for the case of $n = 504\text{day}$. The cutoffs are in annualized terms and are quantiles of the Sharpe ratio distribution. Note how little difference there is between the two corrections.

k	Bonferroni cutoff	Sidak cutoff
1	1.165	1.165
10	1.828	1.823
100	2.341	2.335
1000	2.773	2.769
10000	3.156	3.152
100000	3.502	3.499

Table 5.1.: The Bonferroni and Šidák corrected cutoff values are presented for increasing k for the case of testing the hypothesis that all strategies have zero signal-noise ratio. The cutoffs are for $\alpha = 0.05$ and $n = 504$, and have units $\mathrm{yr}^{-1/2}$.

$\dashv$

5.1.2. Multiple Hypothesis Testing, Correlated Returns

A glaring problem with the simple corrections for multiple hypothesis testing are that they do not take into account possible correlation of strategies' returns. The Bonferroni correction maintains an *upper bound* on the type I error, but is too conservative in the case where returns are mostly positively correlated. (And it is impossible to have many mutually negatively correlated returns, *cf.* Exercise 5.3.)

We illustrate this problem in Example 5.1.4, where we empirically estimate the mean Sharpe ratio of mutually correlated returns. In the presence of large positive mutual correlation of returns, the type I rate will be much lower than the nominal rate. Equivalently, the Bonferroni correction will have low power, since a large signal-noise ratio will have to overcome the 'penalty' paid by the correlation. Effectively, the correlation structure has reduced the number of independent hypothesis tests being performed.

Example 5.1.4 (Overoptimism by selection, correlated returns). We draw k-variate daily returns from a normal distribution with zero mean and covariance

$$(1 - \rho)\,\mathsf{I} + \rho\left(\mathbf{1}\mathbf{1}^{\top}\right)$$

Returns are drawn independently for each of 504 days. We compute the Sharpe ratio of each of the k columns and then compute the maximum. We repeat this 10000 times, and compute the empirical 0.95 quantile of the Sharpe ratios. We let k vary from 1 to 100. We repeat for several values of ρ.

In Figure 5.5, we plot the empirical quantiles of the maximum Sharpe ratio versus k. The line for $\rho = 0$ should correspond approximately to the Šidák column from Table 5.1, which is the cutoff for the null hypothesis that all strategies have signal-noise ratio equal to zero. When ρ is large, the maximum Sharpe ratio is very far from the 0.95 cutoff, meaning the rate of type I errors is much smaller than 0.05. *cf.* Exercise 5.4.

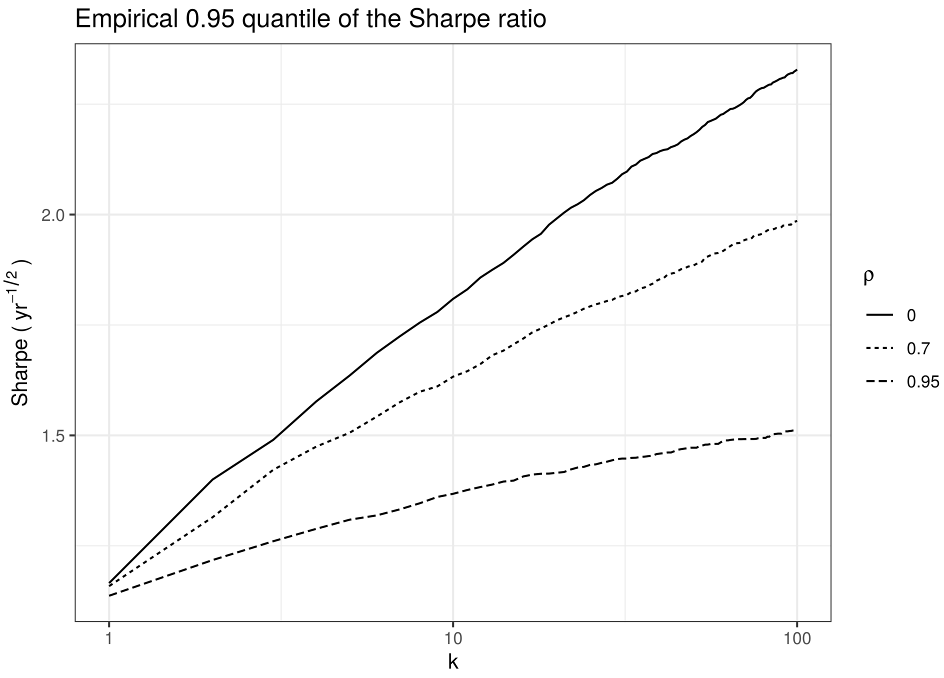

Figure 5.5.: The empirical 0.95 quantile of the maximum Sharpe ratio of k strategies is plotted versus k for returns drawn from a normal distribution with zero mean and covariance $(1-\rho)\,\mathsf{I} + \rho\left(\mathbf{1}\mathbf{1}^{\top}\right)$. Quantiles are computed over 10,000 simulations of 504 days of returns.

⊣

To deal with correlation among returns, consider the normal approximation to the vector of Sharpe ratios,

$$\hat{\boldsymbol{\zeta}} \approx \mathcal{N}\left(\boldsymbol{\zeta}, \frac{1}{n}\Omega\right), \tag{5.2}$$

where Ω will not, in general, be diagonal. To test the null hypothesis that $\boldsymbol{\zeta} = \boldsymbol{\zeta}_0$, convert the above approximation to a standard normal by an inverse square root of Ω:

$$\sqrt{n}\Omega^{-1/2}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right) \approx \mathcal{N}\left(\mathbf{0}, \mathsf{I}\right). \tag{5.3}$$

There are many ways we could test this hypothesis, but we are interested in a one-sided alternative, and wish to focus on the asset which demonstrated maximal Sharpe ratio. Thus a general chi-square test, as suggested in Section 4.2.1, seems inappropriate. Moreover, to apply the technique to the case of very large k, we wish to avoid having to computationally invert $\Omega^{1/2}$.

If each period's returns are elliptical with kurtosis factor κ, and R is the correlation of returns then, recalling Equation 4.29,

$$\Omega = \mathsf{R} + \frac{\kappa - 1}{4}\boldsymbol{\zeta}\boldsymbol{\zeta}^{\top} + \frac{\kappa}{2}\operatorname{Diag}\left(\boldsymbol{\zeta}\right)\left(\mathsf{R}\odot\mathsf{R}\right)\operatorname{Diag}\left(\boldsymbol{\zeta}\right).$$

To tame the computations, consider the simple *rank one updated correlation model*, where we assume that returns have correlation

$$\mathsf{R} = \left(1-\rho\right)\mathsf{I} + \rho\left(\mathbf{1}\mathbf{1}^{\top}\right), \tag{5.4}$$

for $|\rho| < 1$. For this R, under the null hypothesis that $\boldsymbol{\zeta} = \zeta_0\mathbf{1}$,

$$\begin{aligned}
\Omega &= \left(1-\rho\right)\mathsf{I} + \rho\left(\mathbf{1}\mathbf{1}^{\top}\right) + \frac{\kappa-1}{4}\zeta_0^2\mathbf{1}\mathbf{1}^{\top} + \frac{\kappa}{2}\zeta_0^2\left(\left(1-\rho^2\right)\mathsf{I} + \rho^2\left(\mathbf{1}\mathbf{1}^{\top}\right)\right),\\
&= \left[1-\rho+\frac{\kappa}{2}\zeta_0^2\left(1-\rho^2\right)\right]\mathsf{I} + \left[\rho + \frac{\kappa-1}{4}\zeta_0^2 + \frac{\kappa}{2}\zeta_0^2\rho^2\right]\left(\mathbf{1}\mathbf{1}^{\top}\right),\\
&= a_0\,\mathsf{I} + a_2\left(\mathbf{1}\mathbf{1}^{\top}\right).
\end{aligned} \tag{5.5}$$

It then follows (*cf.* Exercise 5.6) that

$$\begin{aligned}
\Omega^{-1/2} &= a_0^{-1/2}\,\mathsf{I} - \left[\frac{\sqrt{a_0 + a_2\mathbf{1}^{\top}\mathbf{1}} - \sqrt{a_0}}{\sqrt{a_0^2 + a_0 a_2\mathbf{1}^{\top}\mathbf{1}}}\right]\frac{\left(\mathbf{1}\mathbf{1}^{\top}\right)}{\mathbf{1}^{\top}\mathbf{1}},\\
&= b_0\,\mathsf{I} + b_2\left(\mathbf{1}\mathbf{1}^{\top}\right).
\end{aligned} \tag{5.6}$$

Note that as a transform, this $\Omega^{-1/2}$ is order-preserving: if $y_i \le y_j$ and $\boldsymbol{z} = \Omega^{-1/2}\boldsymbol{y}$, then $z_i \le z_j$. In particular, it preserves the identity of the maximum element. Thus if the i^{th} element of $\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0$ is the largest, then it is also the largest element of $\sqrt{n}\Omega^{-1/2}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right)$. Note also that $\mathbf{1}/\mathbf{1}^{\top}\mathbf{1}$ computes the mean, thus we can view the transform of $\Omega^{-1/2}$ as a shrinkage to (or away) from the mean.

Thus to test the null in Hypothesis 5.1, let

$$\begin{aligned}
a_0 &= 1-\rho+\frac{\kappa}{2}\zeta_0^2\left(1-\rho^2\right),\\
a_2 &= \rho + \frac{\kappa-1}{4}\zeta_0^2 + \frac{\kappa}{2}\zeta_0^2\rho^2,\\
b_0 &= a_0^{-1/2},\\
b_2 &= \frac{1}{k}\frac{\sqrt{a_0} - \sqrt{a_0 + k a_2}}{\sqrt{a_0^2 + k a_0 a_2}},
\end{aligned}$$

then compute

$$z = \sqrt{n}\left[b_0\left(\hat{\zeta}_{(k)} - \zeta_0\right) + b_2\sum_{1\le i\le k}\left(\hat{\zeta}_i - \zeta_0\right)\right], \tag{5.7}$$

and reject at the α level if $z \geq z_{1-\alpha/k}$, the $1-\alpha/k$ quantile of the normal distribution. Of course, κ and ρ are unknown and have to be estimated from the sample. Note we can rewrite Equation 5.7 as

$$\begin{aligned} z &= \sqrt{n}\left[b_0\left(\hat{\zeta}_{(k)} - \zeta_0\right) + b_2 k\left(\bar{\zeta} - \zeta_0\right)\right],\\ &= \sqrt{n}\left[b_0\left(\hat{\zeta}_{(k)} - \bar{\zeta}\right) + \left(b_0 + b_2 k\right)\left(\bar{\zeta} - \zeta_0\right)\right], \end{aligned} \tag{5.8}$$

where $\bar{\zeta}$ is the average of the sample Sharpe ratios.

See Exercise 5.10 for a 'direct' approach to testing Hypothesis 5.1 that does not require one to compute the sample mean, but which only has approximate type I rate.

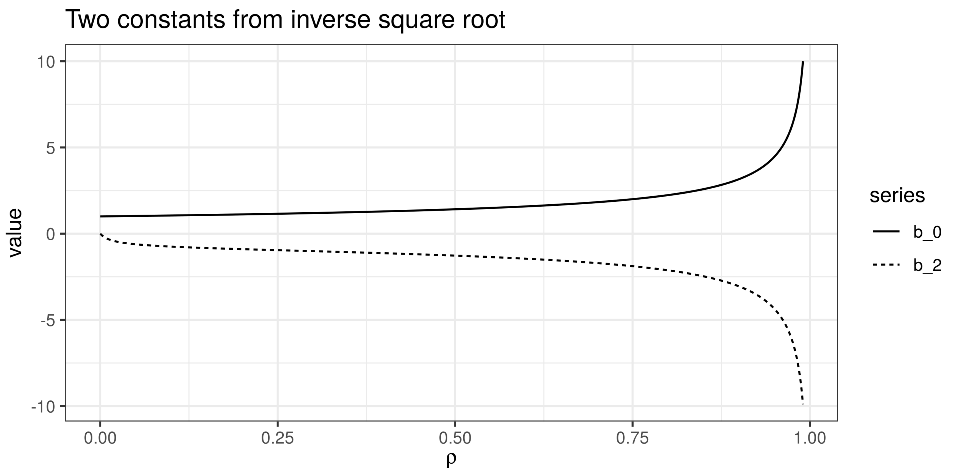

Figure 5.6.: Letting $\Omega^{-1/2} = b_0\,\mathsf{I} + b_2\left(\mathbf{1}\mathbf{1}^{\top}\right)$ for $\Omega = a_0\,\mathsf{I} + a_2\left(\mathbf{1}\mathbf{1}^{\top}\right)$, we plot b_0, b_2 versus ρ for $k = 100$.

In Figure 5.6, we plot b_0, b_2 versus ρ where $\Omega^{-1/2} = b_0\,\mathsf{I} + b_2\left(\mathbf{1}\mathbf{1}^{\top}\right)$, for the case where $k = 100, \kappa = 1$, and $\zeta_0 = 0$. We note that when $\rho \approx 1$ the values of b_0 and b_2 are highly sensitive to ρ. Moreover, as $\rho \to 1$, the procedure would appear to compute z as a large multiple times the difference between $\hat{\zeta}_{(k)}$ and the mean of the $\hat{\zeta}_i$, meaning it would *always* reject for ρ sufficiently close to 1. Noting that ρ has to be estimated from the sample, to keep the test conservative one should bound one's estimate of ρ away from 1. Moreover, as we shall see below, the maximum type I rate is maintained by assuming a smaller ρ, at the expense of power.

Example 5.1.5 (MHT Correction, Correlated Returns). We perform simulations under the null hypothesis, $\zeta = \mathbf{0}$, with $\mathsf{R} = (1-\rho)\,\mathsf{I} + \rho\left(\mathbf{1}\mathbf{1}^{\top}\right)$. We set $k = 100$, and simulate $n = 1008$ days of returns, compute the maximum Sharpe ratio, use a Bonferroni correction to test the null hypothesis, and tabulate the empirical rejection rate at the nominal 0.05 level over 5000 simulations. We allow the correlation correction method to use the population ρ, rather than estimate it from the data.

In Figure 5.7, we plot the empirical rejection rate versus ρ at the nominal 0.05 type I level. The vanilla MHT test is conservative, with near zero rejection rates for large ρ, while the correlation correction yields nominal rejection rates.

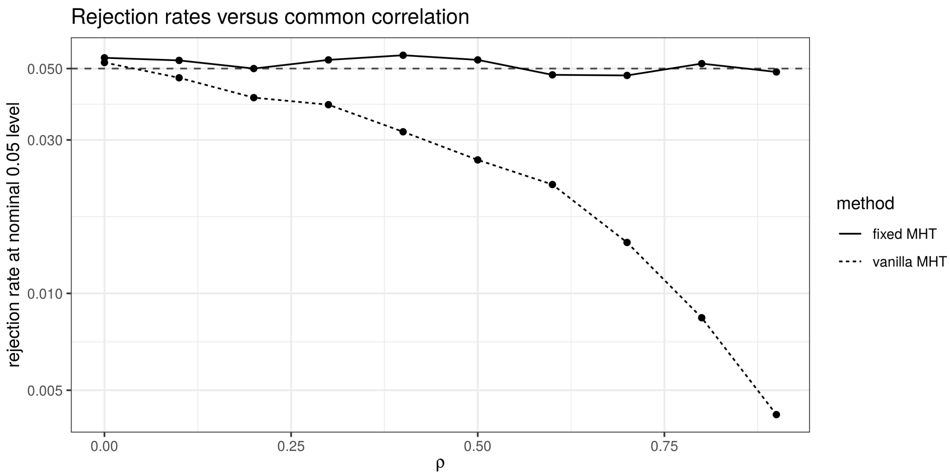

Figure 5.7.: The empirical type I rate under the null hypothesis is plotted versus ρ for the case where $\mathsf{R} = (1-\rho)\,\mathsf{I} + \rho\left(\mathbf{1}\mathbf{1}^{\top}\right)$, for the vanilla Bonferroni correction, and Bonferroni correction with fix for common correlation. Tests are performed with Gaussian returns, for 100 assets over 1008 days. Tests were performed at the 0.05 level, which appears to be maintained by the fixed Bonferroni procedure but not by the regular Bonferroni procedure. Empirical rates are over 5,000 simulations. The y axis is in log scale.

⊣

Alternative Correlation Models Inasmuch as we would like to apply the same analysis to a broader class of correlation matrices, they do not lead to an order-preserving $\Omega^{-1/2}$. Thus it is hard to test $\boldsymbol{z} = \sqrt{n}\Omega^{-1/2}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right)$ via Bonferroni correction, as the largest element cannot easily be identified.

