Conditional inference on the asset with maximum Sharpe ratio

Steven E. Pav *

December 6, 2019

Abstract

We apply the procedure of Lee *et al.* [15] to the problem of performing inference on the signal-noise ratio of the asset which displays maximum sample Sharpe ratio over a set of possibly correlated assets. We find a multivariate analogue of the commonly used approximate standard error of the Sharpe ratio to use in this conditional estimation procedure. We also consider several alternative procedures, including the simple Bonferroni correction for multiple hypothesis testing, which we fix for the case of positive common correlation among assets, the chi-bar square test against one-sided alternatives, Follman's test, and Hansen's asymptotic adjustments. [26, 6, 10]

Testing indicates the conditional inference procedure achieves nominal type I rate, and does not appear to suffer from non-normality of returns. The conditional estimation test has low power under the alternative where there is little spread in the signal-noise ratios of the assets, and high power under the alternative where a single asset has high signal-noise ratio. Unlike the alternative procedures, it appears to enjoy rejection probabilities montonic in the signal-noise ratio of the selected asset.

1 Introduction

The problem of overfitting quantitative investment strategies is certainly as old as the problem of selecting quantitative investment strategies. The choice of a course of action (*e.g.*, making an investment) based on historical observations leads to biased estimates of the value of the selected course of action when one uses the same historical observations to estimate value. That is, the estimates are "biased by selection". This problem is not unique to quantitative finance, and goes by many names: overfitting, p-hacking, data-mining bias, *etc.* To be clear we are interested in the case where one has observed n independent contemporaneous observations of returns from k different "assets" (these can be trading strategy backtest returns, or mutual fund returns, *etc.*), selects one of those assets based on the historical performance, say by selecting the asset with maximum Sharpe ratio; then one wishes to estimate or perform inference

*steven@gilgamath.com. The source code to build this document is available at www.github.com/shabbychef/maxsharpe. The document you are reading was built from commit 58c202f600ddc17a4b0e4b9fbc80f5ee7e3d549c of that repo.

on the true 'value' of the asset, for example its signal-noise ratio, which we define as population analogue of the Sharpe ratio.

Aronson gives a good overview of the problem from the practicioner's point of view, noting the relevant factors are the length of history, the number of strategies tested, the correlation of their historical performance, presence of outliers (or fat-tailedness of returns), and variation in expected true effect size. [2, Chapter 6] White's Reality Check was a pioneering development in the area, giving a generally applicable method of estimating whether a selected model was superior to a benchmark model. [31] White's work was extended and generalized by Romano and Wolf, Hansen, *inter alia*. [25, 9, 11] From a practical point of view the Reality Check and its variants do not scale computationally to hundreds or thousands of assets, as they are based on a (block) bootstrap. However, these methods can be adapted to very general problems, can deal with correlation and autocorrelation of asset returns, and are fairly robust to assumptions.

Recent work by López de Prado and Bailey, adapting standard techniques from Multiple Hypothesis Testing (MHT), has gained attention in the field[1]. [18] They find the asymptotic expected value of the maximum Sharpe ratio of uncorrelated assets with zero signal-noise ratio. While use of simple techniques from MHT (Bonferroni correction, say) can lead to reduced power, and is fragile with respect to assumptions, they are alluring in their simplicity. The Bonferroni correction is very simple to describe and implement, and does not require one to store the historical returns of the assets. It easily scales to millions of tested assets.

In this paper we exploit a result from Lee *et al.* on the problem of *conditional estimation.* [15] The Lee procedure was originally devised for analysis of the Lasso, but is applicable in general to the case of selection from a normally distributed vector conditional to a linear constraint. We simply give an multivariate normal approximation to the vector of Sharpe ratios of k assets, then appeal to the Lee *et al.* procedure.

Our procedure is midway between the simple MHT correction and the Reality Check tests, both computationally and in robustness. Our procedure requires one to estimate the correlation between returns, which would appear to require $\mathcal{O}\left(nk^2\right)$ runtimes. However, only the correlation of the selected asset against all others is required, reducing the burden to $\mathcal{O}\left(nk\right)$. Unlike the bootstrap tests, our procedure is easily adapted to the case of producing confidence intervals on the signal-noise ratio, instead of only supporting hypothesis testing.

2 Conditional Inference on the signal-noise ratio

We consider the following problem: one has observed n *i.i.d.* samples of some k-vector $\boldsymbol{x}$, representing the returns of k different "assets," which could be stocks, trading strategies, *etc.* From the sample one computes the Sharpe ratio of each asset, resulting in a k-vector, $\hat{\boldsymbol{\zeta}}$. One will then choose the asset with maximum Sharpe ratio. One then seeks to perform hypothesis tests or compute confidence intervals on the signal-noise ratio of that asset. Here we define the Sharpe ratio as the sample mean divided by sample standard deviation, and the signal-noise

[1] Though application of MHT corrections to the problem is not new: White's starting assumption was apparently that such simple corrections were inadequate. [31]

ratio as the population analogue. Throughout we use hats to denote sample quantities estimating population parameters.

To simplify the exposition, we will suppose that, conditional on observing the vector $\hat{\boldsymbol{\zeta}}$, one rearranges the indices such that the first asset has demonstrated the highest Sharpe ratio. This is to avoid the cumbersome notation of $\hat{\zeta}_{(1)}$, and we instead can just write $\hat{\zeta}_1$. We note this maximum condition can be written in the form $\mathsf{A}\hat{\boldsymbol{\zeta}} \le \boldsymbol{b}$ for $k-1 \times k$ matrix A defined by

$$\mathsf{A} = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & \dots & 1 \end{bmatrix},$$

and where $\boldsymbol{b}$ is the $(k-1)$-dimensional zero vector. Also note that we are interested in performing inference on ζ_1, which we can express as $\boldsymbol{\eta}^\top \boldsymbol{\zeta}$ for $\boldsymbol{\eta} = \boldsymbol{e}_1$.

Under these conditions, if only $\hat{\boldsymbol{\zeta}}$ were normally distributed, one could use the following theorem due to Lee *et al.*:

Theorem 2.1 (Lee *et al.*, Theorem 5.2 [15])**.** *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 $F(x; a, b, 0, 1)$ 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 $F(x; a, b, \mu, \sigma^2)$ be the CDF of a general truncated normal, defined by

$$F(x; a, b, \mu, \sigma^2) = F\left(\frac{x-\mu}{\sigma}; \frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma}, 0, 1\right).$$

Then, conditional on $\mathsf{A}\boldsymbol{y} \le \boldsymbol{b}$, 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

$$\mathcal{V}^- = \max_{j:(\mathsf{A}\boldsymbol{c})_j < 0} \frac{\boldsymbol{b}_j - (\mathsf{A}\boldsymbol{z})_j}{(\mathsf{A}\boldsymbol{c})_j},$$
$$\mathcal{V}^+ = \min_{j:(\mathsf{A}\boldsymbol{c})_j > 0} \frac{\boldsymbol{b}_j - (\mathsf{A}\boldsymbol{z})_j}{(\mathsf{A}\boldsymbol{c})_j}.$$

This theorem gives us a way to perform hypothesis tests, by comparing $F\left(\boldsymbol{\eta}^\top\boldsymbol{y}; \mathcal{V}^-, \mathcal{V}^+, \boldsymbol{\eta}^\top\boldsymbol{\mu}, \boldsymbol{\eta}^\top\Sigma\boldsymbol{\eta}\right)$ to some cutoff. It also suggests a procedure for computing confidence intervals on $\boldsymbol{\eta}^\top\boldsymbol{\mu}$, namely by univariate search for a value of $\boldsymbol{\eta}^\top\boldsymbol{\mu}$ such that $F\left(\boldsymbol{\eta}^\top\boldsymbol{y}; \mathcal{V}^-, \mathcal{V}^+, \boldsymbol{\eta}^\top\boldsymbol{\mu}, \boldsymbol{\eta}^\top\Sigma\boldsymbol{\eta}\right)$ is equal to some cutoff value.

