Bounds on Portfolio Quality

Steven E. Pav *

November 15, 2019

Abstract

The signal-noise ratio of a portfolio of p assets, its expected return divided by its risk, is couched as an estimation problem on the sphere $\mathcal{S}^{p-1}$. When the portfolio is built using noisy data, the expected value of the signal-noise ratio is bounded from above via a Cramér-Rao bound, for the case of Gaussian returns. The bound holds for 'biased' estimators, thus there appears to be no bias-variance tradeoff for the problem of maximizing the signal-noise ratio. An approximate distribution of the signal-noise ratio for the Markowitz portfolio is given, and shown to be fairly accurate via Monte Carlo simulations, for Gaussian returns as well as more exotic returns distributions. These findings imply that if the maximal population signal-noise ratio grows slower than the universe size to the $\frac{1}{4}$ power, there may be no diversification benefit, rather expected signal-noise ratio can *decrease* with additional assets. As a practical matter, this may explain why the Markowitz portfolio is typically applied to small asset universes. Finally, the theorem is expanded to cover more general models of returns and trading schemes, including the conditional expectation case where mean returns are linear in some observable features, subspace constraints (*i.e.*, dimensionality reduction), and hedging constraints.

1 Introduction

Given p assets with expected return $\boldsymbol{\mu}$ and covariance of return Σ, the portfolio defined as

$$\boldsymbol{\nu}_* =_{\text{df}} \Sigma^{-1}\boldsymbol{\mu}, \tag{1}$$

known, somewhat informally, as the 'Markowitz portfolio', plays a central role in portfolio theory. [21, 3] Up to scaling, it solves the classic mean-variance optimization, as well as the (population) Sharpe ratio maximization problem:

$$\max_{\boldsymbol{\nu}} \frac{\boldsymbol{\nu}^\top \boldsymbol{\mu}}{\sqrt{\boldsymbol{\nu}^\top \Sigma \boldsymbol{\nu}}}. \tag{2}$$

In practice, the Markowitz portfolio has a tarnished reputation, and is infrequently, if ever, used without some modification. The unknown population parameters $\boldsymbol{\mu}$ and Σ must be estimated from samples, resulting in a feasible

*steven@gilgamath.com The author thanks Ramakrishna Kakarala for sharing his research. The code to build this document is available at www.github.com/shabbychef/qbound. This revision was built from commit df649027b33cf04570600863093b0ab58dab2cfd of that repo.

portfolio of dubious value. Michaud went so far as to call mean-variance optimization, "error maximization." [24] In its stead, numerous portfolio construction methodologies have been proposed to replace the Markowitz portfolio, some based on patching conjectured theoretical deficiencies, others relying on simple heuristics. [7, 31, 3]

Praticioners often resort to dimensionality reduction heuristics to mitigate estimation error, effectively reducing the number of free variables in the portfolio optimization problem. One version of this tactic describes the returns of dozens, or even hundreds, of equities as the linear combination of a handful of 'factor' returns (plus some 'idiosyncratic' term); the portfolio problem is then couched as an optimization over factor portfolios. If the population parameters were known with certainty, shrinking the set of feasible portfolios would only result in reducing the optimal portfolio utility. However, the population parameters can typically only be weakly estimated, and dimensionality reduction is common practice.

In this paper, an upper bound is established on the expected value of a feasible portfolio's signal-noise ratio, defined to be the expected return of the portfolio divided by it's risk, with return and risk measured using the (unknown) population parameters, and with the "expected value" taken over realizations of the sample used to estimate the portfolio. This bound balances the 'effect size,' $\sqrt{n\boldsymbol{\mu}^{\top}\Sigma^{-1}\boldsymbol{\mu}}$, with the number of assets, p, and justifies some form of dimensionality reduction. It is established, for example, that if, by adding additional assets to the investment universe, $\sqrt{\boldsymbol{\mu}^{\top}\Sigma^{-1}\boldsymbol{\mu}}$ grows at a rate slower than $p^{1/4}$, the upper bound on expected signal-noise ratio can decrease.