However, we can use the analysis above in the case where R consists of non-negative elements to arrive at a test with bounded type I rate. Consider multivariate normal vectors $\boldsymbol{x}, \boldsymbol{y}$ with $\boldsymbol{x} \sim \mathcal{N}\left(\mathbf{0}, \mathsf{R}\right)$ and $\boldsymbol{y} \sim \mathcal{N}\left(\mathbf{0}, (1-\rho_0)\,\mathsf{I} + \rho_0\left(\mathbf{1}\mathbf{1}^{\top}\right)\right)$, $\mathsf{R}_{i,j} \ge \rho_0 \ge 0$ for all $i \ne j$. By Slepian's lemma, $\Pr\left\{\max_i x_i \ge t\right\} \le \Pr\left\{\max_i y_i \ge t\right\}$ for all t. [153, 179]

Thus we can appeal to analysis above by assuming the correlation matrix has the form $(1-\rho_0)\,\mathsf{I} + \rho_0\left(\mathbf{1}\mathbf{1}^{\top}\right)$ and accepting a biased test, that is one with type I rate (approximately) no greater than α. Thus to test the null in Hypothesis 5.1, supose that $\mathsf{R}_{i,j} \ge \rho_0 \ge 0$ for $i \ne j$. Then compute z as in Equation 5.7, plugging in ρ_0 in the definitions of a_0 and a_2, and reject at the nominal α level if $z \ge z_{1-\alpha/k}$, the $1-\alpha/k$ quantile of the normal distribution. This procedure has (approximate) type I rate *no greater* than α. By "approximate" here, we allude to the normal approximation of Equation 4.29 and the approximate variance-covariance therein.

Example 5.1.6 (MHT Correction, AR(1) Correlated Returns). We perform simulations under the null hypothesis, $\zeta = \mathbf{0}$, where R takes the form of an $AR\left(1\right)$ matrix:

$$\mathsf{R}_{i,j} = \rho^{|i-j|},$$

for some ρ assumed known. We set $k = 100$, and simulate $n = 1008$ days of returns, compute the maximum Sharpe ratio, use a Bonferroni correction to test the null hypothesis, and tabulate the empirical rejection rate at the nominal 0.05 level over $10{,}000$ simulations. We allow the correlation correction method to use the population ρ, rather than estimate it from the data.

In Figure 5.8, we plot the empirical rejection rate versus ρ at the nominal 0.05 type I level. We apply the procedure for Bonferroni correction with correlation matrix $(1-\rho_0)\,\mathsf{I} + \rho_0\left(\mathbf{1}\mathbf{1}^{\top}\right)$ where $\rho_0 = \rho^{k-1}$. The procedure has access to ρ, and need not estimate it. We get near-nominal coverage up until around $\rho \approx 0.75$, after which the procedure is conservative.

See also Exercise 5.11. $\dashv$

The results of Example 5.1.6 indicate that when only a few non-diagonal elements of R are bigger than zero (*e.g.*, the AR(1) correlation matrix when ρ is not near 1), one can still get near-nominal coverage assuming that $\mathsf{R} = \mathsf{I}$. Slepian's lemma only goes one way, giving us a test with type I error at most a fixed value; however, there are known bounds on the other side. For example, Li and Shao prove the following: [97] Suppose that $\boldsymbol{x} \sim \mathcal{N}\left(\mathbf{0}, \mathsf{R}^0\right)$ and $\boldsymbol{y} \sim \mathcal{N}\left(\mathbf{0}, \mathsf{R}^1\right)$, where the diagonal elements of R^0 and R^1 are all one. Then

$$\begin{aligned}\Pr\left\{\max_i x_i > t\right\} \le \Pr\left\{\max_i y_i > t\right\} & \\ + \frac{1}{2\pi}\sum_{1\le i<j\le k}\left(\operatorname{asin} R^1_{i,j} - \operatorname{asin} R^0_{i,j}\right)^{+} \exp\left(-\frac{t^2}{1 + max\left(\left|R^0_{i,j}\right|, \left|R^1_{i,j}\right|\right)}\right).\end{aligned} \tag{5.9}$$

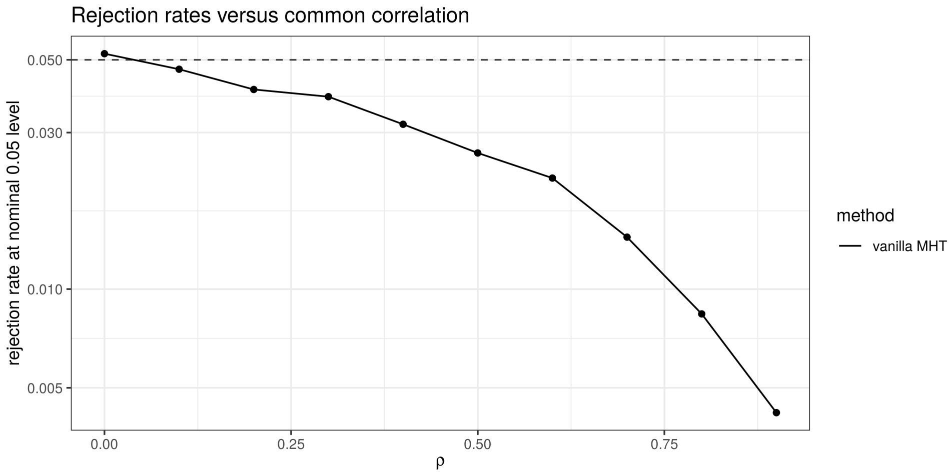

Figure 5.8.: The empirical type I rate under the null hypothesis is plotted versus ρ for the case where R is an $AR(1)$ matrix. That is, $\mathsf{R}_{i,j} = \rho^{|i-j|}$. The test applies a Bonferroni correction with fix for common correlation of ρ^{k-1}. We also plot rejection rates for the Bonferroni test with no correction for correlation. Tests are performed with Gaussian returns, for 100 assets over 1008 days. Tests were performed at the nominal 0.05 level, which appears to be maintained by the bounded correlation Bonferroni procedure for modest, but not large values of ρ. Empirical rates are over 10,000 simulations. The y axis is in log scale.

Thus if there are only $2m$ elements such that $R^1_{i,j} > R^0_{i,j}$ we can roughly transform this bound into

$$\Pr\left\{\max_i x_i > t\right\} \leq \Pr\left\{\max_i y_i > t\right\} + \frac{m}{2}e^{-t^2/2}. \tag{5.10}$$

See also Exercise 5.12.

5.1.3. One-Sided Alternatives

Another approach to the problem of overoptimism by selection is via testing against (multivariate) one-sided alternatives. As with the Bonferroni correction, in this approach we do not test the signal-noise ratio of the asset with maximal Sharpe ratio, *i.e.*, $\zeta_{(k)}$, but instead test the population as a whole. That is, we wish to test the following hypothesis:

$$H_0 : \forall_i \zeta_i = \zeta_0 \quad \text{versus} \quad H_1 : \forall_i \zeta_i \geq \zeta_0 \text{ and } \exists_j \zeta_j > \zeta_0. \tag{5.11}$$

There has been considerable research on testing hypotheses of this form but when the parameter is the expected value of a multivariate normal. [151, 164, 132, 133, 158] By appealing to the normal approximation of Equation 4.29, we can use these tests for tackling overoptimism by selection. That is, assume

$$\hat{\boldsymbol{\zeta}} \approx \mathcal{N}\left(\boldsymbol{\zeta}, \frac{1}{n}\Omega\right), \tag{5.12}$$

where Ω is some function of R, κ and $\boldsymbol{\zeta}$, as in Equation 4.29.

The classical procedure for testing equality of $\boldsymbol{\zeta} = \boldsymbol{\zeta}_0$ in this case, without regard to the form of the alternatives, is via Hotelling's T^2 test. [5, 138] In our case since we have assumed a normal approximation, we are essentially assuming Ω is known, and so a χ^2 test would be used instead. The most basic test against a one-sided alternative is similar in nature, but involves projection to the positive orthant.

We first present the general solution, which is probably only of use when k is small and when Ω can be inverted; then we will consider the large k case for the simple rank-one model of correlation of Equation 5.4. The general case proceeds by computing

$$\bar{x}^2 = n\left[\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right)^\top \Omega^{-1}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right) - \min_{\boldsymbol{x} \geq \boldsymbol{\zeta}_0}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{x}\right)^\top \Omega^{-1}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{x}\right)\right]. \tag{5.13}$$

Here $\boldsymbol{\zeta}_0 = \zeta_0 \mathbf{1}$, and $\boldsymbol{x} \geq \boldsymbol{\zeta}_0$ is taken to be element-wise. Compare Equation 5.13 to Equation 4.31 for testing essentially the same null against an unrestricted alternative.

Under the null hypothesis, $\bar{x}^2$ will follow not a chi-square distribution, but rather a "chi-bar-square distribution." To test Hypothesis 5.11, reject at the α level if

$$\sum_{0 \leq i \leq k} w_i\left(\Omega\right) F_{\chi^2}\left(\bar{x}^2; i\right) \geq 1 - \alpha, \tag{5.14}$$

where $F_{\chi^2}\left(x; i\right)$ is the cumulative distribution of the χ^2 distribution with i degrees of freedom, and $w_i\left(\Omega\right)$ are the *chi-bar-square weights.*

The $w_i\left(\Omega\right)$, which are a function of Ω, are tricky to describe analytically except for some simple cases of Ω: for example, when Ω is diagonal, we have

$$w_i\left(\Omega\right) = \binom{k}{i} 2^{-k}.$$

For non-diagonal Ω, the recommended course of action is to estimate the w_i via simulation. [151] The simulations proceed by repeating the following steps some large number of times (say $N = 10^4$):

1. Generate $\boldsymbol{z} \sim \mathcal{N}\left(\boldsymbol{\zeta}_0, \Omega\right)$.
2. Find the $\boldsymbol{x} \geq \boldsymbol{\zeta}_0$ that minimizes $\left(\boldsymbol{z} - \boldsymbol{x}\right)^\top \Omega^{-1}\left(\boldsymbol{z} - \boldsymbol{x}\right)$.
3. Count the number of elements of $\boldsymbol{x}$ which are strictly greater than ζ_0, call it s.

Estimate $w_i\left(\Omega\right)$ as the proportion of the N simulations where the s equals i.

Example 5.1.7 (Testing signal-noise ratios Fama-French factors). Consider the four Fama French factors from Example 1.2.1. Based on monthly returns from Jan 1927 to Dec 2018, the Sharpe ratios were computed as

$$\hat{\boldsymbol{\zeta}} = \begin{bmatrix} Mkt & SMB & HML & UMD \\ \hline 0.1728 & 0.0657 & 0.1056 & 0.1411 \end{bmatrix} \text{mo.}^{-1/2}.$$

First we test the null hypothesis $H_0 : \forall_i \zeta_i = \zeta_0$ for $\zeta_0 = 0.12\text{mo.}^{-1/2}$, against an unrestricted alternative. This we do via the χ^2 test of Equation 4.30.

We estimate the covariance of the Sharpe ratios (up to n) as

$$\hat{\Omega} \approx \begin{bmatrix} 1.0466 & 0.2912 & 0.1307 & -0.3042 \\ 0.2912 & 0.8970 & 0.0552 & -0.1814 \\ 0.1307 & 0.0552 & 0.8290 & -0.3792 \\ -0.3042 & -0.1814 & -0.3792 & 1.5811 \end{bmatrix} \text{mo.}^{-1}.$$

This is estimated by the procedure described in Section 4.2: namely we stack $\boldsymbol{x}$ with $\boldsymbol{x}^2$, compute the sample mean and covariance of this vector, then apply the delta method.

The test statistic is then computed as

$$x^2 = n\left(\hat{\boldsymbol{\zeta}} - \zeta_0 \mathbf{1}\right)^\top \hat{\Omega}^{-1} \left(\hat{\boldsymbol{\zeta}} - \zeta_0 \mathbf{1}\right) = 10.2.$$

Under the null hypothesis, this is asymptotically distributed as a $\chi^2(4)$, which corresponds to a p-value of of 0.037, and a Frequentist would reject the null hypothesis.

Now we test the null $H_0 : \forall_i \zeta_i = \zeta_0 = 0.12\text{mo.}^{-1/2}$ against the one-sided alternative $H_1 : \forall_i \zeta_i \geq \zeta_0$ and $\exists_j \zeta_j > \zeta_0$. First we estimate the chi-bar-square weights using 100,000 simulations using the estimated $\hat{\Omega}$. The weights are estimated as

$$\begin{bmatrix} w_0 & w_1 & w_2 & w_3 & w_4 \\ \hline 0.027 & 0.186 & 0.348 & 0.368 & 0.071 \end{bmatrix}.$$

The projection of $\hat{\boldsymbol{\zeta}}$ onto the alternative orthant is

$$\hat{\boldsymbol{\zeta}} = \begin{bmatrix} Mkt & SMB & HML & UMD \\ \hline 0.1919 & 0.1200 & 0.1200 & 0.1251 \end{bmatrix} \text{mo.}^{-1/2}.$$

We compute the statistic as

$$\begin{aligned} \bar{x}^2 &= n\left[\left(\hat{\boldsymbol{\zeta}} - \zeta_0 \mathbf{1}\right)^\top \Omega^{-1} \left(\hat{\boldsymbol{\zeta}} - \zeta_0 \mathbf{1}\right) - \min_{\boldsymbol{x} \geq \zeta_0 \mathbf{1}} \left(\hat{\boldsymbol{\zeta}} - \boldsymbol{x}\right)^\top \Omega^{-1} \left(\hat{\boldsymbol{\zeta}} - \boldsymbol{x}\right)\right] \\ &= 10.2279 - 3.7984 = 6.4295. \end{aligned}$$

We then compute the p-value using the simulated weights as 0.0621, and we fail to reject the null at the 0.05 level. In particular, we fail to reject the null hypothesis that the signal-noise ratio of the Mkt is greater than $0.12\text{mo.}^{-1/2}$.

⊣

The chi-bar-square test for general R will not scale well to large k, since it requires solving a quadratic program to perform a projection. Worse, the computation of the chi-bar-square weights requires that we solve that quadratic program thousands of times. Here we describe a simpler form of the test under the assumption that R follows the rank one updated correlation model of Equation 5.4, namely

$$\mathsf{R} = (1-\rho)\,\mathsf{I} + \rho\left(\mathbf{1}\mathbf{1}^{\top}\right), \tag{5.15}$$

for $\rho \geq 0$.

Again, we assume a normal approximation for $\hat{\boldsymbol{\zeta}}$. Under the null hypothesis that $\boldsymbol{\zeta} = \zeta_0\mathbf{1}$, we have

$$\hat{\boldsymbol{\zeta}} \approx \mathcal{N}\left(\boldsymbol{\zeta}, \frac{1}{n}\Omega\right), \text{ where } \Omega = a_0\,\mathsf{I} + a_2\left(\mathbf{1}\mathbf{1}^{\top}\right). \tag{5.16}$$

As noted above (and via Exercise 5.6) the inverse square root of this Ω takes the form

$$\Omega^{-1/2} = b_0\,\mathsf{I} + b_2\left(\mathbf{1}\mathbf{1}^{\top}\right).$$

Letting $\hat{\boldsymbol{\xi}} = \Omega^{-1/2}\hat{\boldsymbol{\zeta}}$ and $\boldsymbol{\xi} = \Omega^{-1/2}\boldsymbol{\zeta}$, the normal approximation can be rewritten as

$$\sqrt{n}\hat{\boldsymbol{\xi}} \approx \mathcal{N}\left(\sqrt{n}\boldsymbol{\xi}, \mathsf{I}\right).$$

Note that $\mathbf{1}$ is an eigenvector of $\Omega^{-1/2}$ with

$$\Omega^{-1/2}\mathbf{1} = c\mathbf{1},$$

for $c = b_0 + kb_2$. Note that the b_2 can be chosen such that $c > 0$. From the order-preserving nature of $\Omega^{-1/2}$, the null and alternative of Hypothesis 5.11 can be expressed as

$$H_0 : \forall_i \xi_i = c\zeta_0 \quad \text{versus} \quad H_1 : \forall_i \xi_i \geq c\zeta_0 \text{ and } \exists_j \xi_j > c\zeta_0.$$

We can now appeal to the simple chi-bar square test. [151, 173] First transform the vector of Sharpe ratios to $\hat{\boldsymbol{\xi}}$ via

$$\hat{\boldsymbol{\xi}} = \mathsf{R}^{-1/2}\hat{\boldsymbol{\zeta}} = c\bar{\zeta}\mathbf{1} + (1-\rho)^{-1/2}\left(\hat{\boldsymbol{\zeta}} - \bar{\zeta}\mathbf{1}\right), \tag{5.17}$$

where $\bar{\zeta}$ is the average of the sample Sharpe ratios. Then compute

$$\bar{x}^2 = n\sum_i \left(\hat{\xi}_i - c\zeta_0\right)_+^2, \tag{5.18}$$

where y_+ is the positive part of y, *i.e.*, $y_+ = y$ if $y > 0$ and zero otherwise. Then compute the CDF of the corresponding chi-bar square distribution as

$$Q = \sum_{i=0}^{k} w_i F_{\chi^2}\left(\bar{x}^2; i\right), \tag{5.19}$$

where $F_{\chi^2}(i; x)$ is the cumulative distribution of the χ^2 distribution with i degrees of freedom, and w_i are the chi-bar square weights, which for diagonal matrix take the value

$$w_i = \binom{k}{i} 2^{-k}.$$

Reject the null hypothesis at the α level if $1 - Q \leq \alpha$.