In the following section we will show that the $\hat{\boldsymbol{\zeta}}$ is *approximately* normally distributed. In the following section we will examine whether the normal approximation is good enough to use the procedure of Lee *et al.* for testing the signal-noise ratio of the asset with maximum Sharpe ratio.

2.1 Normal approximation of the distribution of Sharpe ratios

Here we derive the asymptotic distribution of Sharpe ratio, following Jobson and Korkie *inter alia.* [13, 17, 20, 22] Consider the case of k possibly correlated returns streams, with each observation denoted by the k-vector $\boldsymbol{x}$. Let $\boldsymbol{\mu}$ be the k-vector of population means, and let $\boldsymbol{\alpha}_2$ be the k-vector of the uncentered second moments. Let $\boldsymbol{\zeta}$ be the vector of signal-noise ratios 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 $2k$ 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

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

Under the multivariate central limit theorem [30]

$$\sqrt{n}\left(\left[\hat{\boldsymbol{\mu}}^\top, {\hat{\boldsymbol{\alpha}}_2}^\top\right]^\top - \left[\boldsymbol{\mu}^\top, {\boldsymbol{\alpha}_2}^\top\right]^\top\right) \rightsquigarrow \mathcal{N}\left(0, \Omega\right). \tag{1}$$

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,

$$\sqrt{n}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}\right) \rightsquigarrow \mathcal{N}\left(0, \left(\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, {\boldsymbol{\alpha}_2}^\top]^\top}\right)\Omega\left(\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, {\boldsymbol{\alpha}_2}^\top]^\top}\right)^\top\right). \tag{2}$$

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

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

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. This is impractical for large k, so instead one may wish to impose some distributional assumptions on $\boldsymbol{x}$.

Consider the case where $\boldsymbol{x}$ is drawn from a normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Then, using Isserlis' Theorem [12, 8], we have

$$\Omega = \left[\begin{array}{cc}\Sigma & 2\Sigma\,\mathrm{Diag}\left(\boldsymbol{\mu}\right) \\ 2\,\mathrm{Diag}\left(\boldsymbol{\mu}\right)\Sigma & 2\Sigma\odot\Sigma + 4\,\mathrm{Diag}\left(\boldsymbol{\mu}\right)\Sigma\,\mathrm{Diag}\left(\boldsymbol{\mu}\right)\end{array}\right], \tag{4}$$

where $\odot$ denotes *Hadamard multiplication*.

Let R be the correlation matrix of the returns, defined as

$$\mathsf{R} =_{\mathrm{df}} \operatorname{Diag}\left(\boldsymbol{\sigma}^{-1}\right) \Sigma \operatorname{Diag}\left(\boldsymbol{\sigma}^{-1}\right), \tag{5}$$

where $\boldsymbol{\sigma}$ is the (positive) square root of the diagonal of Σ. Then using the Ω given in Equation 4, Equation 1 becomes

$$\hat{\boldsymbol{\zeta}} \approx \mathcal{N}\left(\boldsymbol{\zeta}, \frac{1}{n}\left(\mathsf{R} + \frac{1}{2}\operatorname{Diag}\left(\boldsymbol{\zeta}\right)\left(\mathsf{R}\odot\mathsf{R}\right)\operatorname{Diag}\left(\boldsymbol{\zeta}\right)\right)\right). \tag{6}$$

(See the appendix.) Note how in the case of scalar Gaussian returns, this reduces to the well known standard error estimate of $\sqrt{\frac{1}{n}\left(1+\frac{\zeta^2}{2}\right)}$. [17, 20, 3, 22] In practice the correlation matrix R and the vector of signal-noise ratios, $\boldsymbol{\zeta}$, have to be estimated and plugged in.

We claim that for the case of *elliptically distributed* $\boldsymbol{x}$, Equation 6 can be generalized to

$$\hat{\boldsymbol{\zeta}} \approx \mathcal{N}\left(\boldsymbol{\zeta}, \frac{1}{n}\left(\mathsf{R} + \frac{\kappa-1}{4}\boldsymbol{\zeta}\boldsymbol{\zeta}^\top + \frac{\kappa}{2}\operatorname{Diag}\left(\boldsymbol{\zeta}\right)\left(\mathsf{R}\odot\mathsf{R}\right)\operatorname{Diag}\left(\boldsymbol{\zeta}\right)\right)\right), \tag{7}$$

where κ is the "kurtosis factor", equal to one third the kurtosis of the marginals. [28] However, elliptically distributed returns have no skew, which makes them less than ideal for modeling returns series. Once again note how this equation reduces to the form of the standard error described by Mertens in the case of $k = 1$. [20]

Corollary 2.2 (to Theorem 2.1)**.** *Let* $\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \Sigma\right)$*, with* $\boldsymbol{\zeta} = \boldsymbol{\mu}\oslash\boldsymbol{\sigma}$*, where* $\boldsymbol{\sigma} = \operatorname{diag}\left(\Sigma\right)$*. Let* R *be the correlation matrix. Suppose you observe* n *independent observations of* $\boldsymbol{x}$ *then construct the Sharpe ratio,* $\hat{\boldsymbol{\zeta}}$*. Then, conditional on* $\mathsf{A}\hat{\boldsymbol{\zeta}} \le \boldsymbol{b}$*, the random variable*

$$u = F\left(\boldsymbol{\eta}^\top\hat{\boldsymbol{\zeta}}; \mathcal{V}^-, \mathcal{V}^+, \boldsymbol{\eta}^\top\boldsymbol{\zeta}, \boldsymbol{\eta}^\top\mathsf{Q}\boldsymbol{\eta}\right)$$

is Uniform on $[0, 1]$*, where* $\mathcal{V}^-$*,* $\mathcal{V}^+$ *and* $F\left(x; a, b, \mu, \sigma^2\right)$ *are as in the theorem, and*

$$\mathsf{Q} = \frac{1}{n}\left(\mathsf{R} + \frac{1}{2}\operatorname{Diag}\left(\boldsymbol{\zeta}\right)\left(\mathsf{R}\odot\mathsf{R}\right)\operatorname{Diag}\left(\boldsymbol{\zeta}\right)\right),$$

as given in Equation 6.

Note that the relationship between $\boldsymbol{\eta}^\top\hat{\boldsymbol{\zeta}}$ and $\mathcal{V}^-$ and $\mathcal{V}^+$ is such that u is unlikely to be strictly monotonic increasing with $\boldsymbol{\eta}^\top\hat{\boldsymbol{\zeta}}$, *ceterus paribus*. However, when $\boldsymbol{\eta}^\top\hat{\boldsymbol{\zeta}} \to \infty$, we expect $u \to 1$, and so to test the null hypothesis

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

one should reject at the α level when

$$F\left(\boldsymbol{\eta}^\top\hat{\zeta}; \mathcal{V}^-, \mathcal{V}^+, c, \boldsymbol{\eta}^\top\mathsf{Q}\boldsymbol{\eta}\right) \ge 1 - \alpha.$$

As stated the procedure requires that one estimate Q, which requires one to estimate R and $\boldsymbol{\zeta}$. However, computing the test statistic only requires access to

$\mathsf{Q}\boldsymbol{\eta}$. In the main inferential task considered here, that vector is the covariance of the asset with maximum Sharpe ratio against all the rest.

Note that Corollary 2.2 has uses beyond the stated problem of performing inference on the asset with the largest Sharpe ratio. For example, suppose you observe the Sharpe ratios of k assets, then select the asset with the largest *absolute* Sharpe ratio, choosing whether to hold it long or short depending on the sign of the Sharpe ratio. You wish to perform inference on your strategy. In this case, again reorder the assets such that the first asset has the highest absolute Sharpe ratio, but also flip the signs of the asset returns as necessary such that all assets have positive Sharpe ratio. Then proceed as in the usual case, but add to A and $\boldsymbol{b}$ the conditional restriction that all elements of $\hat{\boldsymbol{\zeta}}$ are non-negative.