2 Portfolio signal-noise ratio

Let $\boldsymbol{x}$ be the vector of relative returns of p assets, with expectation $\boldsymbol{\mu}$ and covariance Σ. A portfolio $\hat{\boldsymbol{\nu}}$ on these assets has expected return $\hat{\boldsymbol{\nu}}^{\top}\boldsymbol{\mu}$ and variance $\hat{\boldsymbol{\nu}}^{\top}\Sigma\hat{\boldsymbol{\nu}}$. Define the signal-noise ratio of the portfolio $\hat{\boldsymbol{\nu}}$ as the signal-noise ratio of the returns of $\hat{\boldsymbol{\nu}}^{\top}\boldsymbol{x}$:

$$q\left(\hat{\boldsymbol{\nu}}\right) =_{\mathrm{df}} \frac{\hat{\boldsymbol{\nu}}^{\top}\boldsymbol{\mu}}{\sqrt{\hat{\boldsymbol{\nu}}^{\top}\Sigma\hat{\boldsymbol{\nu}}}} \tag{3}$$

One can think of the signal-noise ratio as a kind of 'quality' metric on portfolios, as follows: The Sharpe ratio statistic of the future returns of $\hat{\boldsymbol{\nu}}$ are 'stochastically monotonic' in the signal-noise ratio as so defined, meaning that if $q\left(\hat{\boldsymbol{\nu}}_1\right) \le q\left(\hat{\boldsymbol{\nu}}_2\right)$ then the Sharpe ratio of ${\hat{\boldsymbol{\nu}}_2}^{\top}\boldsymbol{x}$ (first order) stochastically dominates the Sharpe ratio of ${\hat{\boldsymbol{\nu}}_1}^{\top}\boldsymbol{x}$.

Note that the portfolio signal-noise ratio is bounded by the signal-noise ratio achieved by the population Markowitz portfolio, $\boldsymbol{\nu}_*$:

$$\left|q\left(\hat{\boldsymbol{\nu}}\right)\right| \le \zeta_* =_{\mathrm{df}} \sqrt{\boldsymbol{\mu}^{\top}\Sigma^{-1}\boldsymbol{\mu}} = q\left(\boldsymbol{\nu}_*\right) = q\left(\Sigma^{-1}\boldsymbol{\mu}\right). \tag{4}$$

We can interpret portfolio signal-noise ratio geometrically, in 'risk space', by introducing a risk transform:

$$q\left(\hat{\boldsymbol{\nu}}\right) = \frac{\hat{\boldsymbol{\nu}}^{\top}\Sigma\Sigma^{-1}\boldsymbol{\mu}}{\sqrt{\hat{\boldsymbol{\nu}}^{\top}\Sigma\hat{\boldsymbol{\nu}}}} = \frac{\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}\right)^{\top}\Sigma^{\top/2}\boldsymbol{\nu}_*}{\sqrt{\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}\right)^{\top}\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}\right)}}. \tag{5}$$

Now normalize by the maximum absolute value that $q(\hat{\boldsymbol{\nu}})$ can take:

$$\begin{aligned}
\frac{q(\hat{\boldsymbol{\nu}})}{\zeta_*} &= \frac{\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}\right)^{\top}\Sigma^{\top/2}\boldsymbol{\nu}_*}{\sqrt{\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}\right)^{\top}\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}\right)}\sqrt{\left(\Sigma^{\top/2}\boldsymbol{\nu}_*\right)^{\top}\left(\Sigma^{\top/2}\boldsymbol{\nu}_*\right)}},\\
&= \left(\frac{\Sigma^{\top/2}\hat{\boldsymbol{\nu}}}{\sqrt{\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}\right)^{\top}\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}\right)}}\right)^{\top}\left(\frac{\Sigma^{\top/2}\boldsymbol{\nu}_*}{\sqrt{\left(\Sigma^{\top/2}\boldsymbol{\nu}_*\right)^{\top}\left(\Sigma^{\top/2}\boldsymbol{\nu}_*\right)}}\right),\\
&= f_{\mathcal{S}}\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}\right)^{\top} f_{\mathcal{S}}\left(\Sigma^{\top/2}\boldsymbol{\nu}_*\right),
\end{aligned}$$