Note that, as with the Bonferroni correction, the test statistic $\bar{x}^2$ is computed on all elements of $\hat{\boldsymbol{\zeta}}$, and thus the decision to reject the null may not be "about" $\hat{\zeta}_1$. In testing we will see that the one-sided test is highly susceptible to distribution of the $\boldsymbol{\zeta}$, moreso than the Bonferroni correction.

We note that under this setup it is also easy to use Follman's test, which is a very simple procedure with increased power against one-sided alternatives. [48] Here we would compute

$$g^2 = nkc^2 \left(\bar{\zeta} - \zeta_0\right)^2 + nb_0^2 \sum_i \left(\hat{\zeta}_i - \bar{\zeta}\right)^2, \tag{5.20}$$

and reject at the α level if both $1 - F_{\chi^2}\left(g^2; k\right) \leq 2\alpha$ and $\bar{\zeta} > \zeta_0$.

Example 5.1.8 (One-Sided Tests, Five Industry Portfolios). We consider the monthly returns of five industry portfolios, as introduced in Example 1.2.3. These include 1104 months of data on five industries from Jan 1927 to Dec 2018. We compute the Sharpe ratio of the returns of each as follows:

$$\left[\begin{array}{ccccc} Other & Technology & Manufacturing & Consumer & Healthcare \\ \hline 0.140 & 0.170 & 0.172 & 0.187 & 0.193 \end{array}\right] \text{mo.}^{-1/2}.$$

We have reordered the industries in increasing Sharpe ratio. The industry portfolio with the highest Sharpe ratio was Healthcare with a Sharpe ratio of around $0.1927\,\text{mo.}^{-1/2}$ which is approximately $0.6674\,\text{yr}^{-1/2}$.

The correlation of industry returns is high: the pairwise sample correlations range from 0.7081 to 0.8906 with a median value of 0.8014. We assume $\mathsf{R} = (1 - \rho)\,\mathsf{I} + \rho\left(\mathbf{1}\mathbf{1}^{\top}\right)$ and assume this median value as ρ. We compute the median kurtosis factor of industry returns to be 3.5579.

We test the null $H_0 : \forall_i \zeta_i = \zeta_0 = 0.15\text{mo.}^{-1/2}$ against the one-sided alternative $H_1 : \forall_i \zeta_i \geq \zeta_0$ and $\exists_j \zeta_j > \zeta_0$. Under the null we compute

$$\begin{aligned} a_0 &= 0.213, & a_2 &= 0.8415, \\ b_0 &= 2.1669, & b_2 &= -0.3383. \end{aligned}$$

From these we compute $c = 0.4756$. We compute the statistic from Equation 5.18 as $\bar{x}^2 = 5.4172$, corresponding to a p-value of 0.1189. We fail to reject at the 0.05 level.

We compute Follman's statistic (Equation 5.20) as $g^2 = 9.3448$. This corresponds to the upper 0.0961 quantile of the χ^2 distribution with 5 degrees of freedom, which is smaller than 0.10. Moreover, since $\bar{\zeta} = 0.1725\text{mo.}^{-1/2} \geq 0.15\text{mo.}^{-1/2} = \zeta_0$, we reject the null hypothesis at the 0.05 level.

$\dashv$

5.1.4. Hansen's Asymptotic Correction

One failing of many of the approaches considered above is the problem of irrelevant alternatives. That is, instead of testing under the null of equality (Hypothesis 5.11 above) we should test the following

$$H_0 : \forall_i \zeta_i \leq c \quad \text{versus} \quad H_1 : \exists_i \zeta_i > c.$$

Testing such a composite null hypothesis is typically via a *non-similar* test, *i.e.*, one which has a type I rate no greater than the nominal rate for all $\boldsymbol{\zeta}$ in the null, and which achieves that nominal rate for some $\boldsymbol{\zeta}$ under the null. Such tests achieve the nominal rate under the *least favorable configuration* (LFC), which in our case is the null of equality given in Hypothesis 5.11. [151]

Hansen describes a procedure which avoids this problem. The idea is elegant, and ultimately simple to implement. [65, 64] For the problem of overoptimism by selection, it amounts to assuming that the null takes the form

$$H_0 : \forall_i\, \zeta_i \leq c \text{ and } \left|\zeta_i - \hat{\zeta}_i\right| \leq g_n \quad \text{versus} \quad H_1 : \exists_i \zeta_i > c,$$

for some g_n. Note this seems odd since the sample Sharpe ratio appears in the null hypothesis to be tested. Nonetheless, Hansen describes how such a test can be performed while maintaining a maximum type I rate asymptotically, and achieving higher power.

Hansen applied this correction to the chi-bar-square statistic of Equation 5.18, and later to a Studentized version of White's Reality Check statistic, which is rather like the corrected Bonferroni statistic computed in Equation 5.7. [65, 64] Applying Hansen's correction to our problem is simple: compute $\hat{\boldsymbol{\xi}}$ as in Equation 5.17. Let $\tilde{k}$ be the number of elements of $\hat{\boldsymbol{\xi}}$ greater than $c\zeta_0 - \sqrt{(2 \log \log n)/n}$, where $c = (1 + (k-1)\,\rho)^{-1/2}$. If $\tilde{k} = 0$ fail to reject. Otherwise compute the chi-bar-square statistic $\bar{x}^2$ as in Equation 5.18 and reject if

$$\sum_{i=0}^{\tilde{k}} \binom{\tilde{k}}{i} 2^{-\tilde{k}} F_{\chi^2}\left(\bar{x}^2; i\right) \geq 1 - \alpha.$$

This is the same as the chi-bar-square test considered above, but with reduced degrees of freedom *which depend on the observed*.

The same correction is easily applied to the Bonferroni maximum test: again, compute $\hat{\boldsymbol{\xi}}$ and $\tilde{k}$. If $\tilde{k} = 0$ fail to reject. Otherwise reject at the α level if

$$\max_i \hat{\boldsymbol{\xi}}_i - c\zeta_0 \geq z_{1-\alpha/\tilde{k}}.$$

This is similar in spirit to Hansen's SPA test, except it does not use the bootstrap procedure to estimate the standard error as described by Hansen; it is similar in every other regard. [64]

Example 5.1.9 (One-Sided Tests with Hansen's Correction, Five Industry Portfolios). We return to the testing of the five industry portfolios considered in Example 5.1.8.

There we used the chi-bar-square test to test the null $H_0 : \forall_i \zeta_i = 0.15\text{mo.}^{-1/2}$, and failed to reject at the 0.05 level.

We now apply Hansen's correction to the chi-bar-square test. As above, we compute $c = 0.4756$, and $\bar{x}^2 = 5.4172$. The cutoff for the $\hat{\boldsymbol{\xi}}$ is $c\zeta_0 - \sqrt{2 \log \log n / n} = 0.012$. We find that 4 of the elements of $\hat{\boldsymbol{\xi}}$ are above the cutoff. Based on this many degrees of freedom, we compute the p-value of the chi-bar-square statistic to be 0.0813 and we again fail to reject the null at the 0.05 level.

⊣

5.1.5. Conditional Inference

Taming overoptimism via testing Hypothesis 5.1 seems like overkill. One is typically only interested in performing inference on $\zeta_{(k)}$, the signal-noise ratio associated with the strategy that has maximum Sharpe ratio, rather than on *all* the ζ_i. We can do this directly via *conditional inference*. The idea is to perform inference on $\zeta_{(k)}$ conditional on having observed that $\hat{\zeta}_{(k)}$ is the largest Sharpe ratio. One suspects that by directly testing the quantity of interest, one could gain statistical power over the MHT corrections considered above.

Briefly the conditional probability of event A conditional on event B is the probability that both A and B occur, divided by the probability that B occurs:

$$\Pr\left\{A \,|\, B\right\} = \frac{\Pr\left\{A \cap B\right\}}{\Pr\left\{B\right\}}.$$

First we will use conditional probability to analyze some simpler problems, before returning to overoptimism by selection.

Recall the opportunistic strategy introduced in Section 3.5.2: you observe the Sharpe ratio of an asset, then decide to hold it long if positive, otherwise you will hold the asset short. You then wish to perform inference on your strategy, taking into account that the sign depends on the sample. Previously we used symmetric confidence intervals to approach this problem, but it is easily described as a conditional inference problem.

Adjusting the sign of returns to match whether we hold the asset long or short, we are performing inference on ζ conditional on $\hat{\zeta} > 0$. The Sharpe ratio in this case has the (conditional) distribution function

$$F_{SR}\left(\hat{\zeta} \,\middle|\, \hat{\zeta} > 0 \,; \zeta, n\right) = \frac{F_{SR}\left(\hat{\zeta}; \zeta, n\right) - F_{SR}\left(0; \zeta, n\right)}{1 - F_{SR}\left(0; \zeta, n\right)}.$$

More generally, to test the null hypothesis

$$H_0 : \zeta_{(k)} = \zeta_0 \,\middle|\, \hat{\zeta} \geq \zeta_1 \quad \text{versus} \quad H_1 : \zeta_{(k)} > \zeta_0, \tag{5.21}$$

reject at the α level if $F_{SR}\left(\hat{\zeta} \,\middle|\, \hat{\zeta} > \zeta_1 \,; \zeta_0, n\right) \geq 1 - \alpha$. Equivalently, to construct an α lower confidence bound, find ζ_0 such that $F_{SR}\left(\hat{\zeta} \,\middle|\, \hat{\zeta} > \zeta_1 \,; \zeta_0, n\right) = 1 - \alpha$.

Example 5.1.10 (Confidence intervals, Conditional Inference). As in Example 3.5.3, consider the case of 504 daily observations of some hypothetical asset's returns, which result in a measured Sharpe ratio of exactly $0.6\mathrm{yr}^{-1/2}$, assuming 252 days per year. The 'exact' 95% confidence interval is computed as $[-0.7867, 1.9861]\,\mathrm{yr}^{-1/2}$.

Assuming that we are considering holding this asset long because we observed $\hat{\zeta} \geq 0$, we compute 95% conditional confidence interval on ζ as $[-2.6268, 1.9538]\,\mathrm{yr}^{-1/2}$. Note this confidence interval is much wider than the nearly symmetric unconditional confidence interval, which, as outlined in Section 3.5.2, we are justified in applying to the opportunistic strategy.

As a consolation prize, by using conditional inference, we can compute a one-sided conditional confidence interval to this problem. In this case, we compute the 95% one sided conditional confidence interval on ζ as $[-2.0142, \infty]\,\mathrm{yr}^{-1/2}$. This is still lower than the lower bound suggested by the symmetric confidence interval.

⊣

At risk of overgeneralizing from this example, we often find that conditional inference has less power than other inferential approaches. In this case, the conditional inference procedure could not exploit the fact that had we observed $\hat{\zeta} < 0$, we would have opportunistically flipped the sign of the asset and performed another test. That is, the conditioning event does not fully capture our testing procedure and instead assumes that had we observed $\hat{\zeta} < 0$ we would not have performed any test.

Conditional inference for overoptimism by selection To simplify the exposition, we will suppose that, conditional on observing the vector $\hat{\boldsymbol{\zeta}}$, one rearranges the indices in increasing order, so that the first index refers to the asset with the smallest observed Sharpe ratio, and the last index, k, to the asset with the largest. This is to avoid the cumbersome notation of $\hat{\zeta}_{(k)}$, and we instead can just write $\hat{\zeta}_k$. We note this maximum condition can be written in the form $\mathsf{A}\hat{\boldsymbol{\zeta}} \leq \boldsymbol{b}$ for $k-1 \times k$ matrix A defined by

$$\mathsf{A} = \left[\begin{array}{ccccc} 1 & 0 & \ldots & 0 & -1 \\ 0 & 1 & \ldots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -1 \end{array}\right],$$

and where $\boldsymbol{b}$ is the $(k-1)$-dimensional zero vector. Also note that we are interested in performing inference on ζ_k, which we can express as $\boldsymbol{\eta}^{\top}\boldsymbol{\zeta}$ for $\boldsymbol{\eta} = \boldsymbol{e}_k$.

Since, by Equation 4.29, the vector $\hat{\boldsymbol{\zeta}}$ is approximately normally distributed, we can use the following theorem due to Lee *et al.* [90, 129] :

Theorem 5.1.11 (Lee *et al.*, Theorem 5.2 [90])**.** *Suppose* $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. *Define* $\boldsymbol{c} = \Sigma\boldsymbol{\eta}/\boldsymbol{\eta}^{\top}\Sigma\boldsymbol{\eta}$, *and* $\boldsymbol{z} = \boldsymbol{y} - \boldsymbol{c}\boldsymbol{\eta}^{\top}\boldsymbol{y}$. *Let* $\Phi(x)$ *be the CDF of a standard normal, and let Let* $F(x; a, b, 0, 1)$ *be the CDF of a standard normal truncated to* $[a, b]$*:*

$$F(x; a, b, 0, 1) =_{\mathrm{df}} \frac{\Phi(x) - \Phi(a)}{\Phi(b) - \Phi(a)}.$$

Let $F\left(x; a, b, \mu, \sigma^2\right)$ be the CDF of a general truncated normal, defined by

$$F\left(x; a, b, \mu, \sigma^2\right) = F\left(\frac{x-\mu}{\sigma}; \frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma}, 0, 1\right).$$

Then, conditional on $\mathsf{A}\boldsymbol{y} \leq \boldsymbol{b}$, the random variable

$$F\left(\boldsymbol{\eta}^\top \boldsymbol{y}; \mathcal{V}^-, \mathcal{V}^+, \boldsymbol{\eta}^\top \boldsymbol{\mu}, \boldsymbol{\eta}^\top \Sigma \boldsymbol{\eta}\right)$$

is Uniform on $[0,1]$, where $\mathcal{V}^-$ and $\mathcal{V}^+$ are given by

$$\mathcal{V}^- = \max_{j:(\mathsf{A}\boldsymbol{c})_j < 0} \frac{\boldsymbol{b}_j - (\mathsf{A}\boldsymbol{z})_j}{(\mathsf{A}\boldsymbol{c})_j}, \quad \mathcal{V}^+ = \min_{j:(\mathsf{A}\boldsymbol{c})_j > 0} \frac{\boldsymbol{b}_j - (\mathsf{A}\boldsymbol{z})_j}{(\mathsf{A}\boldsymbol{c})_j}.$$

This theorem gives us a way to test the null hypothesis

$$H_0 : \zeta_{(k)} = \zeta_0 \quad \text{versus} \quad H_1 : \zeta_{(k)} > \zeta_0. \tag{5.22}$$

To test this null, set $\boldsymbol{y} = \hat{\boldsymbol{\zeta}}$ and compute $\boldsymbol{c}, \boldsymbol{z}, \mathcal{V}^-$, and $\mathcal{V}^+$ as in the theorem; estimate the covariance, Ω via Equation 4.29; finally, reject the null hypothesis at the α level if $F\left(\hat{\zeta}_{(k)}; \mathcal{V}^-, \mathcal{V}^+, \zeta_0, \frac{1}{n}\Omega_{(k),(k)}\right) > 1 - \alpha$.

Note that we do not have to compute the entire matrix Ω, we only need one column of it to compute the vector $\boldsymbol{c}$ (and one element of that is $\Omega_{(k),(k)}$). That column is

$$\Omega_{:,(k)} = \mathsf{R}_{:,(k)} + \frac{\kappa - 1}{4}\boldsymbol{\zeta}\zeta_0 + \frac{\kappa}{2}\operatorname{Diag}\left(\boldsymbol{\zeta}\right)\left(\mathsf{R}_{:,(k)} \odot \mathsf{R}_{:,(k)}\right)\zeta_0. \tag{5.23}$$

The vector $\boldsymbol{\zeta}$ is unknown; under the null one can estimate it as either $\zeta_0\mathbf{1}$ or $\zeta_0\boldsymbol{e}_{(k)}$.

Example 5.1.12 (Five Industry Portfolios). We consider the monthly returns of five industry portfolios, as introduced in Example 1.2.3. These include 1104 months of data on five industries from Jan 1927 to Dec 2018. We compute the Sharpe ratio of the returns of each as follows:

$$\left[\begin{array}{ccccc} \mathit{Other} & \mathit{Technology} & \mathit{Manufacturing} & \mathit{Consumer} & \mathit{Healthcare} \\ \hline 0.140 & 0.170 & 0.172 & 0.187 & 0.193 \end{array}\right] \mathrm{mo.}^{-1/2}.$$

We have reordered the industries in increasing Sharpe ratio. The industry portfolio with the highest Sharpe ratio was Healthcare with a Sharpe ratio of around $0.1927\,\mathrm{mo.}^{-1/2}$ which is approximately $0.6674\,\mathrm{yr}^{-1/2}$.