One wishes to also use the result for more general problems wherein one will hold a *portfolio* of assets, conditional on some observed properties of $\hat{\boldsymbol{\zeta}}$. For example:

- Suppose you observe the Sharpe ratios of k assets, then select the top m by Sharpe ratio, then you choose to hold an some portfolio of those m assets. In this case set A and $\boldsymbol{b}$ to reflect the "m choose $k-m$" relevant inequalities to condition on.

- Suppose you observe the Sharpe ratios of k assets, then select all assets with Sharpe ratio greater than some minimum value, ζ_*. Then you choose to hold some portfolio of all assets that pass the bar. In this case you need to modify A and $\boldsymbol{b}$ to condition on the passing assets having Sharpe ratios greater than ζ_* and the remaining assets having lower Sharpe ratio.

In these cases, the test vector $\boldsymbol{\eta}$ should reflect the chosen portfolio, but the signal-noise ratio of a portfolio is *not* the portfolio-weighted sum (or average) of the signal-noise ratios of the constituent assets. Indeed the signal-noise ratio of dollar-weighted portfolio $\boldsymbol{\nu}$ is $\boldsymbol{\mu}^\top\boldsymbol{\nu}/\sqrt{\boldsymbol{\nu}^\top\Sigma\boldsymbol{\nu}}$. However, if $\boldsymbol{\nu}$ is expressed in *volatility units*, then the signal-noise ratio is $\boldsymbol{\zeta}^\top\boldsymbol{\nu}/\sqrt{\boldsymbol{\nu}^\top\mathsf{R}\boldsymbol{\nu}}$. Thus assuming you can estimate volatility (and R) without error[2], then one could transform a dollar denominated portfolio into a volatility denominated portfolio. From this one can perform inference using the test vector $\boldsymbol{\eta} = \boldsymbol{\nu}/\sqrt{\boldsymbol{\nu}^\top\mathsf{R}\boldsymbol{\nu}}$.

One could also use the procedure to test the hypothesis that the asset with maximum Sharpe ratio has higher signal-noise ratio than the *average* signal-noise ratio of all assets considered. This is the null commonly tested by the Reality Check and its variants. It may be of limited practical utility, however, since the selected asset may still have inferior signal-noise ratio.

3 Alternative Approaches

3.1 Bonferroni correction with simple correlation fix

The simple MHT approach to the problem is via a Bonferroni correction. [4] In its usual form, it assumes that the returns $\boldsymbol{x}$ are independent and normally

[2]Typically the error in a volatility estimate is less critical than error in the estimate of the mean. [5]

distributed. In this case the marginals of $\hat{\boldsymbol{\zeta}}$ are independent, and distributed as rescaled non-central t random variables. So to test the null hypothesis

$$H_0 : \forall_i \zeta_i = c \quad \text{versus} \quad H_1 : \exists_i \zeta_i > c, \tag{8}$$

compute the Sharpe ratios, $\hat{\boldsymbol{\zeta}}$, then reject at the α level when $\sqrt{n} \max_i \hat{\boldsymbol{\zeta}}_i$ exceeds $t_{1-\alpha/k}\left(\sqrt{n}c, n-1\right)$, the $1-\alpha/k$ quantile of the non-central t-distribution with $n-1$ degrees of freedom and non-centrality parameter $\sqrt{n}c$.

This simple test does does not maintain nominal type I rate in the face of correlated assets. This can be demonstrated empirically, as we do in the sequel. One can get also get a theoretical hint of why this holds by considering the normal approximation of $\hat{\boldsymbol{\zeta}}$ given in Equation 6, then appealing to Slepian's Lemma. Slepian's Lemma establishes that for a normally distributed Gaussian vector with fixed mean and variance, the maximum element is 'stochastically decreasing' as correlations increase. [27] Intuitively the number of true independent assets is decreasing as correlation increases.

Let us consider a simple model for the correlation matrix

$$\mathsf{R} = \rho\left(\mathbf{1}\mathbf{1}^{\top}\right) + (1-\rho)\,\mathsf{I}, \tag{9}$$

where $|\rho| \leq 1$. Now simplify Equation 6 to

$$\hat{\boldsymbol{\zeta}} \approx \mathcal{N}\left(\boldsymbol{\zeta}, \frac{1}{n}\mathsf{R}\right), \tag{10}$$

which is reasonable in the case of the small signal-noise ratios likely to be encountered in practice. Then under the null hypothesis that $\boldsymbol{\zeta} = \boldsymbol{\zeta}_0$, one observes

$$\boldsymbol{z} = \sqrt{n}\mathsf{R}^{-1/2}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right) \approx \mathcal{N}\left(\mathbf{0}, \mathsf{I}\right), \tag{11}$$

where $\mathsf{R}^{-1/2}$ is the inverse of the (symmetric) square root of R.

Under the assumed form for R given in Equation 9, it is simple to confirm that

$$\mathsf{R}^{-1/2} = \frac{1}{k}\left(\frac{1}{\sqrt{1-\rho+k\rho}} - \frac{1}{\sqrt{1-\rho}}\right)\left(\mathbf{1}\mathbf{1}^{\top}\right) + (1-\rho)^{-1/2}\,\mathsf{I}. \tag{12}$$

(This relation holds if we replace $\mathbf{1}\mathbf{1}^{\top}$ in R with $\boldsymbol{w}\boldsymbol{w}^{\top}$ where $\boldsymbol{w}$ is any vector whose elements are ± 1. However in this case we will lose the order-preserving property.)

Now it is simple to confirm that in this case the transform induced by $\mathsf{R}^{-1/2}$ is "order-preserving." That is, if $\boldsymbol{a} = \mathsf{R}^{-1/2}\boldsymbol{b}$ and $b_i \leq b_j$ then $a_i \leq a_j$. As a consequence, if i is the maximal element of $\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right)$, then i is the maximal element of $\boldsymbol{z}$. Let us assume, again, that by convention we have reordered the elements such that the first element of $\hat{\boldsymbol{\zeta}}$ is the maximum. Then to test the null hypothesis $\forall_i \zeta_i = c$, compute

$$z_1 = \frac{\sqrt{n}\hat{\zeta}_1}{\sqrt{1-\rho}} + \left(\frac{1}{\sqrt{1-\rho+k\rho}} - \frac{1}{\sqrt{1-\rho}}\right)\sqrt{n}\left(\frac{\mathbf{1}^{\top}\hat{\boldsymbol{\zeta}}}{k} - c\right), \tag{13}$$

and reject the null hypothesis when z_1 is bigger than $z_{1-\alpha/k}$, the $1-\alpha/k$ quantile of the normal distribution. In practice ρ must be estimated. This could be done by computing the correlations of the first asset against all others, then taking the average.

Note that the test statistic z_1 in Equation 13 depends on elements of $\hat{\boldsymbol{\zeta}}$ other than $\hat{\zeta}_1$. Indeed it depends on the average value among the $\hat{\zeta}_i$. This may not be desireable, as it would reject as one of the $\hat{\zeta}_i$ went to $-\infty$ for $i \neq 1$. Moreover, the statistic z_1 does not seem to be entirely "about" $\hat{\zeta}_1$, but is computed from all elements of $\hat{\boldsymbol{\zeta}}$. To rectify this, one is tempted to rotate the $\boldsymbol{z}$ from Equation 11 to be maximally aligned with $\boldsymbol{e}_1$. This is an area of continued research.

Note that we did not have to make the simplifying assumption that led from Equation 7 to Equation 10 given the form we assumed for R. That is, assuming $\mathsf{R} = a_2 \left(\mathbf{1}\mathbf{1}^{\top}\right) + a_1 \mathsf{I}$, then under the null hypothesis that $\boldsymbol{\zeta} = \zeta_0 \mathbf{1}$, Equation 7 becomes

$$\hat{\boldsymbol{\zeta}} \approx \mathcal{N}\left(\boldsymbol{\zeta}, \frac{1}{n}\left(a_2' \left(\mathbf{1}\mathbf{1}^{\top}\right) + a_1' \mathsf{I}\right)\right), \tag{14}$$

for some constants a_1', a_2' which depend on a_1, a_2, κ and ζ_0. One could then proceed as above, constructing a z_1 statistic.