where

$$f_{\mathcal{S}}(\boldsymbol{x}) =_{\mathrm{df}} \frac{\boldsymbol{x}}{\sqrt{\boldsymbol{x}^{\top}\boldsymbol{x}}} \tag{6}$$

is the projection operator taking non-zero vector $\boldsymbol{x}$ to the unit sphere. That is, $q(\hat{\boldsymbol{\nu}})/\zeta_*$ can be viewed as the dot product of two vectors on the unit sphere (assuming both $\hat{\boldsymbol{\nu}}$ and $\boldsymbol{\nu}_*$ are non-zero vectors), namely $f_{\mathcal{S}}\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}\right)$ and $f_{\mathcal{S}}\left(\Sigma^{\top/2}\boldsymbol{\nu}_*\right)$. Let θ be the angle between $f_{\mathcal{S}}\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}\right)$ and $f_{\mathcal{S}}\left(\Sigma^{\top/2}\boldsymbol{\nu}_*\right)$, and thus $q(\hat{\boldsymbol{\nu}}) = \zeta_* \cos\theta$.

In practice the portfolio $\hat{\boldsymbol{\nu}}$ is built using n *i.i.d.* observations of the random variable $\boldsymbol{x}$. Denote these observations by the $n \times p$ matrix X, and, by abuse of notation, denote the *estimator* that gives $\hat{\boldsymbol{\nu}}$ for a given X by $\hat{\boldsymbol{\nu}}(\mathsf{X})$. By the same abuse of notation, write $\theta(\mathsf{X})$. We will bound the expected value of $\hat{\boldsymbol{\nu}}(\mathsf{X})$.

To appeal to a Cramér-Rao bound, one must typically assume the estimator is unbiased. For this problem a somewhat weaker condition suffices.

Assumption 2.1 (Directional Independence)**.** Assume that

$$\mathrm{E}\left[f_{\mathcal{S}}\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}(\mathsf{X})\right)\right] = c_n\left(\zeta_*^2\right) f_{\mathcal{S}}\left(\Sigma^{\top/2}\boldsymbol{\nu}_*\right) + \boldsymbol{b}_n(\boldsymbol{\mu}, \Sigma), \tag{7}$$

where $\boldsymbol{b}_n(\boldsymbol{\mu}, \Sigma)$ is the 'bias' term, which is orthogonal to $f_{\mathcal{S}}\left(\Sigma^{\top/2}\boldsymbol{\nu}_*\right)$, and which may be an arbitrary function of $\boldsymbol{\mu}$ and Σ.

Note that by orthogonality of $\boldsymbol{b}_n(\boldsymbol{\mu}, \Sigma)$ and $f_{\mathcal{S}}\left(\Sigma^{\top/2}\boldsymbol{\nu}_*\right)$, and linearity of the expectation,

$$\mathrm{E}[\cos\theta(\mathsf{X})] = \mathrm{E}\left[\frac{q(\hat{\boldsymbol{\nu}})}{\zeta_*}\right] = \mathrm{E}\left[f_{\mathcal{S}}\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}(\mathsf{X})\right)^{\top} f_{\mathcal{S}}\left(\Sigma^{\top/2}\boldsymbol{\nu}_*\right)\right] = c_n\left(\zeta_*^2\right). \tag{8}$$

Thus $|c_n(x)| \leq 1$, and we expect $c_n(x) \geq 0$ for a 'sane' portfolio estimator. Moreover, one expects $c_n(x) \to 0$ as $nx \to 0$, and for non-zero x, $c_n(x) \to 1$ as $n \to \infty$.