We are interested in computing 95% upper confidence intervals on the signal-noise ratio of the Healthcare portfolio. We are only considering this portfolio as it is the one with maximum Sharpe ratio in our sample. If we had been interested in testing Healthcare without our conditional selection, we would compute the confidence interval $\left[0.1427\,\mathrm{mo.}^{-1/2}, \infty\right)$ based on inverting the non-central t-distribution. If instead we approximate the standard error by plugging in $0.1927\,\mathrm{mo.}^{-1/2}$ as the signal-noise ratio of Healthcare into Equation 4.29, we estimate the standard error of the Sharpe ratio to be $0.0304\,\mathrm{mo.}^{-1/2}$. Based on this we can compute the naïve confidence interval of the measured Sharpe ratio plus $z_{0.05} = -1.6449$ times the standard error. This also gives the confidence interval $\left[0.1427\,\mathrm{mo.}^{-1/2}, \infty\right)$.

Using the simple Bonferroni correction, however, since we selected Healthcare only for having the maximum Sharpe ratio, we should compute the confidence interval by adding $z_{0.01} = -2.3263$ times the standard error. This yields the confidence interval $\left[0.122\,\mathrm{mo.}^{-1/2}, \infty\right)$.

The correlation of industry returns is high, however. The pairwise sample correlations range from 0.7081 to 0.8906 with a median value of 0.8014. Plugging this value in as ρ, we find the value ζ_0 such that the z_1 from Equation 5.7 is equal to $z_{0.01} = -2.3263$. This leads to the confidence interval $\left[0.1253\,\mathrm{mo.}^{-1/2}, \infty\right)$.

We use this estimate of ρ to compute the chi-bar square test. We invert the test to find the 95% upper confidence interval $\left[0.1406\,\mathrm{mo.}^{-1/2}, \infty\right)$.

Finally we use the conditional estimation procedure, inverting the hypothesis test to find the corresponding population value. We compute Ω by assuming returns are Gaussian, so $\kappa = 1$, and plugging in the sample $\hat{\boldsymbol{\zeta}}$ for $\boldsymbol{\zeta}$. This yields the confidence interval $\left[0.0732\,\mathrm{mo.}^{-1/2}, \infty\right)$.

$\dashv$

This example is consistent with our previous findings suggesting conditional estimation is less powerful than an MHT correction. In the following example we will examine this question directly via simulations under the alternative hypothesis. However, 'the' alternative can take many forms. One interpretation is that we condition on $\zeta_1 > 0$, where again the indexing is such that $\hat{\zeta}_1$ was the maximum over k assets; then we estimate the probability of (correctly) rejecting $\zeta_1 = 0$ versus ζ_1. However, we suspect that the power, as described in this way, would depend on the distribution of values of $\boldsymbol{\zeta}$.

We consider three forms for the alternative: one where all k elements of ζ are equal ("all-equal"), and two where m of k elements of ζ are equal to some positive value, and the remaining $k - m$ are negative that value, for the case $m = 1$ ("one-good"), and $k = 2m$, which we call "half-good".

Note that in the all equal case, since every asset has the same signal-noise ratio, whichever we select will have the same signal-noise ratio, and the test should have the same power as the t-test for a single asset. The conditional estimation procedure, however, may suffer in this case as we may condition on a $\hat{\zeta}_1$ that is very close to being non-optimal, resulting in a small test statistic for which we do not reject. On the other hand, for the one-good case, as the $k-1$ assets may have considerably negative signal-noise ratio, they are unlikely to exhibit the largest Sharpe ratio, and so the MHT is merely testing a single asset, but at the α/k level instead of the α level, resulting in lower power. The conditional estimation procedure, however, should not suffer in this test.

Example 5.1.13 (Power of conditional inference and MHT corrections). We perform simulations under all equal, one-good, and half-good configurations, letting the 'good' signal-noise ratio vary from 0 to $0.15\mathrm{day}^{-1/2}$, which corresponds to an 'annualized' signal-noise ratio of around $2.4\mathrm{yr}^{-1/2}$. We draw Gaussian returns with diagonal covariance for 100 assets, with $n = 1008$. For each setting we perform 10000 simulations then compute the empirical rejection rate of the test at the 0.05 level, conditional on the signal-noise ratio of the *selected* asset, which is to say the one with the largest Sharpe ratio. Note that in some simulations the largest Sharpe ratio is observed in an asset with a negative signal-noise ratio.

In Figure 5.9, we plot the power of the MHT Bonferroni test, the chi-bar square test, Follman's test, Hansen's chi-bar square and MHT ("SPA"), and the conditional estimation procedure versus the signal-noise ratio of the selected asset. We present facet columns for the three configurations, *viz.* all-equal, one-good, half-good. A horizontal line at 0.05 gives the nominal rate under the null, which occurs as $x = 0$ in these plots. As expected from the above explanation, chi-bar square has the highest power for the all-equal alternative, followed by Follman's test, then the MHT, then the conditional estimation test. These relationships are reversed for the one-good case. The chi-bar square test and Follman's test have very low power against the one-good alternative. All tests perform similarly in the half-good and all-equal alternatives, with the exception of Follman's test, which achieves a maximum power of 1/2 in the half-good case, as is to be expected since this is the probability that $\bar{\zeta} > 0$ in the half-good case.

The power of the conditional estimation procedure for the all equal case is rather disappointing. For the case where all assets have a signal-noise ratio of $2.4\mathrm{yr}^{-1/2}$, the test has a power of only around a half. The test suffers from low power because we are conditioning on "$\hat{\zeta}_1$ is the largest Sharpe ratio", where we should actually be conditioning on "the asset with the largest Sharpe ratio."

Note the odd plot in the half-good facet: the MHT correction and one-sided tests have greater than 0.05 rejection rate for *negative* signal-noise ratio. The plot is somewhat misleading in this case, however. We have performed 10000 simulations for each setting of the 'good' signal-noise ratio; in some very small number of them for the half-good case, an asset with negative signal-noise ratio exhibits the maximum Sharpe ratio. We are plotting the rejection rate for the test in this case. But note that the null hypothesis that MHT and the one-sided test are testing *is* violated in this case, because half the assets have positive signal-noise ratio, and the alternative procedures test the null that all assets have zero or lower signal-noise ratio. We have not shown the probability that a 'bad' asset has the highest Sharpe ratio, but note that when the 'good' signal-noise ratio is greater than $0.05\mathrm{day}^{-1/2}$ we do not observe this occuring even once over the 10000 simulations performed for each setting.

We note that the one-sided test appears to have higher power than either of the other tests for the all-equal and half-good populations, but fails to reject at all in the one-good case, except under the null. This is to be expected, since the chi-bar square test depends strongly on all the observed Sharpe ratios, and in this case we expect many of them to be negative.

We note that the chi-bar-square test appears to have higher power than the other tests for the all-equal and half-good populations, but fails to reject at all in the one-good case, except under the null. This is to be expected, since the chi-bar-square test depends strongly on all the observed Sharpe ratios, and in this case we expect many

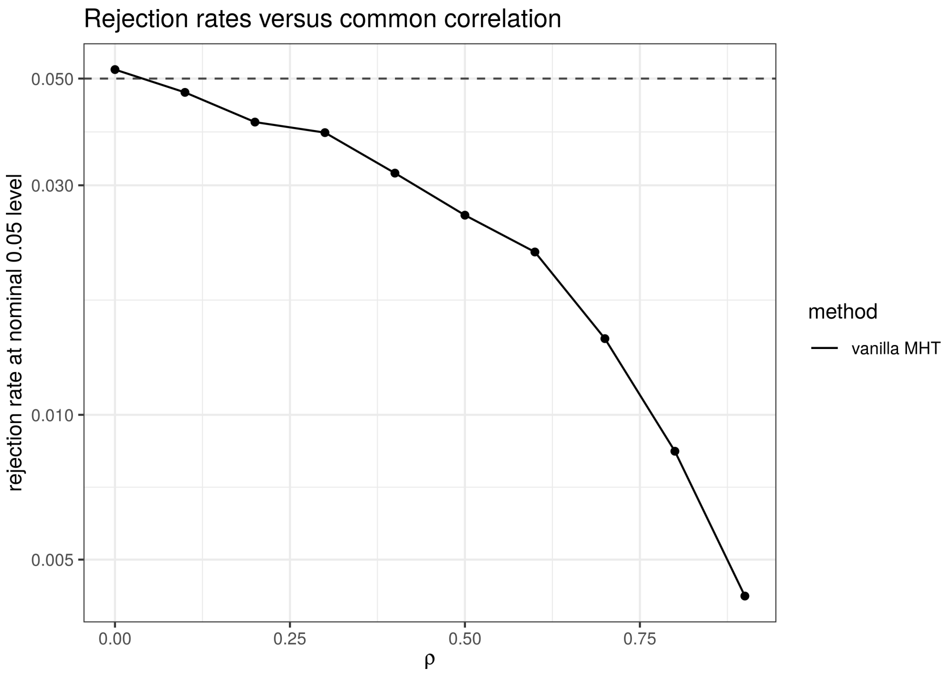

Figure 5.9.: The empirical power of the conditional estimation, MHT corrected test, Follman's test, the chi-bar-square test, and the chi-bar-square and MHT tests with Hansen's $\log\log$ correction are shown versus the signal-noise ratio of the asset with maximum Sharpe ratio under different arrangements of the vector $\boldsymbol{\zeta}$. We jitter the points horizontally to distinguish tests with very similar rejection rates.

of them to be negative. The $\log\log$ adjustment has little effect in the all-equal and half-good case, but greatly improves the power of the MHT and chi-bar square tests in the one-good case. For this sample size, Hansen's adjustment nearly maintains the nominal type I rate under the null, though this is unlikely to hold for smaller n.

It is interesting to note that while the conditional estimation procedure generally has lower power than the other tests (except in the one-good case), it appears to have monotonic rejection probability with respect to the signal-noise ratio of the *selected* asset. That is, in the half-good case, it has low rejection probability in the odd simulations where a 'bad' asset is selected.

We doubt that the simple experiments performed here have revealed all the relevant differences between the various tests or when one dominates the others. ⊣

5.1.6. Subspace Approximation

Another potential approach to the problem which may be useful in the case where returns are highly correlated, as one expects when returns are from backtested quantitative trading strategies, is via a subspace approximation. First we assume that the $n \times k$ matrix of returns, X can be approximated by a p-dimensional subspace

$$\mathsf{X} \approx \mathsf{Y}\mathsf{W},$$

where Y is a $n \times p$ matrix of 'latent' returns, and W is a $k \times k$ 'loading' matrix.

Now the column of X with maximal $\hat{\boldsymbol{\zeta}}$ has Sharpe ratio that is smaller than

$$\hat{\zeta}_* =_{\mathrm{df}} \max_{\boldsymbol{\nu}} \frac{\hat{\boldsymbol{\mu}}^\top \boldsymbol{\nu}}{\sqrt{\boldsymbol{\nu}^\top \hat{\boldsymbol{\Sigma}} \boldsymbol{\nu}}},$$

where $\hat{\boldsymbol{\mu}}$ is the p-vector of the (sample) means of columns of Y and $\hat{\boldsymbol{\Sigma}}$ is the sample covariance matrix. This maximum takes value

$$\hat{\zeta}_* = \sqrt{\hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\mu}}},$$

which is, up to scaling, Hotelling's T^2 statistic.

We shall see in the sequel that under the null hypothesis that the rows of Y are independent draws from a Gaussian random variable with zero mean, then

$$\frac{(n-p)\,\hat{\zeta}_*^2}{p\,(n-1)}$$

follows an F distribution with p and $n-p$ degrees of freedom. Under the alternative it follows a non-central F distribution. [5, 138] Via this upper bound $\hat{\zeta}_1 \leq \hat{\zeta}_*$, one can then perform tests on the null hypothesis $\forall_i \zeta_i = 0$.

However, this approach requires that one estimate p, the dimensionality of the latent subspace. Moreover, the subspace approximation may not be very good. It would seem that to get near equality of $\hat{\zeta}_1$ and $\hat{\zeta}_*$, the columns of X would have to contain both positive and negative exposure to the columns of Y. This in turn should result in mixed correlation of asset returns, which we may not observe in practice. Finally, empirical testing indicates this approach requires further development, as in the following example.

Example 5.1.14 (Overfit of MAC Strategy). Here we analyze, via simulation, the Sharpe ratios of Moving Average Crossover (MAC) strategies. A MAC strategy is simple to describe: for a single asset, compute two moving averages of the price series with different windows. When one moving average is greater than the other, hold the asset long, otherwise hold it short. One selects the two windows by (over)fitting to the available data.

We perform simulations of that process under the null hypothesis, where generated returns are zero mean and independent. Any realization of a MAC strategy in these simulations, with any choice of the windows, will have zero mean return and thus zero

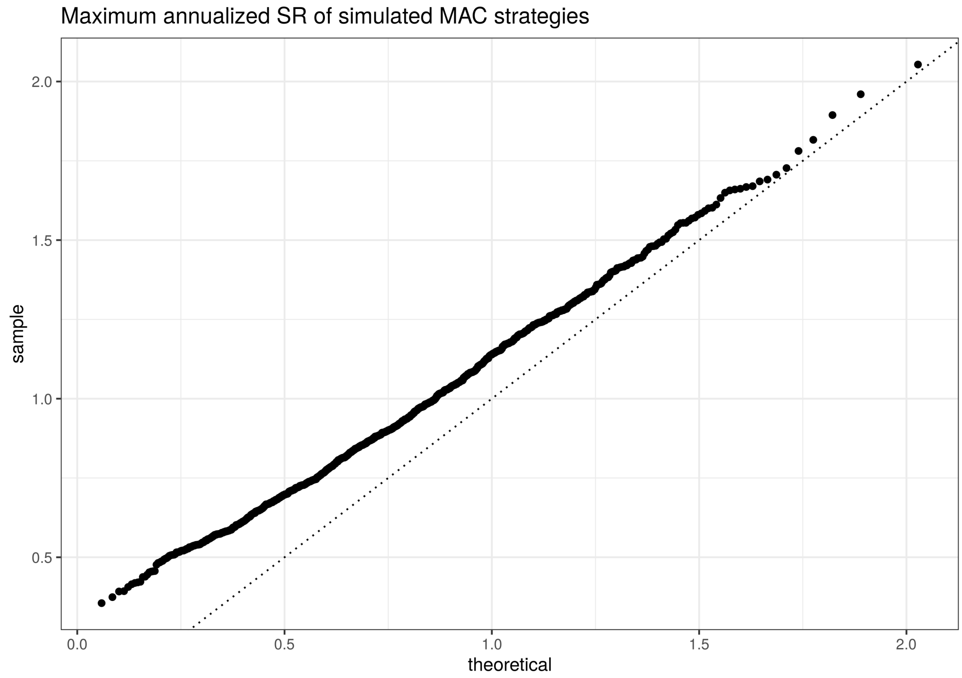

Figure 5.10.: QQ plot of the Sharpe ratio of the optimal 2-window MAC in a simulated backtest under the null. Simulations are over 1100 days. Theoretical quantiles are computed via the F distribution assuming $k = 3$.

Sharpe. We draw 1100 days of returns independently with zero mean. We turn the returns into a price series, and then test MAC strategies with a 'short' window size varying from 4 to 36 days, and the 'long' window size varying from 40 to 280 days. We select the window combination with highest Sharpe ratio and record that Sharpe ratio. We do not force the strategy to be long or short either windowed average, rather we test both. We repeat this process 1024 times.

In Figure 5.10, we Q-Q plot the annualized Sharpe ratio of the selected MAC strategy against the theoretical distribution based on an F distribution with 3 and $1100 - 3$ degrees of freedom. Note that the choice of p is intuitive, but perhaps a different value is more consistent with the F distribution approximation for this data. The Q-Q plot indicates that our approximation is off by an affine shift: we observe higher Sharpe ratios than our subspace approximation supports, or we have underestimated the p.

⊣

5.2. Miscellany

5.2.1. A post hoc test on the Sharpe ratio

In Section 4.3.1, we described the test for the null hypothesis that k different assets have the same signal-noise ratio. This is analogous to the classical ANOVA procedure, which is for testing whether a variable has the same mean across k different groups. In contrast to the ANOVA, we are usually performing a paired test, where returns across the k assets are observed contemporaneously, and we need not assume common volatility of returns.

In the classical setting, if one rejects the null in an ANOVA, a *post hoc* test is then prescribed. [22] This test, also known as *Tukey's Range Test* or the *Honest Significant Difference* test, is based on the range of a number of independent Gaussians, divided by a rescaled chi-square. Tukey's test can easily be adapted to the case of testing the signal-noise ratios of multiple assets. [130]

As in Section 5.1.2, we first consider the case of rank one updated correlation model,

$$\mathsf{R} = (1-\rho)\,\mathsf{I} + \rho\left(\mathbf{1}\mathbf{1}^{\top}\right), \tag{5.24}$$

where $|\rho| \le 1$. Under this correlation structure, recalling Equation 5.3,

$$\sqrt{n}\Omega^{-1/2}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right) \approx \mathcal{N}\left(\mathbf{0}, \mathsf{I}\right).$$

We then showed (*cf.* Equation 5.6 and Exercise 5.6) that

$$\mathsf{R}^{-1/2} = \left((1-\rho)\left(1 + \frac{\kappa}{2}\zeta_0^2\left(1+\rho\right)\right)\right)^{-1/2}\mathsf{I} + c\left(\mathbf{1}\mathbf{1}^{\top}\right). \tag{5.25}$$

for some constant c.