Bonferroni correction for arbitrary correlation structure The Bonferroni correction outlined above is strictly only applicable to the rank-one correlation matrix, $\mathsf{R} = \rho\left(\mathbf{1}\mathbf{1}^{\top}\right) + (1-\rho)\mathsf{I}$. To apply the correction to any correlation matrix with positive entries, Slepian's lemma allows us to appeal to a worst-case rank-one correlation matrix. [27, 33] Let $\boldsymbol{x} \sim \mathcal{N}\left(\mathbf{0}, \mathsf{R}\right)$ and $\boldsymbol{y} \sim \mathcal{N}\left(\mathbf{0}, \rho\left(\mathbf{1}\mathbf{1}^{\top}\right) + (1-\rho)\mathsf{I}\right)$, for $\rho \geq 0$ where $\mathsf{R}_{i,j} \geq \rho$ for all $i \neq j$. By Slepian's lemma, $\Pr\left\{\max_i x_i \geq t\right\} \leq \Pr\left\{\max_i y_i \geq t\right\}.$

Then assume the correlation of returns is R, where $\mathsf{R}_{i,j} \geq \rho$ for $i \neq j$ for some $\rho \geq 0$; to test the null hypothesis $\forall_i \zeta_i = c$, compute z_1 as in Equation 13, plugging in ρ, and reject the null hypothesis when z_1 is bigger than $z_{1-\alpha/k}$, the $1-\alpha/k$ quantile of the normal distribution. This procedure has (approximate) type I rate *no greater* than α.

3.2 Testing against one-sided alternatives

Another obvious approach to the problem is to appeal to the normal approximation of Equation 6 or Equation 7, then use well known techniques in testing of a multivariate normal against a one-sided alternative. [26, 29] That is, the usual multivariate procedure to test the null hypothesis $H_0 : \forall_i \zeta_i = c$ under a normal approximation would be via Hotelling's T^2 test. [1, 24] However we are not interested in the case where some of the ζ_i are less than c.

One-sided tests will not scale well to the case of large k, except perhaps under simple models for correlation. Consider testing the following null

$$H_0 : \forall_i \zeta_i = \zeta_0 \quad \text{versus} \quad H_1 : \forall_i \zeta_i \geq \zeta_0 \text{ and } \exists_j \zeta_j > \zeta_0,$$

subject to the rank one correlation structure of Equation 9 with $\rho \geq 0$. Again assume the approximation of Equation 10,

$$\hat{\boldsymbol{\zeta}} \approx \mathcal{N}\left(\boldsymbol{\zeta}, \frac{1}{n}\mathsf{R}\right).$$

Letting $\hat{\boldsymbol{\xi}} = \mathsf{R}^{-1/2}\hat{\boldsymbol{\zeta}}$ and $\boldsymbol{\xi} = \mathsf{R}^{-1/2}\boldsymbol{\zeta}$, the normal approximation can be rewritten as

$$\sqrt{n}\hat{\boldsymbol{\xi}} \approx \mathcal{N}\left(\sqrt{n}\boldsymbol{\xi}, \mathsf{I}\right).$$

The inverse square root of R, given in Equation 12, is order-preserving. Moreover,

$$\mathsf{R}^{-1/2}\mathbf{1} = c\mathbf{1},$$

for $c = \left(1 + (k-1)\,\rho\right)^{-1/2} > 0$. From the order-preserving nature of $\mathsf{R}^{-1/2}$, the null and alternative hypotheses 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 then appeal to the simple chi-bar square test. [26, 32] First transform the vector of Sharpe ratios to $\hat{\boldsymbol{\xi}}$ via

$$\hat{\boldsymbol{\xi}} = \mathsf{R}^{-1/2}\hat{\boldsymbol{\zeta}} = c\bar{\zeta}\mathbf{1} + (1-\rho)^{-1/2}\left(\hat{\boldsymbol{\zeta}} - \bar{\zeta}\mathbf{1}\right), \tag{15}$$

where $\bar{\zeta}$ is the average of the sample Sharpe ratios. Then compute

$$\bar{x}^2 = n \sum_i \left(\hat{\xi}_i - c\zeta_0\right)_+^2, \tag{16}$$

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

$$Q = \sum_{i=0}^{k} w_i F_{\chi^2}\left(\bar{x}^2; i\right), \tag{17}$$

where $F_{\chi^2}(x; i)$ is the cumulative distribution of the χ^2 distribution with i degrees of freedom, and w_i are the chi-bar square weights. In this case they are defined as

$$w_i = \binom{k}{i} 2^{-k}.$$

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. [6] Here we would compute

$$g^2 = nkc^2\left(\bar{\zeta} - \zeta_0\right)^2 + \frac{n}{1-\rho}\sum_i \left(\hat{\zeta}_i - \bar{\zeta}\right)^2,$$

and reject at the α level if both $1 - F_{\chi^2}\left(g^2; k\right) \leq 2\alpha$ and $\bar{\zeta} > \zeta_0$.

3.3 Hansen's $\log\log$ Corrections

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, (8) 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, or the problem of (8). [26]

Hansen describes a procedure which avoids this problem. The idea is elegant, and ultimately simple to implement. [10, 9] In the terms of the problem we consider, it amounts to assuming that the null takes the form

$$H_0 : \forall_i\, \zeta_i \le c \text{ and } \left|\zeta_i - \hat{\zeta}_i\right| \le g_n \quad \text{versus} \quad H_1 : \exists_i \zeta_i > c,$$

for some g_n. Note this seems odd since the sample Sharpe ratio appears in the null hypothesis to be tested. 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 (*cf.* Equation 16), and later to and to a Studentized version of White's Reality Check statistic, which is rather like the corrected Bonferroni statistic computed in Equation 13. [10, 9] Applying Hansen's correction to our problem is simple: compute $\hat{\boldsymbol{\xi}}$ as in Equation 15. 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 16 and reject if

$$\sum_{i=0}^{\tilde{k}} \binom{\tilde{k}}{i} 2^{-\tilde{k}} F_{\chi^2}\left(\bar{x}^2; i\right) \ge 1 - \alpha.$$

We will refer to this as "Hansen's chi-bar-square." It is 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

$$\max_i \hat{\boldsymbol{\xi}}_i - c\zeta_0 \ge z_{1-\alpha/\tilde{k}}.$$

We will refer to this as "Hansen's SPA," although it does not use the bootstrap procedure to estimate the standard error as described by Hansen, it is similar in every other regard. [9]

3.4 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 back-tested 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 m-dimensional subspace

$$\mathsf{X} \approx \mathsf{Y}\mathsf{W},$$

where Y is a $n \times m$ matrix of 'latent' returns, and W is a $m \times k$ 'loading' matrix.

Now the column of X with maximal $\hat{\boldsymbol{\zeta}}$ has Sharpe ratio that is smaller than

$$\hat{\zeta}_* =_{\text{df}} \max_{\boldsymbol{\nu}} \frac{\hat{\boldsymbol{\mu}}^\top \boldsymbol{\nu}}{\sqrt{\boldsymbol{\nu}^\top \hat{\Sigma} \boldsymbol{\nu}}},$$

where $\hat{\boldsymbol{\mu}}$ is the m-vector of the (sample) means of columns of Y and $\hat{\Sigma}$ is the sample covariance matrix. This maximum takes value

$$\hat{\zeta}_* = \sqrt{\hat{\boldsymbol{\mu}}^\top \hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}}},$$

which is, up to scaling, Hotelling's T^2 statistic.

Under the null hypothesis that the rows of Y are independent draws from a Gaussian random variable with zero mean, then

$$\frac{(n-m)\,\hat{\zeta}_*^2}{m\,(n-1)}$$

follows an F distribution with m and $n-m$ degrees of freedom. Under the alternative it follows a non-central F distribution. [1, 24] Via this upper bound $\hat{\zeta}_1 \le \hat{\zeta}_*$, one can then perform tests on the null hypothesis $\forall_i \zeta_i = 0$.