When $\boldsymbol{b}_n(\boldsymbol{\mu}, \Sigma)$ is the zero vector, the estimator is a 'parallel estimator' in Watson's terminology [16], or 'unbiased' in the sense of Hendricks. [13, 12] Note that Equation 7 is satisfied for any *directionally equivariant* portfolio estimator, *i.e.*, one which, for any orthonormal H, ($\mathsf{H}^{\top}\mathsf{H} = \mathsf{I}_p = \mathsf{H}\mathsf{H}^{\top}$), one has

$$\hat{\boldsymbol{\nu}}\left(\mathsf{X}\mathsf{H}^{\top}\right) = \mathsf{H}\hat{\boldsymbol{\nu}}(\mathsf{X}).$$

However, one should recognize that not all portfolio estimators satisfy this assumption. For example, consider an estimator that never concentrates greater than $p^{-\frac{1}{2}}$ proportion of its total gross allocation in any one asset; this estimator does not exhibit Directional Independence, since it can not capitalize when $\boldsymbol{\nu}_* = \zeta_* \boldsymbol{e}_1$. Neither does the "one over N allocation" estimator. [7]

We must eliminate other 'pathological' cases from consideration.

Assumption 2.2 (Residual Independence)**.** Assume that the distribution of the residual

$$f_{\mathcal{S}}\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right) - \mathrm{E}\left[f_{\mathcal{S}}\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)\right]$$

is independent of $\Sigma^{\top/2}\,\boldsymbol{\nu}_*$.

This assumption prevents us from making false assertions about *e.g.*, the $1/N$ allocation in the case where it happens to nearly equal $\boldsymbol{\nu}_*$. [7]

Let $\boldsymbol{y}$ be a p-variate random variable. Then

$$\begin{aligned} \mathrm{tr}\left(\mathrm{Var}\left(\boldsymbol{y}\right)\right) &= \mathrm{tr}\left(\mathrm{E}\left[\left(\boldsymbol{y} - \mathrm{E}\left[\boldsymbol{y}\right]\right)\left(\boldsymbol{y} - \mathrm{E}\left[\boldsymbol{y}\right]\right)^{\top}\right]\right), \\ &= \mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{y}\boldsymbol{y}^{\top}\right]\right) - \mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{y}\right]\mathrm{E}\left[\boldsymbol{y}\right]^{\top}\right), \\ &= \mathrm{E}\left[\boldsymbol{y}^{\top}\boldsymbol{y}\right] - \mathrm{E}\left[\boldsymbol{y}\right]^{\top}\mathrm{E}\left[\boldsymbol{y}\right]. \end{aligned} \tag{9}$$

By Equation 7, and using orthogonality of $\boldsymbol{b}_n\left(\boldsymbol{\mu}, \Sigma\right)$ and $f_{\mathcal{S}}\left(\Sigma^{\top/2}\boldsymbol{\nu}_*\right)$, we then have

$$\begin{aligned} \mathrm{tr}\left(\mathrm{Var}\left(f_{\mathcal{S}}\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)\right)\right) &= 1 - \left(c_n^2\left(\zeta_*^2\right) + \boldsymbol{b}_n^{\top}\left(\boldsymbol{\mu}, \Sigma\right)\boldsymbol{b}_n\left(\boldsymbol{\mu}, \Sigma\right)\right) \\ &\leq 1 - c_n^2\left(\zeta_*^2\right), \end{aligned} \tag{10}$$

We will bound the variance of $f_{\mathcal{S}}\left(\Sigma^{\top/2}\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)$ by a Cramér-Rao lower bound, thus establishing an upper bound on $c_n\left(\zeta_*^2\right)$.

Define

$$\boldsymbol{\eta} =_{\mathrm{df}} \Sigma^{\top/2}\boldsymbol{\nu}_* = \Sigma^{-1/2}\boldsymbol{\mu}. \tag{11}$$

Note that $\boldsymbol{\eta}^{\top}\boldsymbol{\eta} = \boldsymbol{\mu}^{\top}\Sigma^{-1}\boldsymbol{\mu} = \zeta_*^2$. Using the Cramér-Rao lower bound for the left hand side of Equation 10, and then using the definition of $\boldsymbol{\eta}$ in the expectation, we have [25]

$$\frac{1}{n}\,\mathrm{tr}\left(\mathsf{D}\mathcal{I}_{\boldsymbol{\eta}}^{-1}\mathsf{D}^{\top}\right) \leq 1 - c_n^2\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right), \tag{12}$$

where

$$\mathsf{D} =_{\mathrm{df}} \frac{\mathrm{d}c_n\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right)\frac{\boldsymbol{\eta}}{\sqrt{\boldsymbol{\eta}^{\top}\boldsymbol{\eta}}}}{\mathrm{d}\boldsymbol{\eta}}. \tag{13}$$