Now we consider the difference in Sharpe ratios of two assets, indexed by i and j. Let $\boldsymbol{v} = \boldsymbol{e}_i - \boldsymbol{e}_j$, where $\boldsymbol{e}_i$ is the i^{th} column of the identity matrix. From Equation 5.2.1 we have

$$\begin{aligned}
\boldsymbol{v}^{\top}\boldsymbol{z} &= \boldsymbol{v}^{\top}\sqrt{n}\mathsf{R}^{-1/2}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right),\\
&= \sqrt{n}\boldsymbol{v}^{\top}\left[\left((1-\rho)\left(1 + \frac{\kappa}{2}\zeta_0^2\left(1+\rho\right)\right)^{-1/2}\right)\mathsf{I} + c\left(\mathbf{1}\mathbf{1}^{\top}\right)\right]\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right),\\
&= \sqrt{\frac{n}{(1-\rho)\left(1 + \frac{\kappa}{2}\zeta_0^2\left(1+\rho\right)\right)}}\boldsymbol{v}^{\top}\hat{\boldsymbol{\zeta}},\\
&\approx \sqrt{\frac{n}{1-\rho}}\boldsymbol{v}^{\top}\hat{\boldsymbol{\zeta}}.
\end{aligned}$$

Here we have used that $\boldsymbol{v}^{\top}\mathbf{1} = 0$ and under the null hypothesis, $\boldsymbol{\zeta}_0$ is some constant times $\mathbf{1}$. The last approximation follows because ζ_0^2/n is likely to be very small for most practical work. Then

$$\hat{\zeta}_i - \hat{\zeta}_j = \sqrt{\frac{1-\rho}{n}}\left(z_i - z_j\right). \tag{5.26}$$

Now note that the $\boldsymbol{z}$ is distributed as a standard multivariate normal. So the *range* of $\hat{\boldsymbol{\zeta}}$, which is to say $\max_{i,j}\left(\hat{\zeta}_i - \hat{\zeta}_j\right)$, is distributed as $\sqrt{\left(1-\rho\right)/n}$ times the range of a standard k-variate normal.

To quote this as a hypothesis test,

$$\max_{i,j}\left|\hat{\zeta}_i - \hat{\zeta}_j\right| \geq HSD = q_{1-\alpha,k,\infty}\sqrt{\frac{\left(1-\rho\right)}{n}}, \tag{5.27}$$

with probability α, where the $q_{1-\alpha,k,l}$ is the upper α-quantile of the Tukey distribution with k and l degrees of freedom. In the R language, this quantile may be computed via the `qtukey` function. [139, 116] With $l = \infty$, the cutoff HSD is the rescaled upper α quantile of the range of k independent Gaussians. That is, $q_{1-\alpha,k,\infty}$ is the number such that

$$1 - \alpha = k\int_{-\infty}^{\infty} \phi\left(x\right)\left(\Phi\left(x + q_{1-\alpha,k,\infty}\right) - \Phi\left(x\right)\right)^{k-1} \mathrm{d}x.$$

The normal approximation of Equation 4.29 may be too rough of an approximation for computation of HSD, even if the covariance is approximately correct. The distributional shape of $\hat{\boldsymbol{\zeta}}$ may be far enough from multivariate normal that we cannot use Tukey's distribution for a cutoff, especially when n is small and k is large. In that case, one is tempted to compare the observed range to

$$HSD = q_{1-\alpha,k,n-1}\sqrt{\frac{\left(1-\rho\right)}{n-1}}. \tag{5.28}$$

The reasoning behind this heuristic is that we are computing the range of (non-independent) t statistics, up to scaling, which is almost the same as the Tukey distribution, which is the ratio of the range of normals divided by a pooled χ variable. We will refer to the cutoff of Equation 5.27 as "$df = \infty$" and the cutoff of Equation 5.28 as the "$df = n-1$" cutoff. While we note the $df = n-1$ cutoff stands on perhaps shakier theoretical grounds, experimental evidence suggests that it maintains the nominal type I rate much better than the $df = \infty$ cutoff, especially for the small n case. [130]

Bonferroni Cutoff: We note that an alternative calculation provides a very similar cutoff value. Consider two assets with correlation ρ, and with signal-noise ratios of $\zeta\left(1+\epsilon\right)$ and ζ. Recalling Equation 4.45, the difference in Sharpe ratios can be shown to be approximately normal with

$$\left[\hat{\zeta}_1 - \hat{\zeta}_2\right] \rightsquigarrow \mathcal{N}\left(\epsilon\zeta, \frac{2}{n}\left(1-\rho\right) + \frac{\zeta^2}{2n}\left(1 + \left(1+\epsilon\right)^2 - 2\rho^2\left(1+\epsilon\right)\right)\right). \tag{5.29}$$

Again, assuming that ζ^2/n will be very small for most practical work, one can compute a "Bonferroni Cutoff" as

$$BC = \sqrt{\frac{2\left(1-\rho\right)}{n}} z_{1-\alpha/\binom{k}{2}}, \tag{5.30}$$

where z_α is the α quantile of the standard normal distribution. This cutoff is based on a Bonferroni correction that recognizes we are performing $\binom{k}{2}$ pairwise comparison tests. The cutoff BC is typically very similar to HSD (for $df = \infty$) or slightly smaller (and is easier to compute), *cf.* Exercise 5.15. We note that since BC is based on a normal approximation, it may suffer from the same issues that the HSD cutoff does for small samples. However, there is hope one can compute an exact small n Bonferroni cutoff.

Compact Letter Displays: One way to summarize the results of the *post hoc* test is via a compact letter display. [22, 136] Here one assigns each asset to one or more groups, with the groups identified by a letter. The group assignment is chosen such that two assets with Sharpe ratio greater than HSD are not assigned to the same group. While this could be done trivially, usually an assignment with few groups is achievable. We illustrate in Example 5.2.1.

Arbitrary correlation structure: The test outlined above is strictly only applicable to the rank-one correlation matrix, $\mathsf{R} = (1-\rho)\,\mathsf{I} + \rho\left(\mathbf{1}\mathbf{1}^\top\right)$. To apply the test to assets with arbitrary correlation matrices, one would like to appeal to a stochastic dominance result. For example, if one could adapt Slepian's lemma to the distribution of the *range*, then the above analysis could be applied where ρ is the smallest off-diagonal correlation, to give a test with maximum type I rate of α. However, it is not immediately clear that Slepian's lemma can be so modified. [153, 179, 177] The Bonferroni Cutoff, however, is easily adapted to this kind of worst-case analysis, however.

Example 5.2.1 (Five Industry Portfolios, *post hoc* test). We consider the returns of the 5 industry portfolios introduced in Example 1.2.3. The data consist of 1104 months of returns, from Jan 1927 to Dec 2018. Over this period we estimate the correlation of returns to be

$$\mathsf{R} = \begin{bmatrix} & Consumer & Manufacturing & Technology & Healthcare & Other \\ \hline Consumer & 1.00 & 0.87 & 0.81 & 0.78 & 0.88 \\ Manufacturing & 0.87 & 1.00 & 0.81 & 0.74 & 0.89 \\ Technology & 0.81 & 0.81 & 1.00 & 0.71 & 0.80 \\ Healthcare & 0.78 & 0.74 & 0.71 & 1.00 & 0.74 \\ Other & 0.88 & 0.89 & 0.80 & 0.74 & 1.00 \end{bmatrix}.$$

This is fairly well approximated by $(1-\rho)\,\mathsf{I} + \rho\left(\mathbf{1}\mathbf{1}^\top\right)$, with $\rho = 0.8$. As listed in Example 2.2.5, the Sharpe ratios range from $0.4852\text{yr}^{-1/2}$ for Other to $0.6674\text{yr}^{-1/2}$ for Healthcare.

First we perform the hypothesis test of equality of signal-noise ratios, via the χ^2 test of Equation 4.42. We compute a statistic of 12.2 which should be distributed as a $\chi^2(4)$ under the null. This corresponds to a p-value of 0.0159, and we reject the null of equality of all signal-noise ratios.

Using the $df = n-1$ formulation and the estimated ρ, we compute $HSD = 0.1796\text{yr}^{-1/2}$ for $\alpha = 0.05$, and narrowly reject the equality of signal-noise ratios for Other and Healthcare. Based on the observed Sharpe ratios and this HSD cutoff, a compact letter display of the five industries might look as follows:

$$\left[\begin{array}{ccccc} Healthcare & Consumer & Manufacturing & Technology & Other \\ \hline a & ab & ab & ab & b \end{array}\right]$$

The assignment to groups recognizes that we found a significant difference between Other and Healthcare, but not between any of the other pairs of assets.

⊣

Example 5.2.2 (Fama-French factors, *post hoc* test). Consider the four Fama French factors from Example 1.2.1. In Example 4.3.1, based on monthly returns from Jan 1927 to Dec 2018, we narrowly rejected the null hypothesis of equal signal-noise ratios among the four factors. The correlation of returns was estimated as

$$\mathsf{R} = \left[\begin{array}{c|cccc} & Mkt & SMB & HML & UMD \\ \hline Mkt & 1.00 & 0.32 & 0.24 & -0.33 \\ SMB & 0.32 & 1.00 & 0.12 & -0.14 \\ HML & 0.24 & 0.12 & 1.00 & -0.41 \\ UMD & -0.33 & -0.14 & -0.41 & 1.00 \end{array}\right].$$

Given the negative correlation of UMD to the other three factors, $(1-\rho)\,\mathsf{I}+\rho\left(\mathbf{1}\mathbf{1}^{\top}\right)$ is not a good model for R. Moreover, we cannot easily use such a correlation matrix as a lower bound. Instead, we perform the $\binom{4}{2}$ tests for equality of signal-noise ratio via the χ^2 test of Wright *et al.*, as given in Equation 4.42. We perform a Bonferroni correction on these, rejecting only if the p-value is less than $\alpha/6$. Under this test, we only reject equality of Mkt and SMB. The Sharpe ratios and compact letter display of the four factors then might look as follows:

$$\left[\begin{array}{ccc} factor & Sharpe & letters \\ \hline Mkt & 0.60 & a \\ UMD & 0.49 & ab \\ HML & 0.37 & ab \\ SMB & 0.23 & b \end{array}\right]$$

⊣

Exercises

Ex. 5.1 **The Šidák correction** Establish the correctness of the Šidák correction.

1. Let $p_1, p_2, \ldots, p_k$ be independent random variables uniformly on $[0,1]$. Prove that
$$\Pr\left\{\min_i p_i \le x\right\} = 1 - (1-x)^k.$$

2. Find x such that $1-(1-x)^k = \alpha$.

Ex. 5.2 Bonferroni and Šidák The Šidák correction yields a slightly larger critical region than the Bonferroni correction. What happens to the ratio of their sub-test type I errors,

$$\frac{1-(1-\alpha)^{1/k}}{\alpha/k},$$

for $\alpha \approx 0.05$ as $k \to \infty$?

Ex. 5.3 Limits to anti-correlation Let R be a $k \times k$ correlation matrix.

1. Prove that R is positive semi-definite.
2. Prove that the average off-diagonal element of R is no less than $-1/(k-1)$.

Ex. 5.4 Bonferroni power loss Extend Example 5.1.4 by finding the signal-noise ratio such that the Bonferroni correction achieves the nominal 0.05 type I rate for the k, n and ρ considered in that example.

Ex. 5.5 Rank one updated correlation

1. Show that the matrix $\mathsf{R} = (1-\rho)\,\mathsf{I} + \rho\left(\mathbf{1}\mathbf{1}^{\top}\right)$ is a valid correlation matrix if and only if $|\rho| \le 1$. (A valid correlation matrix is symmetric and positive semidefinite, has elements in $[-1, 1]$ and unit diagonal.)
2. Let $\boldsymbol{q}$ be a k-vector whose elements are ± 1. Show that the matrix $\mathsf{R} = (1-\rho)\,\mathsf{I} + \rho\left(\boldsymbol{q}\boldsymbol{q}^{\top}\right)$ is a valid correlation matrix if and only if $|\rho| \le 1$.
3. Let $\boldsymbol{q}$ be a k-vector whose elements are no greater than 1 in absolute value. Show that
$$\operatorname{Diag}\left(\mathbf{1} - \boldsymbol{q} \odot \boldsymbol{q}\right) + \boldsymbol{q}\boldsymbol{q}^{\top}$$
is a valid correlation matrix.

Ex. 5.6 Rank one update matrix powers Let $\boldsymbol{q}$ be a k-vector whose elements are -1, 0 or 1.

1. Show that the vectors $\mathbf{1} - |\boldsymbol{q}|$ and $\boldsymbol{q}$ are othogonal to each other, and thus
$$\operatorname{Diag}\left(\mathbf{1} - |\boldsymbol{q}|\right)\boldsymbol{q}\boldsymbol{q}^{\top} = \mathsf{0}.$$
2. Show that $\operatorname{Diag}\left(\mathbf{1} - |\boldsymbol{q}|\right)$ is idempotent:
$$\operatorname{Diag}\left(\mathbf{1} - |\boldsymbol{q}|\right)\operatorname{Diag}\left(\mathbf{1} - |\boldsymbol{q}|\right) = \operatorname{Diag}\left(\mathbf{1} - |\boldsymbol{q}|\right).$$
3. Let
$$\begin{aligned}\mathsf{A} &= a_0\mathsf{I} + a_1\operatorname{Diag}\left(\mathbf{1} - |\boldsymbol{q}|\right) + a_2\boldsymbol{q}\boldsymbol{q}^{\top},\\ \mathsf{B} &= b_0\mathsf{I} + b_1\operatorname{Diag}\left(\mathbf{1} - |\boldsymbol{q}|\right) + b_2\boldsymbol{q}\boldsymbol{q}^{\top},\end{aligned}$$

with $a_0 > 0$. Show that $\mathsf{BA} = \mathsf{I}$ if

$$\begin{aligned} b_0 &= a_0^{-1}, \\ b_1 &= \frac{-a_1}{a_0^2 + a_0 a_1}, \\ b_2 &= \frac{-a_2}{a_0^2 + a_0 a_2 \boldsymbol{q}^\top \boldsymbol{q}}. \end{aligned}$$

4. Let

$$\begin{aligned} \mathsf{A} &= a_0 \mathsf{I} + a_1 \operatorname{Diag}\left(\mathbf{1} - |\boldsymbol{q}|\right) + a_2 \boldsymbol{q}\boldsymbol{q}^\top, \\ \mathsf{B} &= b_0 \mathsf{I} + b_1 \operatorname{Diag}\left(\mathbf{1} - |\boldsymbol{q}|\right) + b_2 \boldsymbol{q}\boldsymbol{q}^\top, \end{aligned}$$

with $a_0 > 0$. Show that $\mathsf{BB} = \mathsf{A}$ if

$$\begin{aligned} b_0 &= \pm\sqrt{a_0}, \\ b_1 &= -b_0 \pm \sqrt{a_0 + a_1}, \\ b_2 &= \frac{-b_0 + \sqrt{a_0 + a_2 \boldsymbol{q}^\top \boldsymbol{q}}}{\boldsymbol{q}^\top \boldsymbol{q}}. \end{aligned}$$

5. Let

$$\begin{aligned} \mathsf{A} &= a_0 \mathsf{I} + a_1 \operatorname{Diag}\left(\mathbf{1} - |\boldsymbol{q}|\right) + a_2 \boldsymbol{q}\boldsymbol{q}^\top, \\ \mathsf{B} &= b_0 \mathsf{I} + b_1 \operatorname{Diag}\left(\mathbf{1} - |\boldsymbol{q}|\right) + b_2 \boldsymbol{q}\boldsymbol{q}^\top, \end{aligned}$$

with $a_0 > 0$. Show that $\mathsf{BBA} = \mathsf{I}$ if

$$\begin{aligned} b_0 &= a_0^{-1/2}, \\ b_1 &= -\frac{1}{\sqrt{a_0}} \pm \sqrt{\frac{1}{a_0 + a_1}} = -\frac{\sqrt{a_0 + a_1} \mp \sqrt{a_0}}{\sqrt{a_0^2 + a_0 a_1}}, \\ b_2 &= -\frac{1}{\boldsymbol{q}^\top \boldsymbol{q} \sqrt{a_0}} \pm \frac{1}{\boldsymbol{q}^\top \boldsymbol{q}} \sqrt{\frac{1}{a_0 + a_2 \boldsymbol{q}^\top \boldsymbol{q}}} = -\frac{1}{\boldsymbol{q}^\top \boldsymbol{q}} \frac{\sqrt{a_0 + a_2 \boldsymbol{q}^\top \boldsymbol{q}} \mp \sqrt{a_0}}{\sqrt{a_0^2 + a_0 a_2 \boldsymbol{q}^\top \boldsymbol{q}}}. \end{aligned}$$

Ex. 5.7 Order Preserving Matrices Say that a square matrix A is *order preserving* if for every $\boldsymbol{y} = \mathsf{A}\boldsymbol{x}$ if $x_i \le x_j$ then $y_i \le y_j$

1. Prove that A is order preserving if and only if all eigenvalues are non-negative and equal, with the possible exception of the eigenvalue associated with the eigenvector $\mathbf{1}$.
2. Prove that the matrix $a_0 \, \mathsf{I} + a_1 \left(\mathbf{1}\mathbf{1}^\top\right)$ is order preserving if and only if $a_0 \ge 0$.