However, this approach requires that one estimate m, 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. [23]

4 Empirical Results

4.1 Simulations under the null

4.1.1 Gaussian returns, infeasible estimator

First we seek to establish if, and under what circumstances, the normal approximation of Equation 6 is sufficiently accurate to give nominal coverage under the conditional estimation procedure. First we test a single case of $k = 100$ using a correlation matrix that is $\rho = 0.7$ on the off-diagonals: $\mathsf{R} = 0.7\left(\mathbf{1}\mathbf{1}^\top\right) + 0.3\,\mathsf{I}$. We generate Gaussian returns over 1260 days, approximately 5 years worth for equity returns. We let $\boldsymbol{\zeta}$ range uniformly from $-0.1\text{day}^{-1/2}$ to $0.1\text{day}^{-1/2}$. We compute the Sharpe ratios of each asset's simulated returns, find the asset with maximum Sharpe ratio, then compute a p-value using Theorem 2.1. Since we wish to assess the accuracy of the normal approximation, we use the actual population value of R, and the $\boldsymbol{\zeta}$ to compute the covariance via Equation 6. This is not, of course, how the test would be applied in practice since R and $\boldsymbol{\zeta}$ have to be estimated. Moreover, we with to check coverage of the procedure under the null, so we use the actual $\boldsymbol{\eta}^\top\boldsymbol{\zeta}$ in our test.

We repeat this experiment 10^5 times and collect the resultant putative p-values. We would like to Q-Q or P-P plot these p-values, as evidence that they are near uniform, but the large sample size presents some challenges. Instead, we choose some selected small q (like 0.05 or 0.01), and compute the proportion of our p-values $\le q$. We then subtract q. This value, call it Δ, should be near zero. We plot Δ against q, with errorbars around the x axis reflecting the area where we would expect the points to fall roughly 95% of the time. Those errorbars are computed via the Binomial law, but do not always have width of exactly 95% because of the finite sample size. However, the plot suggests that the p-values are indeed uniformly distributed, and that the procedure has near nominal type I rate when selecting a cutoff near the selected q.

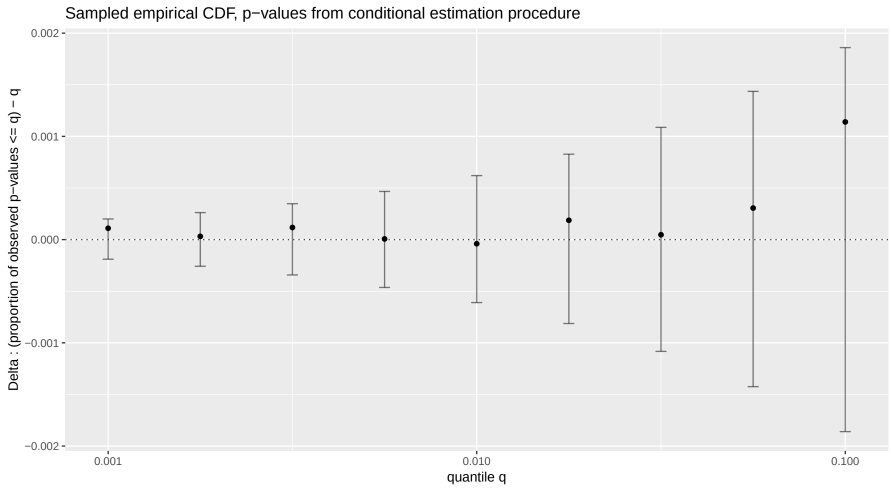

Figure 1: The computed p-values from the conditional estimation procedure over 1e+05 simulations are compared to a uniform law. Given the large sample size, a Q-Q plot is visually hard to interpret. Instead, for selected q, we compute the proportion of our putative p-values less than q. We then compute $\Delta = (\text{Prop } \leq q) - q$ against q. Using the binomial law, we plot approximate error bars around the x axis that indicate where the points should fall with approximately 95% confidence. Simulations use the exact R and $\boldsymbol{\zeta}$ to compute the covariance matrix. The plot supports uniformity of the putative p-values.

4.1.2 Gaussian returns, feasible estimator

While these experiments suggest the normal approximation leads to nearly uniform p-values under the null, they use the (unknown) population values of R and $\boldsymbol{\zeta}$ to compute the variance-covariance of $\hat{\boldsymbol{\zeta}}$. So we repeat the experiments, but plug in the usual sample estimate of covariance and the vector of Sharpe ratios into Equation 6 to estimate the covariance matrix of $\hat{\boldsymbol{\zeta}}$. Other than this change, we repeat the previous experiment, performing 10^5 simulations, setting $k = 100$, $\mathsf{R} = 0.7\left(\mathbf{1}\mathbf{1}^{\top}\right) + 0.3\,\mathsf{I}$. $n = 1260\text{day}$, *etc.* In Figure 2 we present a sampled CDF plot of the log p values, as above. Again the simulations are consistent with the procedure having nominal coverage. This is not surprising, because, as noted above, the statistical test only requires us to estimate the standard error of the Sharpe ratio of the asset with maximum Sharpe ratio, and so does not greatly rely on the k^2 elements of the estimate of R.

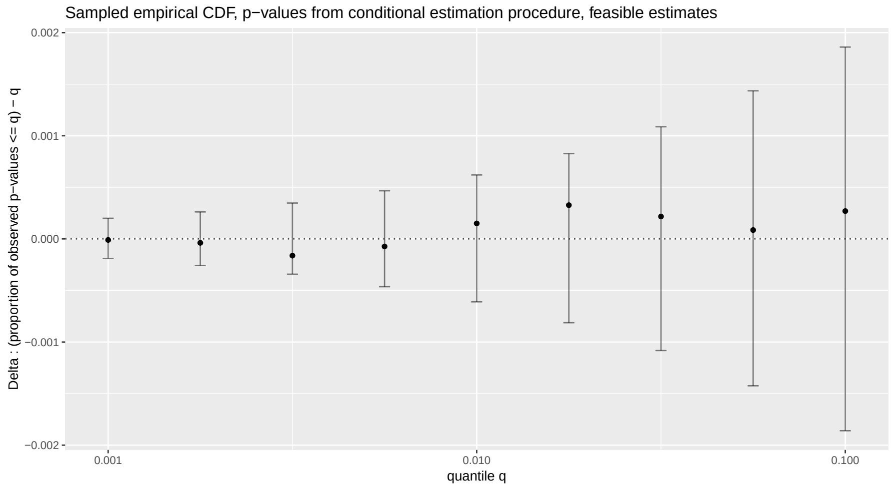

Figure 2: The computed p-values from the conditional estimation procedure using sample estimates of the covariance matrix, R, and the vector of signal-noise ratios, $\boldsymbol{\zeta}$ are used in this sampled CDF plot. P values are computed over 1e+05 simulations. For selected q, we compute the proportion of our putative p-values less than q. We then compute $\Delta = (\text{Prop } \le q) - q$ against q. Using the binomial law, we plot approximate error bars around the x axis that indicate where the points should fall with approximately 95% confidence. Simulations use the exact R and $\boldsymbol{\zeta}$ to compute the covariance matrix. The plot supports uniformity of the putative p-values.

4.1.3 Gaussian returns, feasible estimator, sensitivity

This kind of "proof by eyeball" is somewhat unsatisfying, and does not scale up to the task of finding where the approximation is accurate. To measure the uniformity of our p-values, we generate some via simulations as described above, then compute the Kolmogorov-Smirnov statistic against a uniform distribution. [19] You can think of the K-S statistic as the maximum absolute deviance of a point away from the $y = x$ line in a P-P plot like Figure 1.

So we repeat the previous experiments, using a feasible estimator of the covariance matrix of $\hat{\boldsymbol{\zeta}}$. Again we draw returns from a Gaussian distribution. We let n vary from 126 to 2016; we let k vary from 50 to 200; we let ρ vary from 0 to 0.8 where we take $\mathsf{R} = \rho \left(\mathbf{1}\mathbf{1}^{\top}\right) + (1 - \rho)\,\mathsf{I}$; we take $\boldsymbol{\zeta}$ to be a uniform sequence from 0 to $0.1\text{day}^{-1/2}$. For each setting of the parameters in the Cartesian product we perform 5×10^4 simulations, computing p values from the feasible estimator.