Here we take the derivative to follow the 'numerator layout' convention, meaning a gradient is a row vector. This derivative takes the form

$$\mathsf{D} = \frac{c_n'\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right)}{\sqrt{\boldsymbol{\eta}^{\top}\boldsymbol{\eta}}}\boldsymbol{\eta}\boldsymbol{\eta}^{\top} + c_n\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right)\left(\frac{\mathsf{I}}{\sqrt{\boldsymbol{\eta}^{\top}\boldsymbol{\eta}}} - \frac{\boldsymbol{\eta}\boldsymbol{\eta}^{\top}}{\boldsymbol{\eta}^{\top}\boldsymbol{\eta}^{\frac{3}{2}}}\right). \tag{14}$$

To compute the Fisher information, $\mathcal{I}_{\boldsymbol{\eta}}$, we must fix the likelihood of the returns, $\boldsymbol{x}$. While the normal distribution is a poor fit for asset returns [6], it is a convenient distribution to work with.

Assumption 2.3 (Normal Returns)**.** Assume that $\boldsymbol{x}$ are multivariate normally distributed, $\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \Sigma\right)$.

For multivariate normal returns, and conditional on $\mathsf{\Sigma}$, the log likelihood takes the form

$$\begin{aligned}\log \mathcal{L}\left(\boldsymbol{\eta}\,|\boldsymbol{x}\,\right) &= c_1 - \frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^{\top}\mathsf{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right),\\ &= c\left(\boldsymbol{x}\right) + \boldsymbol{\eta}^{\top}\mathsf{\Sigma}^{-1/2}\boldsymbol{x} - \frac{1}{2}\boldsymbol{\eta}^{\top}\boldsymbol{\eta},\end{aligned} \tag{15}$$

dropping the 'nuisance parameters' from the likelihood function. The Fisher Information is negative the expectation of the Hessian of the log likelihood with respect to $\boldsymbol{\eta}$. In this case we have simply

$$\mathcal{I}_{\boldsymbol{\eta}} = -\operatorname{E}\left[\frac{\partial^2 \log \mathcal{L}\left(\boldsymbol{\eta}\,|\boldsymbol{x}\,\right)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^{\top}}\right] = \mathsf{I}_p. \tag{16}$$

This radically simplifies the exposition, as the Cramér-Rao bound of Equation 12 can now be expressed as

$$\frac{1}{n}\operatorname{tr}\left(\mathsf{D}\mathsf{D}^{\top}\right) \le 1 - c_n^2\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right). \tag{17}$$

Using the form of D given in Equation 14, and noting that the cross terms are orthogonal, we have

$$\begin{aligned}\operatorname{tr}\left(\mathsf{D}\mathsf{D}^{\top}\right) &= \operatorname{tr}\left(\left[c_n'\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right)\right]^2\boldsymbol{\eta}\boldsymbol{\eta}^{\top} + c_n^2\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right)\left[\frac{\mathsf{I}}{\boldsymbol{\eta}^{\top}\boldsymbol{\eta}} - \frac{\boldsymbol{\eta}\boldsymbol{\eta}^{\top}}{\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right)^2}\right]\right),\\ &= \left[c_n'\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right)\right]^2\boldsymbol{\eta}^{\top}\boldsymbol{\eta} + c_n^2\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right)\frac{p-1}{\boldsymbol{\eta}^{\top}\boldsymbol{\eta}},\end{aligned} \tag{18}$$

using the fact that $\operatorname{tr}\left(\boldsymbol{y}\boldsymbol{y}^{\top}\right) = \boldsymbol{y}^{\top}\boldsymbol{y}$. With Equation 17, this gives

$$\left[c_n'\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right)\right]^2\boldsymbol{\eta}^{\top}\boldsymbol{\eta} + c_n^2\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right)\frac{p-1}{\boldsymbol{\eta}^{\top}\boldsymbol{\eta}} \le n\left(1 - c_n^2\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right)\right). \tag{19}$$