Ex. 5.8 Monotonicity of inference Show that when using the rank-one model of correlation, it is conservative to keep ρ bounded away from 1. Let

$$a_0 = 1 - \rho + \frac{\kappa}{2}\zeta_0^2\left(1 - \rho^2\right), \qquad a_2 = \rho + \frac{\kappa - 1}{4}\zeta_0^2 + \frac{\kappa}{2}\zeta_0^2\rho^2,$$
$$b_0 = a_0^{-1/2}, \qquad b_2 = \frac{1}{k}\frac{\sqrt{a_0} - \sqrt{a_0 + ka_2}}{\sqrt{a_0^2 + ka_0a_2}},$$

for $\rho \geq 0, \kappa \geq 1$.

1. Show that $\frac{\mathrm{d}a_0}{\mathrm{d}\rho} \leq 0$ and $\frac{\mathrm{d}a_2}{\mathrm{d}\rho} \geq 0$.
2. Show that $b_2 \leq 0$, and that $\frac{\mathrm{d}b_0}{\mathrm{d}\rho} \geq 0$ and $\frac{\mathrm{d}b_2}{\mathrm{d}\rho} \leq 0$.
3. Let $c = b_0 + kb_2$. Show that $c > 0$ and $\frac{\mathrm{d}c}{\mathrm{d}\rho} \leq 0$.
4. Show that
$$\lim_{\rho \to 1^-} b_0 = \infty.$$
5. Show that the z in Equation 5.8 goes to ∞ as $\rho \to 1$.
6. Show that the summands of Equation 5.18 take the form
$$b_0\left(\hat{\zeta}_i - \bar{\zeta}\right) + \left(b_0 + b_2k\right)\left(\bar{\zeta} - \zeta_0\right).$$
Show that the $\bar{x}^2$ of Equation 5.18 goes to ∞ as $\rho \to 1$.
7. Show that Follman's statistic, the g^2 of Equation 5.20, goes to ∞ as $\rho \to 1$.

Ex. 5.9 Follman's test, half-good case In Example 5.1.13, we saw that Follman's test achieved maximum power of around 1/2 for the half-good case.

1. Why does this happen?
2. Suppose you expected that 1/4 of the assets' signal-noise ratios were greater than ζ_0. How would you adapt Follman's test to have higher power under this alternative?

Ex. 5.10 Bonferroni Correction, Direct Estimation Suppose that $\boldsymbol{y} = \left(b_0\,\mathsf{I} + b_1\left(\mathbf{1}\mathbf{1}^\top\right)\right)\boldsymbol{z}$, where $\boldsymbol{z} \sim \mathcal{N}\left(\mathbf{0}, \mathsf{I}\right)$. In this exercise you will construct an approximate CDF for the maximum element $y_{(k)}$.

1. Show empirically that the correlation between $z_{(k)}$ and $\mathbf{1}^\top\boldsymbol{z}$ goes to zero for large k.
2. Suppose that $\boldsymbol{y} = b_0\boldsymbol{z} + b_1kz_0$, where $z_0 \sim \mathcal{N}\left(0, 1\right)$ is independent of $\boldsymbol{z} \sim \mathcal{N}\left(\mathbf{0}, \mathsf{I}\right)$. Show that
$$\Pr\left\{y_{(k)} \leq t\right\} = \int \Phi^k\left(\frac{t - b_1kx}{b_0}\right)\phi\left(x\right)\mathrm{d}x$$

3. Implement code to approximately compute that integral given k, b_0, and b_1. Use a basic quadrature scheme like Trapezoid rule.
4. Test your code empirically: spawn such a $\boldsymbol{y}$ for large k, compute $y_{(k)}$, compute the probability that you would see a value so large using your code, and repeat a thousand times. Are your putative p-values approximately uniform?

Ex. 5.11 Bonferroni Correction, Bounded Correlation Repeat the experiment of Example 5.1.6:

1. Repeat the experiment, but assume that the correlation of returns is R with

$$\mathsf{R}_{i,j} = \begin{cases} \rho^{|i-j|} & \text{if } |i-j| \leq 1, \\ \frac{\rho}{2} & \text{otherwise.} \end{cases}$$

2. Repeat the experiment assuming that

$$\mathsf{R} = (1-\rho)\,\mathsf{I} + \rho\left(\boldsymbol{q}\boldsymbol{q}^{\top}\right),$$

where exactly half the elements of $\boldsymbol{q}$ are -1 and the other half are $+1$.

Ex. 5.12 Slepian Bounds Suppose that $\boldsymbol{x} \sim \mathcal{N}\left(\mathbf{0}, (1-\rho)\,\mathsf{I} + \rho\left(\boldsymbol{q}\boldsymbol{q}^{\top}\right)\right)$, and $\boldsymbol{y} \sim \mathcal{N}\left(\mathbf{0}, (1-\rho)\,\mathsf{I} + \rho\left(\mathbf{1}\mathbf{1}^{\top}\right)\right)$, for some $\rho > 0$, where k of the k elements of $\boldsymbol{q}$ are -1 and the rest are $+1$.
From the Li and Shao bound of Equation 5.9, prove that

$$\Pr\left\{\max_i x_i > t\right\} \leq \Pr\left\{\max_i y_i > t\right\} + \frac{k\,(k-k)\,\operatorname{asin}\rho}{\pi} e^{-t^2/(1+\rho)}. \tag{5.31}$$

Ex. 5.13 Conditional Confidence Interval, Opportunistic Strategy Consider how large a Sharpe ratio you have to observe such that a lower one-sided conditional confidence interval is exactly zero.

1. Show that $[0, \infty]$ is a conditional $1-\alpha$ interval for the signal-noise ratio of the opportunistic strategy exactly when

$$F_{SR}\left(\hat{\zeta}; 0, n\right) = 1 - \frac{\alpha}{2}.$$

2. Confirm that $[0, \infty]$ is an unconditional $1-\alpha$ interval for the signal-noise ratio of an asset exactly when

$$F_{SR}\left(\hat{\zeta}; 0, n\right) = 1 - \alpha.$$

Ex. 5.14 Overoptimism by opportunistic selection Consider a mashup of overoptimism by selection and the opportunistic strategy. That is, suppose you observe the Sharpe ratios of k strategies, then select the asset with the largest *absolute* Sharpe ratio, which you will hold long or short depending on the sign of its Sharpe ratio. How would you use Theorem 5.1.11 for inference on this problem?

Ex. 5.15 BC and HSD bounds Compare the *HSD* and *BC* cutoffs from Equation 5.27 and Equation 5.30. Compute both for the $\alpha = 0.05$ level, $n = 504\,\text{day}$, $\rho = 0.8$ and vary the number of assets, k from 4 to 200. Plot both cutoffs.

* **Ex. 5.16 Research Problem: shrinkage after selection** §
Hwang describes an estimator of the mean of selected populations for the case of independent multivariate normal errors that has reduced Bayesian Risk. [73, 52] This could easily be used to describe the problem of Bayesian inference on the signal-noise ratio of the asset with maximal Sharpe ratio, via the normal approximation of Equation 4.29. However, it would only apply for $\mathsf{R} = \sigma^2\mathsf{I}$. Generalize Hwang's procedure to the case of the rank-one correlation matrix of Equation 5.4, $\mathsf{R} = (1-\rho)\,\mathsf{I} + \rho\left(\mathbf{1}\mathbf{1}^{\top}\right)$.

* **Ex. 5.17 Research Problem: CIs on selected signal-noise ratios** §
Fuentes, Casella and Wells describe a procedure for computing confidence intervals on the means of selected populations for the case where errors are independent and homoskedastic. [53] This could easily be used to compute confidence intervals on the signal-noise ratios of, say, the top p assets as selected by Sharpe ratio, via the normal approximation of Equation 4.29 for the case where $\mathsf{R} = \sigma^2\mathsf{I}$. Generalize this procdure to the case of the rank-one correlation matrix of Equation 5.4, $\mathsf{R} = (1-\rho)\,\mathsf{I} + \rho\left(\mathbf{1}\mathbf{1}^{\top}\right)$.

A. Glossary

> Mathematics, a veritable sorcerer in our computerized society, while assisting the trier of fact in the search for truth, must not cast a spell over him.
>
> *(Supreme Court of California,* People v. Collins*)*

> *Trust but verify*
>
> *(Russian Proverb)*

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.

Σ The population covariance matrix of p assets.

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

ζ_* The SNR of $\boldsymbol{\nu}_*$.

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

$\hat{\Sigma}$ 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: $\hat{\boldsymbol{\nu}}_* =_{\mathrm{df}} \hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}$.

$\hat{\zeta}_*$ 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.

μ_3 the skew of a random variable.

μ_4 the excess kurtosis of a random variable.

κ the kurtosis factor of an elliptical distribution. This is one third the (regular, not excess) kurtosis of the marginals.

α_i the i^{th} uncentered moment of a random variable, $\mathrm{E}\left[(x)^i\right]$.

$\hat{\alpha}_i$ the i^{th} uncentered sample moment of a sample, typically computed as the simple mean of the observations to the i^{th} power.

μ_i the i^{th} centered moment of a random variable, defined as $\mathrm{E}\left[(x - \mathrm{E}\left[x\right])^i\right]$.

$\hat{\mu}_i$ the i^{th} centered sample moment of a sample.

κ_i the $i + 2^{\text{th}}$ raw cumulant of a random variable.

γ_i the $i + 2^{\text{th}}$ standardized cumulant of a random variable, equal to the $i + 2^{\text{th}}$ raw cumulant divided by σ^{i+2}.

γ_1 the skew of a random variable.

γ_2 the excess kurtosis of a random variable.

B. Miscellanea

Example B.0.1 (Notebook for Homoscedastic, independent returns and Sharpe ratio standard error). This is the sympy notebook for Example 4.4.2:

```
In [1]: from __future__ import division
        from sympy import *
        from sympy.physics.quantum import TensorProduct
        from sympy.assumptions.assume import global_assumptions
        init_printing()
        # set up symbols
        mu, zeta, r0 = symbols('\mu \zeta r_0')
        sigma = symbols('\sigma',positive=True)
        r0 = 0.0
        mu = sigma * zeta + r0
        # needed matrices
        Elim = eye(4)[[0,1,3],:]
        Komm = eye(4)[[0,2,1,3],:]
        vmu  = Matrix(2,1,[mu,1])
        vsig = Matrix(2,2,[sigma**2,0,0,0])

        Theta = Matrix(2,2,[sigma**2 + mu**2,mu,mu,1])
        iTheta = Theta.inv()
        ihTheta = iTheta.cholesky()
        simplify(ihTheta)
```

Out[1]:

$$\begin{bmatrix} \frac{1}{\sigma} & 0 \\ -\zeta & 1 \end{bmatrix}$$

```
In [2]: EThetahat = vmu * vmu.transpose() + vsig
        EThetahat
```

Out[2]:

$$\begin{bmatrix} \sigma^2\zeta^2 + \sigma^2 & \sigma\zeta \\ \sigma\zeta & 1 \end{bmatrix}$$

```
In [3]: Vpart1 = TensorProduct(vmu*vmu.transpose(),vsig) + TensorProduct(vsig,vmu*vmu.transpose()) + \
          TensorProduct(vsig,vsig)
        VThetahat = (eye(4) + Komm) * Vpart1
        LVThetahat = Elim * VThetahat * Elim.transpose()
        simplify(LVThetahat)
```

Out[3]:

$$\begin{bmatrix} \sigma^4\left(4\zeta^2 + 2\right) & 2\sigma^3\zeta & 0 \\ 2\sigma^3\zeta & \sigma^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

```
In [4]: InDeriv0 = TensorProduct(Theta,Theta) * (eye(4)+Komm) * TensorProduct(ihTheta,eye(2))
        InDeriv  = Elim * InDeriv0 * Elim.transpose()

        e1r0v = Matrix(3,1,[r0,1,0])
        DerivPart = (e1r0v).transpose() * InDeriv.inv()
        simplify(DerivPart)
```

Out [4]:

$$\left[-\frac{\zeta}{2\sigma^2} \quad \frac{1}{\sigma}\left(\zeta^2+1\right) \quad -\frac{\zeta^3}{2}-\zeta\right]$$

```
In [5]: vcov = DerivPart * (Elim * VThetahat * (Elim.transpose())) * DerivPart.transpose()
        simplify(vcov)
```

Out [5]:

$$\left[\frac{\zeta^2}{2}+1\right]$$

⊣

Example B.0.2 (Notebook for Homoscedastic, independent returns and ex-factor Sharpe ratio standard error with a Market). This is the sympy notebook to confirm Equation 4.71 first in the Gaussian case via Equation 4.65, and then in the *i.i.d.* Elliptical case from Equation 4.69.

```
In [1]: from __future__ import division
        from sympy import *
        from sympy.physics.quantum import TensorProduct
        from sympy.assumptions.assume import global_assumptions
        init_printing()
        mu, zetah, mzeta, beta, mmu, rho, r0 = symbols('\mu \zeta_h \zeta_m \\beta \mu_m \rho r_0')
        kurty = symbols('\kappa',positive=True)
        sigma = symbols('\sigma',positive=True)
        msigma = symbols('\sigma_m',positive=True)
        r0 = 0.0
        mu = sigma * zetah + r0
        mmu = msigma * mzeta

In [2]: Elim = eye(9)[[0,1,2,4,5,8],:]
        Komm = eye(9)[[0,3,6,1,4,7,2,5,8],:]
        twoSimm = (eye(9) + Komm)

        vmu = Matrix(3,1,[beta*mmu + mu,1,mmu])
        betvec = Matrix(2,1,[mu,beta])
        gammamat = Matrix(2,2,[1,mmu,mmu,mmu**2 + msigma**2])

        Theta = zeros(3,3)
        Theta[0,0] = Matrix([(sigma**2)]) + (betvec.transpose() * (gammamat * betvec))
        Theta[1:3,0] = gammamat * betvec
        Theta[0,1:3] = betvec.transpose() * gammamat
        Theta[1:3,1:3] = gammamat
        vsig = Theta - vmu * vmu.transpose()

        iTheta = Theta.inv()
        ihTheta = iTheta.cholesky()
        ihTheta

        # take Gamma and invert, cholesky, transpose, and invert:
        Gam_icti = Matrix(2,2,[1/sqrt(1 + mzeta**2),mzeta/sqrt(1+mzeta**2),0,msigma*sqrt(1+mzeta**2)])
        Theta_icti = zeros(3,3)
```

```
Theta_icti[0,0] = sigma
Theta_icti[0,1:3] = betvec.transpose() * Gam_icti
Theta_icti[1:3,1:3] = Gam_icti
Theta_icti

# confirm we have computed Theta_icti correctly:
Foo = Theta_icti.inv().transpose()
Foo2 = Foo * Foo.transpose()
Foo3 = Foo2.inv()
Foo4 = Foo3 - Theta
simplify(Foo4)
```

Out[2]:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

```
In [3]: EThetahat = simplify(vmu * vmu.transpose() + vsig)
        EThetahat
```

Out[3]:

$$\begin{bmatrix} \beta\sigma_m\left(\beta\sigma_m\left(\zeta_m^2+1\right)+\sigma\zeta_h\zeta_m\right)+\sigma^2+\sigma\zeta_h\left(\beta\sigma_m\zeta_m+\sigma\zeta_h\right) & \beta\sigma_m\zeta_m+\sigma\zeta_h & \sigma_m\left(\beta\sigma_m\left(\zeta_m^2+1\right)+\sigma\zeta_h\zeta_m\right) \\ \beta\sigma_m\zeta_m+\sigma\zeta_h & 1 & \sigma_m\zeta_m \\ \sigma_m\left(\beta\sigma_m\left(\zeta_m^2+1\right)+\sigma\zeta_h\zeta_m\right) & \sigma_m\zeta_m & \sigma_m^2\left(\zeta_m^2+1\right) \end{bmatrix}$$

```
In [4]: InDeriv0 = twoSimm * TensorProduct(Theta_icti,Theta)
        InDeriv1 = Elim * InDeriv0 * Elim.transpose()
        e1r0v = Matrix(6,1,[r0,1,0, 0,0,0])
        DerivPart = (e1r0v).transpose() * InDeriv1.inv()
        sdp = simplify(DerivPart)

In [5]: # Omega here comes from Gaussian Isserlis equation
        Vpart1 = TensorProduct(vmu*vmu.transpose(),vsig) + TensorProduct(vsig,vmu*vmu.transpose()) + \
          TensorProduct(vsig,vsig)
        VThetahat = twoSimm * Vpart1
        LVThetahat = simplify(Elim * VThetahat * Elim.transpose())

        vcov = sdp * LVThetahat * sdp.transpose()
        simplify(vcov)
```

Out[5]:

$$\left[\frac{\zeta_h^2}{2}+\zeta_m^2+1\right]$$

```
In [6]: # Omega here comes from Independent Elliptical equation
        uu = vmu * vmu.transpose()
        vector_sigma = Matrix(9,1,vsig)

        Vpart1 = (kurty-1) * simplify(vector_sigma * vector_sigma.transpose() + twoSimm * TensorProduct(vsig,vsig))
        Vpart2 = twoSimm * (simplify(TensorProduct(Theta,Theta) - TensorProduct(uu,uu)))
        VThetahat = simplify(Vpart1 + Vpart2)
        LVThetahat = simplify(Elim * VThetahat * Elim.transpose())

        vcov = sdp * LVThetahat * sdp.transpose()
        simplify(vcov)
```

Out[6]:

$$\left[\frac{3\kappa}{4}\zeta_h^2+\kappa\zeta_m^2-\frac{\zeta_h^2}{4}+1\right]$$

⊣

Bibliography

Good artists borrow. Great artists steal.