In Figure 3 we plot those K-S statistics against n, with facets for ρ and k. All else equal, we expect the approximation to be worse, and thus the K-S statistics to be higher, for smaller n and larger k. This pattern is somewhat visible in the plots, although large ρ seems to reduce the number of 'pseudo-assets' in that relationship. However, with the given limited evidence, we cannot claim to have definitively established where our procedure breaks down, but warn users that the $k \gg n$ cases are likely to be problematic in the sense that nominal type I rates may not be maintained.

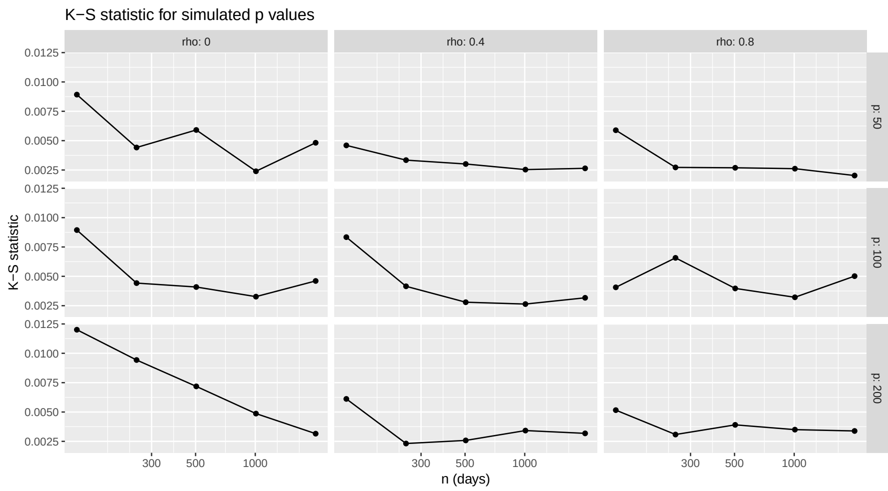

Figure 3: Kolmogorov-Smirnov statistics summarizing uniformity of the test statistic u are plotted versus n with facets for k and ρ. Broadly we see that the test statistic is less uniform in the regime where $k \gg n$, but that a large positive ρ perhaps reduces the number of assets in that relationship.

4.1.4 t returns, feasible estimator

The simulations above were carried out assuming Gaussian returns, and using Equation 6 to compute the covariance matrix of $\hat{\boldsymbol{\zeta}}$. Gaussian returns are not a good model for real asset returns, so we repeat those simulations with returns drawn from a multivariate t-distribution with 5 degrees of freedom. [16, 14] Again we perform 10^4 simulations with $k = 100$, $\mathsf{R} = 0.7\left(\mathbf{1}\mathbf{1}^\top\right) + 0.3\,\mathsf{I}$, $n = 1260\text{day}$, *etc.* We perform inference twice, once using Equation 6, and once using Equation 7 where we have estimated the kurtosis factor, κ, by taking the median of the sample marginal kurtosises of the assets. In Figure 4 we present the subsampled empirical CDF plots on the log-transformed p-values under the two methods of estimating the covariance of $\hat{\boldsymbol{\zeta}}$. There is little difference in the performance of the two sets of simulations, though without the correction, the procedure is slightly anti-conservative, while with the correction it is slightly conservative. We remain cautiously optimistic that for large n, one need not correct Equation 6 to account for non-normal returns.

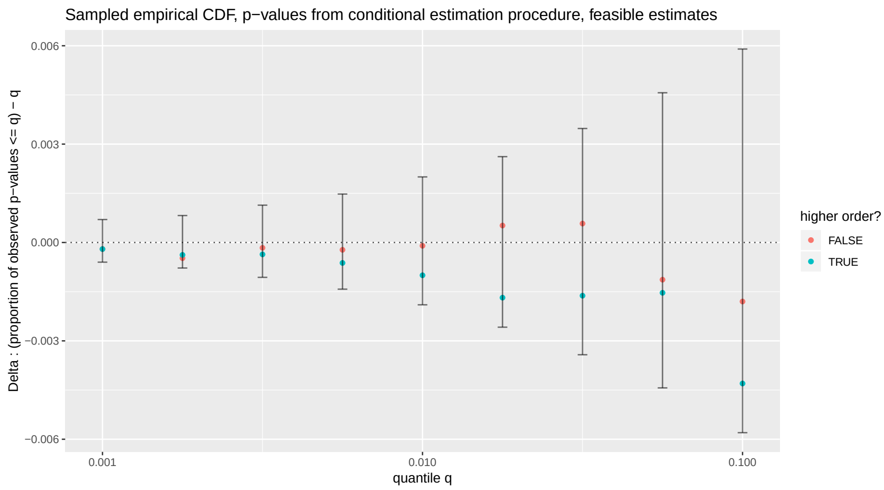

Figure 4: The computed p-values from the conditional estimation procedure over 10000 simulations are plotted using our sampled CDF procedure. Returns are drawn from a $t\,(5)$ distribution. Sample estimates are used to construct the variance-covariance matrix of $\hat{\boldsymbol{\zeta}}$. The experiments are performed twice: first assuming that returns are Gaussian; then assuming returns are elliptical with unknown kurtosis factor that we estimate from the sample. Both procedures give near-uniform p values. Without correcting for higher order moments, the procedure has higher than nominal type I rate; with the correction it has slightly smaller than the nominal rate.

4.2 Simulations under the alternative

We wish to test the power of the method under the alternative hypothesis. However, it is hard to state exactly what constitutes the alternative. 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 will consider following alternatives: 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 $m = 1$, we call this the "one-good" alternative; for $m = k/2$, the "half-good" alternative.

We compare the power of the conditional estimation procedure to that of a simple MHT correction, and the one-sided test. In our experiments, we draw returns from a Gaussian distribution with diagonal covariance. Under this assumption, one can use the distribution of the t statistic to perform inference on the signal-noise ratio. [22] We then use the Bonferroni correction to account for the multiple tests performed. We also perform the one-sided test based on the chi-bar square statistic, Follman's test, and the chi-bar and MHT tests with Hansen's $\log\log$ adjustment.

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 Bonferroni-corrected 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 chi-bar square test and Follman's test are also unlikely to reject. The conditional estimation procedure, however, should not suffer under this alternative. We also expect the $\log\log$ adjustment to have little effect when the Sharpe ratios are all nearly equal, and more effect when they are different.

Our suspicions are born out by the simulations. We perform simulations under all-equal, one-good, and half-good configurations, letting the 'good' signal-noise ratio vary from 0 to $0.15\text{day}^{-1/2}$, which corresponds to an 'annualized' signal-noise ratio of around $2.4\text{yr}^{-1/2}$. We draw Gaussian returns with diagonal covariance for 100 assets, with $n = 1008$. For each setting we perform 10^4 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. We hope our tests to have lower power when this occurs.

In Figure 5, 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'st 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 zero 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.

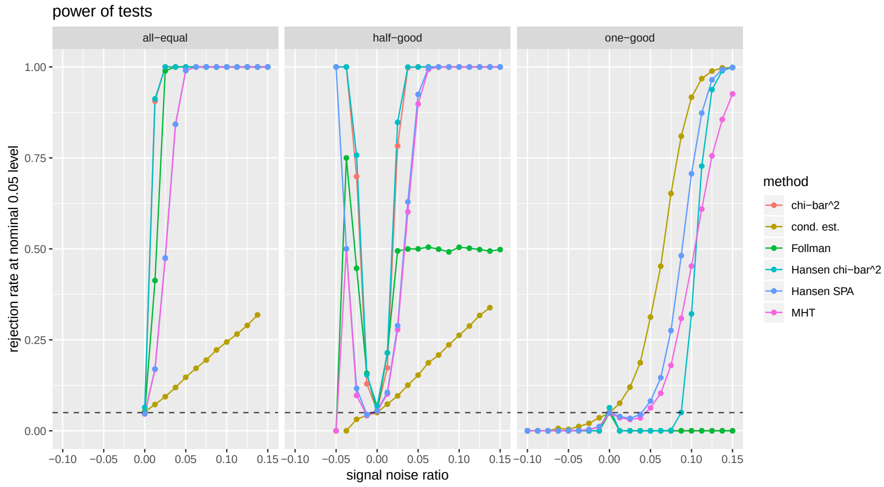

Figure 5: The empirical power of the conditional estimation, MHT corrected test, chi-bar-square test, Follman's test, Hansen's SPA and corrected chi-bar-square procedures are shown versus the signal-noise ratio of the asset with maximum Sharpe ratio under different arrangements of the vector $\boldsymbol{\zeta}$. For the one-good case, the chi-bar-square and Follman's test have the same power, which is essentially zero under the alternative.