The term $\left[c_n'\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right)\right]^2\boldsymbol{\eta}^{\top}\boldsymbol{\eta}$ is non-negative, so we may discard it to get a coarser bound that does not involve the derivative of c_n:

$$c_n^2\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right)\frac{p-1}{\boldsymbol{\eta}^{\top}\boldsymbol{\eta}} \le n\left(1 - c_n^2\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right)\right). \tag{20}$$

This yields

$$c_n^2\left(\boldsymbol{\eta}^{\top}\boldsymbol{\eta}\right) \le \frac{n\boldsymbol{\eta}^{\top}\boldsymbol{\eta}}{p-1+n\boldsymbol{\eta}^{\top}\boldsymbol{\eta}}, \tag{21}$$

proving the following theorem.

Theorem 2.4. *Let $\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)$ be an estimator based on n* i.i.d. *observations of multivariate Gaussian returns, X, satisfying the assumptions of directional independence and residual independence. Then*

$$\operatorname{E}\left[q\left(\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)\right] \le \frac{\sqrt{n}\zeta_*^2}{\sqrt{p-1+n\zeta_*^2}}. \tag{22}$$

Theorem 2.4 balances the "degrees of freedom" of the estimator, $p-1$, with one lost because only direction matters, and the "observable effect size", $n\zeta_*^2$. The effect size is a unitless quantity. If ζ_* is measured in trading days, then n should be the number of trading days; if ζ_* is measured in 'annualized' terms, then n should be the number of years.

This bound is fairly harsh. Consider a typical actively managed portfolio. Generously, we can estimate $\zeta_* = 1\text{yr}^{-1/2}$ over $p = 10$ assets, using $n = 5\text{yr}$ of historical data. Then the expected value of $q\left(\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)$ is bounded by $0.6\text{yr}^{-1/2}$; the event of having a year-over-year loss is then a "0.6-sigma" event.

Theorem 2.4 suggests that for comparing investments, the magnitude of the *squared* Sharpe ratio is a limiting factor, rather than the Sharpe ratio itself (assuming it is positive). That is, under the bound of the theorem, $\zeta_* = 2\text{yr}^{-1/2}$ is *four* times as 'good' as $\zeta_* = 1\text{yr}^{-1/2}$, in the sense that such an effect size can 'balance' four times as many degrees of freedom.

3 Approximate distribution of the signal-noise ratio of the Markowitz portfolio

Here we establish an approximate distribution of the quantity $q\left(\hat{\boldsymbol{\nu}}\right)/\zeta_* = \cos\theta$ for the sample Markowitz portfolio, $\hat{\boldsymbol{\nu}}_* =_{\text{df}} \hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}$, with $\hat{\Sigma}, \hat{\boldsymbol{\mu}}$ the usual sample estimates of Σ and $\boldsymbol{\mu}$. The approximation is constructed by assuming that misestimation of Σ contributes no error to the portfolio.

Assuming that $\hat{\Sigma} = \Sigma$, then

$$\Sigma^{\top/2}\hat{\boldsymbol{\nu}}_* = \Sigma^{-1/2}\hat{\boldsymbol{\mu}} = \Sigma^{-1/2}\boldsymbol{\mu} + \frac{1}{\sqrt{n}}\boldsymbol{z}, \tag{23}$$

where $\boldsymbol{z} \sim \mathcal{N}\left(\mathbf{0}, \mathsf{I}\right)$. Then, with $q\left(\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)/\zeta_* = \cos\left(\theta\left(\mathsf{X}\right)\right)$, we should have

$$\cot\left(\theta\left(\mathsf{X}\right)\right) = \frac{\left\|\Sigma^{-1/2}\boldsymbol{\mu}\right\|_2 + \frac{1}{\sqrt{n}}z_1}{\sqrt{\frac{1}{n}\sum_{2\le i\le p} z_i^2}}, \tag{24}$$

where the z_i are independent standard normal random variables. This can be expressed as

$$\tan\left(\arcsin\left(\frac{q\left(\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)}{\zeta_*}\right)\right) \sim \frac{1}{\sqrt{p-1}}t\left(\sqrt{n}\zeta_*, p-1\right), \tag{25}$$