(Variously attributed)

[1] Bet 362. Privately Published, 2007. URL http://longbets.org/362/.

[2] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition, 1964. URL http://people.math.sfu.ca/~cbm/aands/toc.htm.

[3] Masafumi Akahira. A higher order approximation to a percentage point of the noncentral t-distribution. *Communications in Statistics - Simulation and Computation*, 24(3):595–605, 1995. doi: 10.1080/03610919508813261. URL http://www.tandfonline.com/doi/abs/10.1080/03610919508813261.

[4] Masafumi Akahira, Michikazu Sato, and Norio Torigoe. On the new approximation to non-central t-distributions. *Journal of the Japan Statistical Society*, 25(1):1–18, 1995. URL https://www.jstage.jst.go.jp/article/jjss1995/25/1/25_1_1/_pdf.

[5] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. Wiley, 2003. ISBN 9780471360919. URL http://books.google.com/books?id=Cmm9QgAACAAJ.

[6] Albert Ando and G. M. Kaufman. Bayesian analysis of the independent multinormal process. neither mean nor precision known. *Journal of the American Statistical Association*, 60(309):pp. 347–358, 1965. ISSN 01621459. URL http://www.jstor.org/stable/2283159.

[7] David R. Aronson. *Evidence-Based Technical Analysis: Applying the Scientific Method and Statistical Inference to Trading Signals*. Wiley, Hoboken, NJ, nov 2006. URL http://www.evidencebasedta.com/.

[8] Cliff Asness. Our model goes to six *and* saves value from redundancy along the way. Privately Published, December 2014. URL https://www.aqr.com/cliffs-perspective/our-model-goes-to-six-and-saves-value-from-redundancy-along-the-way.

[9] Clifford S. Asness, Andrea Frazzini, and Lasse Heje Pedersen. Quality minus junk. Privately Published, October 2013. URL http://ssrn.com/abstract=2312432.

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

[11] L. J. Bain and M. Engelhardt. *Introduction to Probability and Mathematical Statistics*. Classic Series. Cengage Learning, 1992. ISBN 9780534380205. URL http://books.google.com/books?id=MkFRIAAACAAJ.

[12] Yong Bao. Estimation risk-adjusted Sharpe ratio and fund performance ranking under a general return distribution. *Journal of Financial Econometrics*, 7(2): 152–173, 2009. doi: 10.1093/jjfinec/nbn022. URL https://doi.org/10.1093/jjfinec/nbn022.

[13] W. Baumgartner, P. Weiss, and H. Schindler. A nonparametric test for the general two-sample problem. *Biometrics*, 54(3):1129–1135, 1998.

[14] 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.

[15] Roger L. Berger. Likelihood ratio tests and intersection-union tests. Technical Report 2288, Department of Statistics, North Carolina State University, September 1996. URL http://www.stat.ncsu.edu/information/library/mimeo.archive/ISMS_1996_2288.pdf. Institute of Statistics, Mimeo Series.

[16] 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.

[17] 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.

[18] 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.

[19] Jacob Boudoukh, Ronen Israel, and Matthew P. Richardson. Long Horizon Predictability: A Cautionary Tale. *SSRN eLibrary*, 2018. doi: 10.2139/ssrn.3142575. URL https://ssrn.com/paper=3142575.

[20] Kris Boudt, Brian Peterson, and Christophe Croux. Estimation and decomposition of downside risk for portfolios with non-normal returns, 2008. URL http://ssrn.com/abstract=1024151.

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

[22] Frank Bretz, Torsten Hothorn, and Peter Westfall. *Multiple comparisons using R*. Chapman and Hall/CRC, 2016. URL http://www.ievbras.ru/ecostat/Kiril/R/Biblio_N/R_Eng/Bretz2011.pdf.

[23] Mark Britten-Jones and Anthony Neuberger. Improved inference and estimation in regression with overlapping observations, 2007. URL http://www2.warwick.ac.uk/fac/soc/wbs/subjects/finance/faculty1/anthony_neuberger/improved.pdf.

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

[25] Warren E. Buffett. Berkshire Hathaway shareholder letter 2016. Privately Published, February 2017. URL http://www.berkshirehathaway.com/letters/2016ltr.pdf.

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

[27] CBOE. The CBOE volatility index - VIX. Privately Published, 2009. URL http://www.cboe.com/micro/vix/vixwhite.pdf.

[28] CBOE. Vix index. Privately Published, 2014. URL http://www.quandl.com/YAHOO/INDEX_VIX.

[29] CBOE. CBOE VIX data. Privately Published, 2017. URL http://www.cboe.com/products/vix-index-volatility/vix-options-and-futures/vix-index/vix-historical-data.

[30] Damien Challet. Sharper asset ranking from total drawdown durations, 2015. URL http://arxiv.org/abs/1505.01333.

[31] R. Chattamvelli and R. Shanmugam. An enhanced algorithm for noncentral t-distribution. *Journal of Statistical Computation and Simulation*, 49(1-2):77–83, 1994. doi: 10.1080/00949659408811561. URL http://www.tandfonline.com/doi/abs/10.1080/00949659408811561.

[32] Long Chen, Robert Novy-Marx, and Lu Zhang. An alternative three-factor model. Privately Published, April 2011. URL http://ssrn.com/abstract=1418117.

[33] Victor Chernozhukov, Iván Fernández-Val, and Alfred Galichon. Rearranging Edgeworth-Cornish-Fisher expansions. Privately Published, 2007. URL http://arxiv.org/abs/0708.1627.

[34] Gregory C. Chow. Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28(3):pp. 591–605, 1960. ISSN 00129682. URL http://www.jstor.org/stable/1910133.

[35] John Howland Cochrane. *Asset pricing*. Princeton Univ. Press, Princeton [u.a.], 2001. ISBN 0691074984. URL http://press.princeton.edu/titles/7836.html.

[36] Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001. doi: 10.1080/713665670. URL http://personal.fmipa.itb.ac.id/khreshna/files/2011/02/cont2001.pdf.

[37] James M. Dickey. Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. *The Annals of Mathematical Statistics*, 38(2):511–518, 04 1967. doi: 10.1214/aoms/1177698967. URL http://dx.doi.org/10.1214/aoms/1177698967.

[38] DLMF. Digital library of mathematical functions, 2011. URL http://dlmf.nist.gov/5.11#iii. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, eds.

[39] D. Duffie. *Dynamic Asset Pricing Theory*. Princeton Series in Finance. Princeton University Press, 3rd edition, 2010. ISBN 9781400829200. URL https://books.google.com/books?id=f2Wv-LDpsoUC.

[40] Dirk Eddelbuettel and Romain François. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18, 2011. URL http://www.jstatsoft.org/v40/i08/.

[41] Edwin J. Elton, Martin J. Gruber, and Christopher R. Blake. Fundamental economic variables, expected returns, and bond fund performance. *The Journal of Finance*, 50(4):pp. 1229–1256, 1995. ISSN 00221082. URL http://www.jstor.org/stable/2329350.

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

[43] 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.

[44] Eugene F. Fama and Kenneth R. French. Dissecting Anomalies with a Five-Factor Model. *SSRN eLibrary*, 2014. doi: 10.2139/ssrn.2503174. URL http://ssrn.com/paper=2503174.

[45] Mark Finkelstein and Edward O Thorp. Nontransitive dice with equal means. *Optimal Play: Mathematical Studies in Games and Gambling*, pages 293–310, 2000. URL http://www.math.uci.edu/~mfinkels/dice9.pdf.

[46] Ronald Aylmer Fisher. The sampling error of estimated deviates, together with other illustrations of the properties and applications of the integrals and derivatives of the normal error function. In *Introduction to British A.A.S. Math. Tables, I.* 1931. URL http://hdl.handle.net/2440/15209.

[47] Ronald Aylmer Fisher. *Properties and applications of* Hh *functions*, pages xxvi–xxxv. 1931. URL http://scholar.google.com/scholar?cites=8550436429828449531&as_sdt=4005&sciodt=4000&hl=en.

[48] Dean Follmann. A simple multivariate test for one-sided alternatives. *Journal of the American Statistical Association*, 91(434):854–861, 1996. ISSN 01621459. URL http://www.jstor.org/stable/2291680.

[49] Andrea Frazzini, David Kabiller, and Lasse Heje Pedersen. Buffett's alpha. Privately Published, 2012. URL http://www.econ.yale.edu/~af227/pdf/Buffett%27s%20Alpha%20-%20Frazzini,%20Kabiller%20and%20Pedersen.pdf.

[50] Kenneth French. Data library. Privately Published, 2017. URL http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[51] Kenneth French. 5 industry portfolios. Privately Published, 2019. URL http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_5_ind_port.html.

[52] Claudio Fuentes and Vik Gopal. A constrained conditional likelihood approach for estimating the means of selected populations, 2017. URL http://arxiv.org/abs/1702.07804. cite arxiv:1702.07804.

[53] Claudio Fuentes, George Casella, and Martin T. Wells. Confidence intervals for the means of the selected populations. *Electron. J. Statist.*, 12(1):58–79, 2018. doi: 10.1214/17-EJS1374. URL https://doi.org/10.1214/17-EJS1374.

[54] M. Gardner. *The Colossal Book of Mathematics: Classic Puzzles, Paradoxes, and Problems : Number Theory, Algebra, Geometry, Probability, Topology, Game Theory, Infinity, and Other Topics of Recreational Mathematics*. Norton, 2001. ISBN 9780393020236. URL https://books.google.com/books?id=orz0SDEakpYC.

[55] 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.

[56] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013. ISBN 9781439840955. URL http://books.google.com/books?id=ZXL6AQAAQBAJ.

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

[58] 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*, oct 2010. URL http://arxiv.org/abs/1010.2265.

[59] Georg M. Goerg. *LambertW: An R package for Lambert W x F Random Variables*, 2014. R package version 0.5.

[60] 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".

[61] R. Grinold and R. Kahn. *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Selecting Superior Returns and Controlling Risk*. McGraw-Hill Library of Investment and Finance. McGraw-Hill Education, 1999. ISBN 9780070248823. URL http://books.google.com/books?id=a1yB8LTQn0EC.

[62] William C. Guenther. Evaluation of probabilities for the noncentral distributions and the difference of two t-variables with a desk calculator. *Journal of Statistical Computation and Simulation*, 6(3-4):199–206, 1978. doi: 10.1080/00949657808810188. URL http://dx.doi.org/10.1080/00949657808810188.

[63] Peter Hall. Chi squared approximations to the distribution of a sum of independent random variables. *The Annals of Probability*, 11(4):pp. 1028–1036, 1983. ISSN 00911798. URL http://www.jstor.org/stable/2243514.

[64] P. R. Hansen. A test for superior predictive ability. *Journal of Business and Economic Statistics*, 23(4), 2005. doi: 10.1198/073500105000000063. URL http://pubs.amstat.org/doi/abs/10.1198/073500105000000063.

[65] Peter Reinhard Hansen. Asymptotic tests of composite hypotheses. Working Paper 2003-09, Brown University, Department of Economics, Providence, RI, 2003. URL http://hdl.handle.net/10419/80222.

[66] Campbell R Harvey and Yan Liu. Backtesting. *The Journal of Portfolio Management*, 42(1):13–28, 2015. URL https://www.cmegroup.com/education/files/backtesting.pdf.

[67] Joel Hasbrouck. *Empirical Market Microstructure:The Institutions, Economics, and Econometrics of Securities Trading: The Institutions, Economics, and Econometrics of Securities Trading*. Oxford University Press, USA, 2007. ISBN 9780195301649. URL http://books.google.com/books?id=KIEHmQEACAAJ.

[68] 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.

[69] G. W. Hill. Algorithm 396: Student's t-quantiles. *Commun. ACM*, 13(10): 619–620, oct 1970. ISSN 0001-0782. doi: 10.1145/355598.355600. URL http://doi.acm.org/10.1145/355598.355600.

[70] G. W. Hill. Remark on "Algorithm 396: Student's t-quantiles [s14]". *ACM Trans. Math. Softw.*, 7(2):250–251, jun 1981. ISSN 0098-3500. doi: 10.1145/355945.355956. URL http://doi.acm.org/10.1145/355945.355956.

[71] S. Hodges and Financial Options Research Centre. *A Generalization of the Sharpe Ratio and Its Applications to Valuation Bounds and Risk Measures*. FORC preprint: Financial Options Research Centre. Financial Options Research Centre, Warwick Business School, University of Warwick, 1998. URL http://www2.warwick.ac.uk/fac/soc/wbs/subjects/finance/research/wpaperseries/1998/98-88.pdf.

[72] D. Hogben, R. S. Pinkham, and M. B. Wilk. The moments of the non-central t-distribution. *Biometrika*, 48(3/4):pp. 465–468, 1961. ISSN 00063444. URL http://www.jstor.org/stable/2332772.

[73] Jiunn T. Hwang. Empirical Bayes estimation for the means of the selected populations. *Sankkhyâ: The Indian Journal of Statistics*, 55:285–304, 1993. URL https://www.jstor.org/stable/25050934.

[74] L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12 (1/2):pp. 134–139, 1918. ISSN 00063444. URL http://www.jstor.org/stable/2331932.

[75] Stefan R. Jaschke. The Cornish-Fisher-expansion in the context of Delta - Gamma - Normal approximations. Technical Report 2001,54, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, 2001. URL http://www.jaschke-net.de/papers/CoFi.pdf.

[76] 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.

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

[78] Zura Kakushadze and Willie Yu. Notes on fano ratio and portfolio optimization, 2017. URL http://arxiv.org/abs/1711.10640.

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

[80] Jongphil Kim and Anthony J. Hayter. Efficient confidence interval methodologies for the non centrality parameter of a non central t distribution. *Communications in Statistics - Simulation and Computation*, 37(4):660–678, 2008. doi: 10.1080/03610910701739605. URL http://www.tandfonline.com/doi/abs/10.1080/03610910701739605.

[81] S. Kotz and S. Nadarajah. *Multivariate t-Distributions and Their Applications*. Cambridge University Press, 2004. ISBN 9780521826549. URL http://drsmorey.org/bibtex/upload/Kotz:Nadarajah:2004.pdf.

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

[83] K. Krishnamoorthy. *Handbook of Statistical Distributions with Applications*. Statistics: A Series of Textbooks and Monographs. CRC Press, 2006. ISBN 9781420011371. URL http://www.crcnetbase.com/isbn/9781420011371.

[84] William Kruskal. The monotonicity of the ratio of two noncentral t density functions. *The Annals of Mathematical Statistics*, 25(1):pp. 162–165, 1954. ISSN 00034851. URL http://www.jstor.org/stable/2236523.

[85] A. M. Kshirsagar. Some extensions of the multivariate t-distribution and the multivariate generalization of the distribution of the regression coefficient. *Mathematical Proceedings of the Cambridge Philosophical Society*, 57(01):80–85, 1 1961. ISSN 1469-8064. doi: 10.1017/S0305004100034885. URL http://journals.cambridge.org/article_S0305004100034885.

[86] Daniel Lakens. Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses. 2016. doi: 10.1177/1948550617697177. URL https://osf.io/preprints/psyarxiv/97gpc/.

[87] Bruno Lecoutre. Two useful distributions for Bayesian predictive procedures under normal models. *Journal of Statistical Planning and Inference*, 79:93–105, 1999. URL http://www.researchgate.net/publication/242997315_Two_useful_distributions_for_Bayesian_predictive_procedures_under_normal_models/file/5046352b9cba661c83.pdf.

[88] 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.

[89] 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.