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 $2.4\text{yr}^{-1/2}$, 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 under this alternative, however. We have performed 10^4 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 $0.05\text{day}^{-1/2}$ we do not observe this occuring even once over the 10^4 simulations performed for each setting.

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 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.

4.2.1 Simulations under the null versus with correlated returns

The naïve MHT test cannot maintain the nominal type I rate in the face of correlated assets. To demonstrate this, we repeat the experiments above, but setting $\mathsf{R} = \rho\left(\mathbf{1}\mathbf{1}^{\top}\right) + (1-\rho)\,\mathsf{I}$ and $\zeta = \mathbf{0}$. Again we consider $k = 20$, $n = 504$, and perform 10^4 simulations to estimate the empirical rejection rate. In Figure 6, we plot the empirical rejection rate versus ρ at the nominal 0.05 type I level. While the conditional estimation procedure and one-sided tests appear to maintain the nominal rejection rate, the MHT test is conservative, with near zero rejection rates for large ρ. The fix for common correlation described in Section 3.1 is also tested, yielding nominal rejection rates.

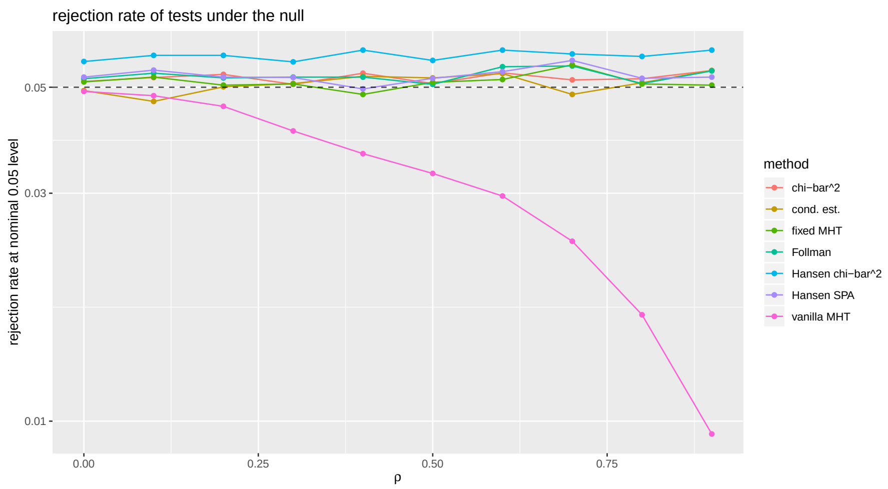

Figure 6: The empirical type I rate under the null hypothesis is plotted versus ρ for the case where $\mathsf{R} = \rho\left(\mathbf{1}\mathbf{1}^{\top}\right) + (1-\rho)\,\mathsf{I}$, for several testing procedures: the conditional estimation, the chi-bar square test, Follman's test, the vanilla Bonferroni correction, Hansen's corrected chi-bar-square and SPA tests, and Bonferroni correction with fix for common correlation. Tests are performed with Gaussian returns, for 20 assets over 504 days. Tests were performed at the 0.05 level, which appears to be maintained by all procedures except the regular Bonferroni procedure, and Hansen's chi-bar-square test, which has slightly higher than nominal rejection rates for this n. Empirical rates are over 10,000 simulations. The y axis is in log scale.

4.3 Five Industry Portfolios

We download the monthly returns of five industry portfolios from Ken French's data library. [7] We consider the 1104 months of data on five industries from Jan 1927 to Dec 2018. We compute the Sharpe ratio of the returns of each, and present them in Table 1. We have reordered the industries in decreasing Sharpe ratio. The industry portfolio with the highest Sharpe ratio was Healthcare with a Sharpe ratio of around $0.193\,\mathrm{mo.}^{-1/2}$ which is approximately $0.667\,\mathrm{yr}^{-1/2}$.

industry	Sharpe Ratio
Healthcare	$0.193\,\mathrm{mo.}^{-1/2}$
Consumer	$0.187\,\mathrm{mo.}^{-1/2}$
Manufacturing	$0.172\,\mathrm{mo.}^{-1/2}$
Technology	$0.170\,\mathrm{mo.}^{-1/2}$
Other	$0.140\,\mathrm{mo.}^{-1/2}$

Table 1: The Sharpe ratios of the five industry portfolios are shown.

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 $\left[0.143\,\mathrm{mo.}^{-1/2}, \infty\right)$ based on inverting the non-central t-distribution. [22, 21] If instead we approximate the standard error by plugging in $0.193\,\mathrm{mo.}^{-1/2}$ as the signal-noise ratio of Healthcare into Equation 6, we estimate the standard error of the Sharpe ratio to be $0.03\,\mathrm{mo.}^{-1/2}$. Based on this we can compute the naïve confidence interval of the measured Sharpe ratio plus $z_{0.05} = -1.645$ times the standard error. This also gives the confidence interval $\left[0.143\,\mathrm{mo.}^{-1/2}, \infty\right)$.

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 $z_{0.01} = -2.326$ 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 0.708 to 0.891 with a median value of 0.801. Plugging this value in as ρ, we find the value c such that the z_1 from Equation 13 is equal to $z_{0.01} = -2.326$. This leads to the confidence interval $\left[0.125\,\mathrm{mo.}^{-1/2}, \infty\right)$.

We use this estimate of ρ to compute the chi-bar square test. We invert the test to find the 95% upper confidence interval $\left[0.141\,\mathrm{mo.}^{-1/2}, \infty\right)$.

Finally we use the conditional estimation procedure, inverting the hypothesis test to find the corresponding population value. This yields the confidence interval $\left[0.073\,\mathrm{mo.}^{-1/2}, \infty\right)$.

5 Conclusions and Future Work

The conditional estimation procedure appears to achieve nominal type I rates under the null, and does not seem unduly harmed by assuming the vector $\hat{\boldsymbol{\zeta}}$ is normally distributed. Nor does it seem to suffer greatly from using sample estimates of the correlation matrix, R, nor from the presence of kurtotic returns. The procedure can be used for other test configurations beyond the asset with the maximum Sharpe ratio, and can be used to construct confidence intervals. It appears to have low power when many assets have the same signal-noise ratio, but appears to have low rejection probability when the selected asset has low signal-noise ratio, which sets it apart from all of the other procedures tested.

Clearly the low power of the test gives us reason to seek improvements. Perhaps the conditioning procedure can be adapted to recognize that the strategist would have been testing another asset if the Sharpe ratio of the currently selected asset had been lower. However, it would seem that in so doing, one would lose the desireable property of testing the signal-noise ratio of the selected asset, instead of testing the more general hypothesis of (8). Perhaps its power can be increased using Hansen's $\log\log$ adjustment.

References

[1] 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.

[2] 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/.

[3] 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.

[4] 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.

[5] Vijay Kumar Chopra and William T. Ziemba. The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, 19(2):6–11, 1993. URL http://faculty.fuqua.duke.edu/~charvey/Teaching/BA453_2006/Chopra_The_effect_of_1993.pdf.

[6] 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.

[7] 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.

[8] J. B. S. Haldane. Moments of the distributions of powers and products of normal variates. *Biometrika*, 32(3/4):pp. 226–242, 04 1942. ISSN 0006-3444. doi: 10.1093/biomet/32.3-4.226. URL https://doi.org/10.1093/biomet/32.3-4.226.