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

Approximation 25 implies the following approximation:

$$q^2\left(\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right) \sim \zeta_*^2\mathcal{B}\left(n\zeta_*^2, \frac{1}{2}, \frac{p-1}{2}\right), \tag{26}$$

where $\mathcal{B}\left(\delta, p, q\right)$ is a non-central Beta distribution with non-centrality δ, and 'shape' parameters p and q. [32] However, by describing the distribution of the *square* of $q\left(\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)$, we cannot easily model the (sometimes significant) probability that it is negative. This form, does, however, give bounds on the variance of $q\left(\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)$ under the approximation of Approximation 25, since the moments of the non-central Beta are known. [32, sec 30.3] Under Approximation 25, we have

$$\mathrm{E}\left[q^2\left(\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)\right] = \zeta_*^2 e^{-\frac{n\zeta_*^2}{2}} \frac{\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} \frac{\Gamma\left(\frac{p}{2}\right)}{\Gamma\left(\frac{p+2}{2}\right)} {}_2F_2\left(\frac{p}{2}, \frac{3}{2}; \frac{1}{2}, \frac{2+p}{2}; \frac{n\zeta_*^2}{2}\right), \tag{27}$$

where ${}_2F_2\left(\cdot,\cdot;\cdot,\cdot;\cdot\right)$ is the Generalized Hypergeometric function. [27, sec 16.2] This is a rough upper bound on the variance of $q^2\left(\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)$; a lower bound can be had using the upper bound on the mean from Theorem 2.4.

Because the median value of the non-central t-distribution is approximately equal to the non-centrality parameter, [15, 17] the median value of $q\left(\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)$ for the sample Markowitz portfolio, via Approximation 25, is approximately

$$m \approx \zeta_* \sin\left(\arctan\left(\frac{\sqrt{n}\zeta_*}{\sqrt{p-1}}\right)\right) = \frac{\sqrt{n}\zeta_*^2}{\sqrt{p-1+n\zeta_*^2}}, \tag{28}$$

which is exactly the upper bound of Theorem 2.4!

3.1 Monte Carlo simulations

The accuracy of Approximation 25 is checked by Monte Carlo simulations: 10^6 simulations were performed of construction of the Markowitz portfolio using $n = 1012$ (4 years of daily observations), $p = 6$ and $\zeta_* = 1.25\mathrm{yr}^{-1/2}$; the returns are normally distributed. Since the population Markowitz portfolio is known, the portfolio signal-noise ratio can be computed exactly. The Q-Q plot in Figure 1 confirms that Approximation 25 is very good for this choice of n, p, ζ_*.

Rather than rely on 'proof by graph', the Kolmogorov-Smirnov test was computed for the values of signal-noise ratio generated under Gaussian returns. [22] The statistic, the maximal difference between empirical CDF and theoretical CDF under the approximation, was computed to be 0.005 over the 10^6 simulations. While this seems small, the computed p-value under the null underflows to 0 because the sample size is so large.

The experiment is then repeated using returns drawn from a uniform distribution, a t-distribution with 4 degrees of freedom, from a Tukey h-distribution with parameter $h = 0.15$, and from a Lambert W $\times$ Gaussian with parameter $\gamma = -0.2$. [9, 10] Returns are generated by first generating *i.i.d.* p-variate draws from a zero mean, identity covariance distribution whose marginals follow the so-named laws, then scaling and shifting to have the appropriate ζ_*. For each simulation, the $\boldsymbol{\Sigma}$ is a random draw from a Wishart random variable.

The uniform distribution is not a realistic model of market returns, but is included to check the approximation on platykurtic returns. The t and more exotic distributions are more realistic models of market returns, and are leptokurtotic. The Lambert W has non-zero skew. Again, 10^6 simulations are performed under each of these distributions with the same values of n, p, ζ_* as above. Some of the empirical quantiles from these simulations are shown in Table 1, along with the approximate quantiles from Approximation 25. The Kolmogorov-Smirnov test statistics for the different distributions are presented in Table 2. For this choice of n, p, ζ_*, the approximation is very good, across the tested returns distributions.