[90] Jason D. Lee, Dennis L. Sun, Yuekai Sun, and Jonathan E. Taylor. Exact post-selection inference, with application to the lasso, 2013. URL http://arxiv.org/abs/1311.6238. cite arxiv:1311.6238 Comment: Published at http://dx.doi.org/10.1214/15-AOS1371 in the Annals of Statistics (http://www.imstat.org/aos/) by the Institute of Mathematical Statistics (http://www.imstat.org).

[91] Yoong-Sin Lee and Ting-Kwong Lin. Algorithm AS 269: High order Cornish-Fisher expansion. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(1):pp. 233–240, 1992. ISSN 00359254. URL http://www.jstor.org/stable/2347649.

[92] 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.

[93] Russell V. Lenth. Algorithm AS 243: Cumulative distribution function of the non-central t distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 38(1):pp. 185–189, 1989. ISSN 00359254. URL http://www.jstor.org/stable/2347693.

[94] Russell V. Lenth. Statistical algorithms: Algorithm AS 243: Cumulative distribution function of the non-central t distribution. *Journal of Applied Statistics*, 38 (1):185–189, March 1989. ISSN 0035-9254 (print), 1467-9876 (electronic). URL http://lib.stat.cmu.edu/apstat/243.

[95] Gérard Letac and Hélène Massam. A tutorial on the non central Wishart distribution. Privately Published, 2004. URL http://www.math.univ-toulouse.fr/~letac/Wishartnoncentrales.pdf.

[96] 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.

[97] Wenbo V Li and Qi-Man Shao. A normal comparison inequality and its applications. *Probability Theory and Related Fields*, 122(4):494–508, 2002.

URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.147.1513&rep=rep1&type=pdf.

[98] 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.

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

[100] Marcos López de Prado and David H. Bailey. The false strategy theorem: A financial application of experimental mathematics. *American Mathematical Monthly, forthcoming*, 2018.

[101] Xingguo Luo and Jin E. Zhang. The term structure of vix. *Journal of Futures Markets*, 32(12):1092–1123, 2012. ISSN 1096-9934. doi: 10.1002/fut.21572. URL http://dx.doi.org/10.1002/fut.21572.

[102] Hal Lux. The secret world of Jim Simons. *Institutional Investor*, 34(11), Nov 2000.

[103] 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.

[104] Jan R. Magnus and H. Neudecker. The elimination matrix: Some lemmas and applications. *SIAM Journal on Algebraic Discrete Methods*, 1(4):422–449, dec 1980. doi: 10.1137/0601049. URL http://www.janmagnus.nl/papers/JRM008.pdf.

[105] Jan R. Magnus and H. Neudecker. Symmetry, 0-1 matrices and Jacobians. *Econometric Theory*, 2:157–190, 1986. URL http://www.janmagnus.nl/papers/JRM014.pdf.

[106] 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.

[107] Burton G. Malkiel. The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, 17(1):59–82, 2003. doi: 10.1257/089533003321164958. URL http://eml.berkeley.edu/~craine/EconH195/Fall_14/webpage/Malkiel_Efficient%20Mkts.pdf.

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

[109] 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.

[110] 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.

[111] 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.

[112] 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 https://doi.org/10.2307/2330636.

[113] C. Munk. *Financial Asset Pricing Theory*. OUP Oxford, 2nd edition, 2013. ISBN 9780199585496. URL https://books.google.com/books?id=3WVs9pFkimIC.

[114] Kevin P. Murphy. Conjugate Bayesian analysis of the Gaussian distribution. Privately Published, 2007. URL http://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf.

[115] Daniel B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):pp. 347–370, 1991. ISSN 00129682. URL http://finance.martinsewell.com/stylized-facts/distribution/Nelson1991.pdf.

[116] R. E. Odeh and J. O. Evans. Algorithm AS 70: The percentage points of the normal distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 23(1):96–97, 1974. ISSN 00359254, 14679876. doi: 10.2307/2347061. URL http://www.jstor.org/stable/2347061.

[117] 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.

[118] 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.

[119] 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.

[120] D. B. Owen. Survey of properties and applications of the noncentral t-distribution. *Technometrics*, 10(3):445–478, 1968. doi: 10.1080/00401706.1968.10490594. URL http://dx.doi.org/10.1080/00401706.1968.10490594.

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

[122] Steven E. Pav. *SharpeR: Statistical Significance of the Sharpe Ratio*, 2014. URL https://github.com/shabbychef/SharpeR. R package version 0.1501.

[123] Steven E. Pav. Bounds on portfolio quality. Privately Published, 2014. URL http://arxiv.org/abs/1409.5936.

[124] Steven E. Pav. *MarkowitzR: Statistical Significance of the Markowitz Portfolio*, 2015. URL https://github.com/shabbychef/MarkowitzR. R package version 0.1502.

[125] Steven E. Pav. Safety third: Roy's criterion and higher order moments. Privately Published, 2015. URL http://arxiv.org/abs/1506.04227.

[126] Steven E. Pav. Inference on the Sharpe ratio via the upsilon distribution. Privately Published, 2015. URL http://arxiv.org/abs/1505.00829.

[127] Steven E. Pav. *BWStest: Baumgartner Weiss Schindler Test of Equal Distributions*, 2016. URL https://CRAN.R-project.org/package=BWStest. R package version 0.2.0.

[128] Steven E. Pav. *aqfb.data: Data for the AQFB Book*, 2017. URL https://github.com/shabbychef/aqfb_data. R package version 0.1.1.

[129] Steven E. Pav. Conditional inference on the asset with maximum Sharpe ratio. Privately Published, 2019. URL http://arxiv.org/abs/1906.00573.

[130] Steven E. Pav. A *post hoc* test on the Sharpe ratio. Privately Published, 2019. URL http://arxiv.org/abs/1911.04090. code available from https://github.com/shabbychef/posthoc_sr.

[131] Steven E. Pav. The Sharpe ratio: Statistics and applications, 2021.

[132] 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.

[133] Michael D Perlman and Lang Wu. The emperors new tests. *Statistical Science*, 14(4):355–369, 1999. URL https://projecteuclid.org/euclid.ss/1009212517.

[134] Kaare Brandt Petersen and Michael Syskind Pedersen. The matrix cookbook, nov 2012. URL http://www2.imm.dtu.dk/pubdb/p.php?3274. Version 20121115.

[135] Brian G. Peterson and Peter Carl. *PerformanceAnalytics: Econometric tools for performance and risk analysis*, 2014. URL http://CRAN.R-project.org/package=PerformanceAnalytics. R package version 1.4.3541.

[136] Hans-Peter Piepho. An algorithm for a letter-based representation of all-pairwise comparisons. *Journal of Computational and Graphical Statistics*, 13(2):456–466, 2004.

[137] Jacques Poitevineau and Bruno Lecoutre. Implementing Bayesian predictive procedures: The K-prime and K-square distributions. *Computational Statistics and Data Analysis*, 54(3):724–731, 2010. URL http://arxiv.org/abs/1003.4890v1.

[138] 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.

[139] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL http://www.R-project.org/. ISBN 3-900051-07-0.

[140] 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.

[141] Adam Rej, Philip Seager, and Jean-Philippe Bouchaud. You are in a drawdown. when should you start worrying?, 2017. URL http://arxiv.org/abs/1707.01457.

[142] Igor Rivin. What is the sharpe ratio, and how can everyone get it wrong?, 2018. URL http://arxiv.org/abs/1802.04413.

[143] Stephen A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, 1976. URL http://linkinghub.elsevier.com/retrieve/pii/0022053176900466.

[144] A. D. Roy. Safety first and the holding of assets. *Econometrica*, 20(3):pp. 431–449, 1952. ISSN 00129682. URL http://www.jstor.org/stable/1907413.

[145] Jeffrey A. Ryan and Joshua M. Ulrich. *xts: eXtensible Time Series*, 2014. URL http://CRAN.R-project.org/package=xts. R package version 0.9-7.

[146] 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.

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

[148] Pranab Kumar Sen. Union-intersection principle and constrained statistical inference. *Journal of Statistical Planning and Inference*, 137(11):3741–3752, 2007. ISSN 0378-3758. doi: 10.1016/j.jspi.2007.03.046. URL http://www.sciencedirect.com/science/article/pii/S0378375807001310. Special Issue: In Celebration of the Centennial of The Birth of Samarendra Nath Roy (1906-1964).

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

[150] William F. Sharpe. Likely gains from market timing. *Financial Analysts Journal*, 31(2):pp. 60–69, 1975. ISSN 0015198X. URL http://www.jstor.org/stable/4477805.

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

[152] Kesar Singh, Minge Xie, and William E. Strawderman. Confidence distribution (cd) – distribution estimator of a parameter. In Regina Liu, William Strawderman, and Cun-Hui Zhang, editors, *Complex Datasets and Inverse Problems*, volume Volume 54 of *Lecture Notes–Monograph Series*, pages 132–150. Institute of Mathematical Statistics, Beachwood, Ohio, USA, 2007. doi: 10.1214/074921707000000102. URL http://dx.doi.org/10.1214/074921707000000102.

[153] David Slepian. The one-sided barrier problem for Gaussian noise. *Bell System Technical Journal*, 41(2):463–501, 1962. doi: 10.1002/j.1538-7305.1962.tb02419.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1538-7305.1962.tb02419.x.

[154] Kent Smetters and Xingtan Zhang. A sharper ratio: A general measure for correctly ranking non-normal investment risks. Working Paper 19500, National Bureau of Economic Research, October 2013. URL http://www.nber.org/papers/w19500.

[155] M. R. Spiegel and L. J. Stephens. *Schaum's Outline of Statistics*. Schaum's Outline Series. Mcgraw-hill, 2007. ISBN 9780071594462. URL http://books.google.com/books?id=qdcBmgs3N3AC.

[156] Stephen M. Stigler. Stigler's law of eponymy. *Transactions of the New York Academy of Sciences*, 39(1 Series II):147–157, 1980. ISSN 2164-0947. doi: 10.1111/j.2164-0947.1980.tb02775.x. URL http://dx.doi.org/10.1111/j.2164-0947.1980.tb02775.x.

[157] Edward J. Sullivan. A.D. Roy: The forgotten father of portfolio theory. *Research in the History of Economic Thought and Methodology*, 29:pp. 73–82, 2011. doi: 10.1108/S0743-4154(2011)000029A008. URL http://www.emeraldinsight.com/books.htm?chapterid=1943472.

[158] Dei-In Tang. Uniformly more powerful tests in a one-sided multivariate problem. *Journal of the American Statistical Association*, 89(427):pp. 1006–1011, 1994. ISSN 01621459. URL http://www.jstor.org/stable/2290927.

[159] N. H. Timm. *Applied multivariate analysis: methods and case studies*. Springer Texts in Statistics. Physica-Verlag, 2002. ISBN 9780387227719. URL http://amzn.to/TMdgaE.

[160] Vladimir V. Ulyanov. Cornish Fisher expansions. In Miodrag Lovric, editor, *International Encyclopedia of Statistical Science*, pages 312–315. Springer Berlin Heidelberg, 2014. ISBN 978-3-642-04897-5. doi: 10.1007/978-3-642-04898-2_193. URL http://dx.doi.org/10.1007/978-3-642-04898-2_193.

[161] Rossen Valkanov. Long-horizon regressions: theoretical results and applications. *Journal of Financial Economics*, 68(2):201 – 232, 2003. ISSN 0304-405X. doi: http://dx.doi.org/10.1016/S0304-405X(03)00065-5. URL http://www.sciencedirect.com/science/article/pii/S0304405X03000655.

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

[163] C. Vignat and S. Bhatnagar. An extension of Wick's theorem, 2007. URL http://arxiv.org/abs/0709.1999.

[164] Michael Vock. *"One-sided" Statistical Inference for a Multivariate Location Parameter*. PhD thesis, University of Bern, 2007. URL http://www.imsv.unibe.ch/e19451/e226662/e309213/e309217/e309218/files309221/Diss_Vock_3rd_eng.pdf.

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

[166] 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.

[167] Halbert White. A reality check for data snooping. *Econometrica*, 68:1097–1127, 2000. URL https://www.ssc.wisc.edu/~bhansen/718/White2000.pdf.

[168] whuber (https://stats.stackexchange.com/users/919/whuber). What is the distribution of a bivariate normal component conditional on the max of the other component? Cross Validated, 2019. URL https://stats.stackexchange.

com/q/416791. URL:https://stats.stackexchange.com/q/416791 (version: 2019-07-10).

[169] Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN 978-0-387-98140-6. URL http://had.co.nz/ggplot2/book.

[170] Hadley Wickham and Romain Francois. *dplyr: A Grammar of Data Manipulation*, 2015. URL http://CRAN.R-project.org/package=dplyr. R package version 0.4.1.

[171] 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].

[172] Viktor Witkovsky. A note on computing extreme tail probabilities of the noncentral t distribution with large noncentrality parameter. Privately Published, 2013. URL http://arxiv.org/abs/1306.5294.

[173] Frank A Wolak. An exact test for multiple inequality and equality constraints in the linear regression model. *Journal of the American Statistical Association*, 82(399):782–793, 1987. URL https://web.stanford.edu/group/fwolak/cgi-bin/sites/default/files/files/An%20Exact%20Test%20for%20Multiple%20Equality%20and%20Inequality%20Constraints%20in%20the%20Linear%20Regression%20Model_Wolak(1).pdf.

[174] John Alexander Wright, Sheung Chi Phillip Yam, and Siu Pang Yung. A test for the equality of multiple Sharpe ratios. *Journal of Risk*, 16 (4), 2014. URL http://www.risk.net/journal-of-risk/journal/2340044/latest-issue-of-the-journal-of-risk-volume-16-number-4-2014.

[175] Min-ge Xie and Kesar Singh. Confidence distribution, the frequentist distribution estimator of a parameter: A review. *International Statistical Review*, 81(1):3–39, 2013. ISSN 1751-5823. doi: 10.1111/insr.12000. URL http://dx.doi.org/10.1111/insr.12000.

[176] Yihui Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2015. URL http://yihui.name/knitr/. R package version 1.9.

[177] Chuancun Yin. Stochastic orderings of multivariate elliptical distributions, 2019. URL https://arxiv.org/abs/1910.07158.

[178] Valeri Zakamouline and Steen Koekebakker. Portfolio Performance Evaluation with Generalized Sharpe Ratios: Beyond the Mean and Variance. *SSRN eLibrary*, 2008. doi: 10.2139/ssrn.1028715. URL http://ssrn.com/paper=1028715.

[179] Ofer Zeitouni. GAUSSIAN FIELDS notes for lectures. unpublished notes, 2017. URL http://www.wisdom.weizmann.ac.il/~zeitouni/notesGauss.pdf.

[180] Jian Zhang. *Path-Dependence Properties of Leveraged Exchange-Traded Funds: Compounding, Volatility and Option Pricing*. PhD thesis, New York University, 2010. URL http://math.cims.nyu.edu/faculty/avellane/thesis_Zhang.pdf.

[181] Jin E. Zhang and Yingzi Zhu. VIX Futures. *Journal of Futures Markets*, 26(6): 521–531, 2006.

Index

Index I copy from old
Vladivostok telephone directory.

(Tom Lehrer, Lobachevsky*)*

Version History

1, 2, stupid.
3, 4, dumb.
5, 6, idiotic.
7, 8, seat bee sate.

(Negativland, Theme from "A Big 10-8 Place"*)*

- version 0.2.001, 2019-12-09; *Overoptimism.* Chapter on overoptimism.
- version 0.1.005, 2019-06-29; *Second SSRN Release.* Fama French factor data now through 2018. Adding more clearly marked hypothesis tests. Confidence intervals on the ex-factor Sharpe ratio. Fix mixed up references to equality of SR testing (F test vs chi-square). Material on prediction intervals for ex-factor Sharpe ratio.
- version 0.1.004, 2019-03-31; *Symmetric CIs.* Adding material about symmetric confidence intervals of Sharpe.
- version 0.1.003, 2018-10-06; *Higher Order Moments.* Adding material on the higher order estimates of bias and standard error of the Sharpe, due to Bao. Adding some material on elliptical distributions. Sourcing data from the **aqfb.data** package. Embedding data in the PDF document.
- version 0.1.002, 2017-10-18; *Reweighting.* More details earlier on inverse volatility reweighting, the "market clock," and "pessimistic models" of heteroskedasticity.
- version 0.1.001, 2017-08-26; *Hello World!* First release, warts and all.

GNU Free Documentation License

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or non-commercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "**Document**", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "**you**". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "**Modified Version**" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "**Secondary Section**" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "**Invariant Sections**" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "**Cover Texts**" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "**Transparent**" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "**Opaque**".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "**Title Page**" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "**publisher**" means any person or entity that distributes copies of the Document to the public.

A section "**Entitled XYZ**" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "**Acknowledgements**", "**Dedications**", "**Endorsements**", or "**History**".) To "**Preserve the Title**" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See `http://www.gnu.org/copyleft/`.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with ... Texts." line with this:

> with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.