[9] 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.

[10] Peter Reinhard Hansen. Asymptotic tests of composite hypotheses. Working Paper 2003-09, Providence, RI, 2003. URL http://hdl.handle.net/10419/80222.

[11] Po-Hsuan Hsu, Yu-Chin Hsu, and Chung-Ming Kuan. Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias. *Journal of Empirical Finance*, 17(3):471 – 484, 2010. ISSN 0927-5398. doi: 10.1016/j.jempfin.2010.01.001. URL http://ssrn.com/abstract=1087044.

[12] 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.

[13] 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.

[14] 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.

[15] 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).

[16] Pi-Erh Lin. Some characterizations of the multivariate t distribution. *Journal of Multivariate Analysis*, 2(3):339 – 344, 1972. ISSN 0047-259X. doi: http://dx.doi.org/10.1016/0047-259X(72)90021-8. URL http://www.sciencedirect.com/science/article/pii/0047259X72900218.

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

[18] Marcos López de Prado and David H. Bailey. The false strategy theorem: A financial application of experimental mathematics. *American Mathematical Monthly, forthcoming*, 2018.

[19] George Marsaglia, Wai Wan Tsang, and Jingbo Wang. Evaluating Kolmogorov's distribution. *Journal of Statistical Software*, 8(18):1–4, 11 2003. ISSN 1548-7660. URL http://www.jstatsoft.org/v08/i18.

[20] 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.

[21] Steven E. Pav. *SharpeR: Statistical significance of Sharpe ratio*, 2014-2019. URL https://github.com/shabbychef/SharpeR. R package version 1.2.0.

[22] Steven E. Pav. A short Sharpe course. Privately Published, 2017. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3036276.

[23] Steven E. Pav. Distribution of maximal Sharpe, the Markowitz approximation. Privately Published, 2018. URL http://www.sharperat.io/max-sharpe-two.html.

[24] 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.

[25] Joseph P Romano and Michael Wolf. Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282, 2005. URL https://www.econ.uzh.ch/dam/jcr:ffffffff-935a-b0d6-ffff-ffffa286d4d1/etca.pdf.

[26] 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.

[27] 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.

[28] C. Vignat and S. Bhatnagar. An extension of Wick's theorem, 2007. URL http://arxiv.org/abs/0709.1999.

[29] 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.

[30] 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.

[31] Halbert White. A reality check for data snooping. *Econometrica*, 68:1097–1127, 2000. URL https://www.ssc.wisc.edu/~bhansen/718/White2000.pdf.

[32] 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.

[33] Ofer Zeitouni. GAUSSIAN FIELDS notes for lectures. unpublished notes, 2017. URL http://www.wisdom.weizmann.ac.il/~zeitouni/notesGauss.pdf.

A Establishing Equation 7.

Let V be the variance covariance to be computed. Then from Equation 2,

$$\begin{aligned}
\mathsf{V} &= \left(\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, \boldsymbol{\alpha}_2{}^\top]^\top}\right) \Omega \left(\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}[\boldsymbol{\mu}^\top, \boldsymbol{\alpha}_2{}^\top]^\top}\right)^\top, \\
&= \operatorname{Diag}\left(\frac{1}{\boldsymbol{\sigma}^2}\right) \left[\begin{array}{cc} \operatorname{Diag}(\boldsymbol{\sigma} + \boldsymbol{\mu}\boldsymbol{\zeta}) & \operatorname{Diag}\left(\frac{-\boldsymbol{\zeta}}{2}\right) \end{array}\right] \Omega \left[\begin{array}{c} \operatorname{Diag}(\boldsymbol{\sigma} + \boldsymbol{\mu}\boldsymbol{\zeta}) \\ \operatorname{Diag}\left(\frac{-\boldsymbol{\zeta}}{2}\right) \end{array}\right] \operatorname{Diag}\left(\frac{1}{\boldsymbol{\sigma}^2}\right).
\end{aligned}$$

Plugging in Ω from Equation 4, we have

$$\begin{aligned}
\mathsf{V} = \operatorname{Diag}\left(\frac{1}{\boldsymbol{\sigma}^2}\right) \Bigg\{ & \operatorname{Diag}(\boldsymbol{\sigma} + \boldsymbol{\mu}\boldsymbol{\zeta})\, \Sigma \operatorname{Diag}(\boldsymbol{\sigma} + \boldsymbol{\mu}\boldsymbol{\zeta}) + \operatorname{Diag}(\boldsymbol{\sigma} + \boldsymbol{\mu}\boldsymbol{\zeta})\, 2\Sigma \operatorname{Diag}(\boldsymbol{\mu}) \operatorname{Diag}\left(\frac{-\boldsymbol{\zeta}}{2}\right) \\
& + \operatorname{Diag}\left(\frac{-\boldsymbol{\zeta}}{2}\right) 2 \operatorname{Diag}(\boldsymbol{\mu})\, \Sigma \operatorname{Diag}(\boldsymbol{\sigma} + \boldsymbol{\mu}\boldsymbol{\zeta}) + \operatorname{Diag}\left(\frac{-\boldsymbol{\zeta}}{2}\right) 2\Sigma \odot \Sigma \operatorname{Diag}\left(\frac{-\boldsymbol{\zeta}}{2}\right) \\
& + \operatorname{Diag}\left(\frac{-\boldsymbol{\zeta}}{2}\right) 4 \operatorname{Diag}(\boldsymbol{\mu})\, \Sigma \operatorname{Diag}(\boldsymbol{\mu}) \operatorname{Diag}\left(\frac{-\boldsymbol{\zeta}}{2}\right) \Bigg\} \operatorname{Diag}\left(\frac{1}{\boldsymbol{\sigma}^2}\right), \\
= \operatorname{Diag}\left(\frac{1}{\boldsymbol{\sigma}^2}\right) \{ & \operatorname{Diag}(\boldsymbol{\sigma})\, \Sigma \operatorname{Diag}(\boldsymbol{\sigma}) - \operatorname{Diag}(\boldsymbol{\mu}\boldsymbol{\zeta})\, \Sigma \operatorname{Diag}(\boldsymbol{\mu}\boldsymbol{\zeta}) \\
& + \operatorname{Diag}\left(\frac{-\boldsymbol{\zeta}}{2}\right) 2\Sigma \odot \Sigma \operatorname{Diag}\left(\frac{-\boldsymbol{\zeta}}{2}\right) \\
& + \operatorname{Diag}(\boldsymbol{\mu}\boldsymbol{\zeta})\, \Sigma \operatorname{Diag}(\boldsymbol{\mu}\boldsymbol{\zeta}) \} \operatorname{Diag}\left(\frac{1}{\boldsymbol{\sigma}^2}\right), \\
= \operatorname{Diag}\left(\frac{1}{\boldsymbol{\sigma}^2}\right) & \left\{ \operatorname{Diag}(\boldsymbol{\sigma})\, \Sigma \operatorname{Diag}(\boldsymbol{\sigma}) + \frac{1}{2} \operatorname{Diag}(\boldsymbol{\zeta})\, \Sigma \odot \Sigma \operatorname{Diag}(\boldsymbol{\zeta}) \right\} \operatorname{Diag}\left(\frac{1}{\boldsymbol{\sigma}^2}\right), \\
= \mathsf{R} + \frac{1}{2} & \operatorname{Diag}(\boldsymbol{\zeta}) \operatorname{Diag}\left(\frac{1}{\boldsymbol{\sigma}^2}\right) \Sigma \odot \Sigma \operatorname{Diag}\left(\frac{1}{\boldsymbol{\sigma}^2}\right) \operatorname{Diag}(\boldsymbol{\zeta}), \\
= \mathsf{R} + \frac{1}{2} & \operatorname{Diag}(\boldsymbol{\zeta})\, \mathsf{R} \odot \mathsf{R} \operatorname{Diag}\left(\frac{1}{\boldsymbol{\sigma}^2}\right).
\end{aligned}$$

Proving Equation 7 is similar, and is left as an exercise for the reader.