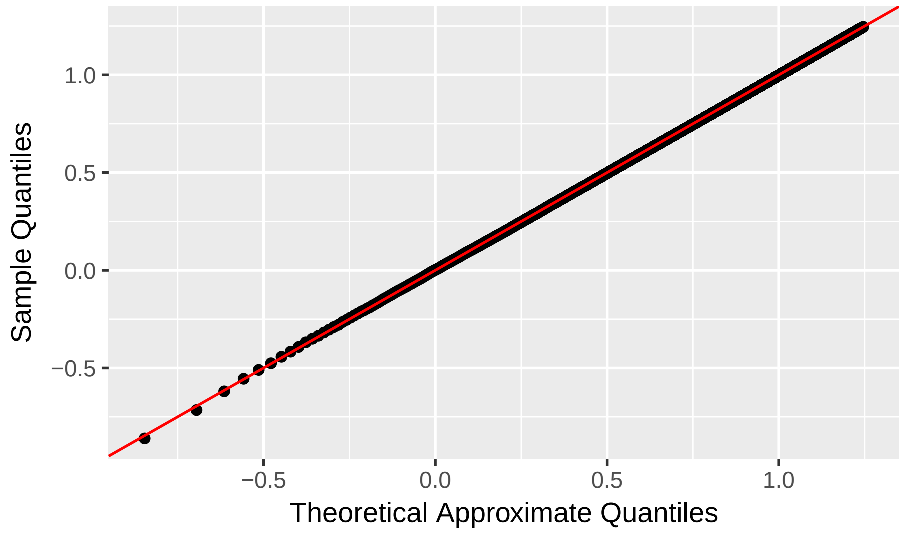

Figure 1: Q-Q plot of 10^6 simulated signal-noise ratio values versus Approximation 25 is shown. Units are 'annual', *i.e.*, $\mathrm{yr}^{-1/2}$. Since the number of samples is very large, only a subset of 10^4 points, uniformly selected by sample quantile, are plotted.

In Table 3, the empirical mean value of $q\left(\hat{\boldsymbol{\nu}}_*\right)$, over the 10^6 simulations, is presented for the five returns distributions, along with the upper bound given by Theorem 2.4. It seems that there is a small gap between the empirical mean for the case of Gaussian returns, and the theoretical upper bound, a gap on the order of 4%. Perhaps this gap is caused by discarding the derivative term from Equation 19, or because the sample Markowitz portfolio is not efficient for finite samples.

In Table 4, the empirical mean value of $q^2\left(\hat{\boldsymbol{\nu}}_*\right)$, over the 10^6 simulations, is presented for the five returns distributions, along with the theoretical value from Equation 27, which is valid only under Approximation 25. The approximate value is decent, meaning an estimate of the variance of $q\left(\hat{\boldsymbol{\nu}}_*\right)$ could be had by combining Equation 27 and the upper bound of Theorem 2.4.

Of course, these simulations are conducted using only a single choice of the parameters n, p and ζ_*. To check the robustness of this approximation to these parameters, 10^5 Monte Carlo simulations were conducted for each combination of $n = 0.5, 1, 2, 4, 8$ years of daily observations, $p = 2, 4, 8, 16$, and $\zeta_* = 0.35, 0.5, 0.71, 1, 1.41\mathrm{yr}^{-1/2}$, all under Gaussian returns. The Kolmogorov-Smirnov test statistic is then computed on the empirically observed quantiles of portfolio signal-noise ratio, under the distribution of Approximation 25.

Plots of the Kolmogorov-Smirnov statistic are given in Figure 2, and Figure 3, which suggest that the quality of Approximation 25 is a function of the quantity $\zeta_*\left(p-1\right)/\sqrt{n}$. As a rough guide, when $\zeta_*\left(p-1\right)/\sqrt{n} \le 5\mathrm{yr}^{-1}$, for daily observations, Approximation 25 is an acceptable approximation to the distribution of signal-noise ratio of the sample Markowitz portfolio.

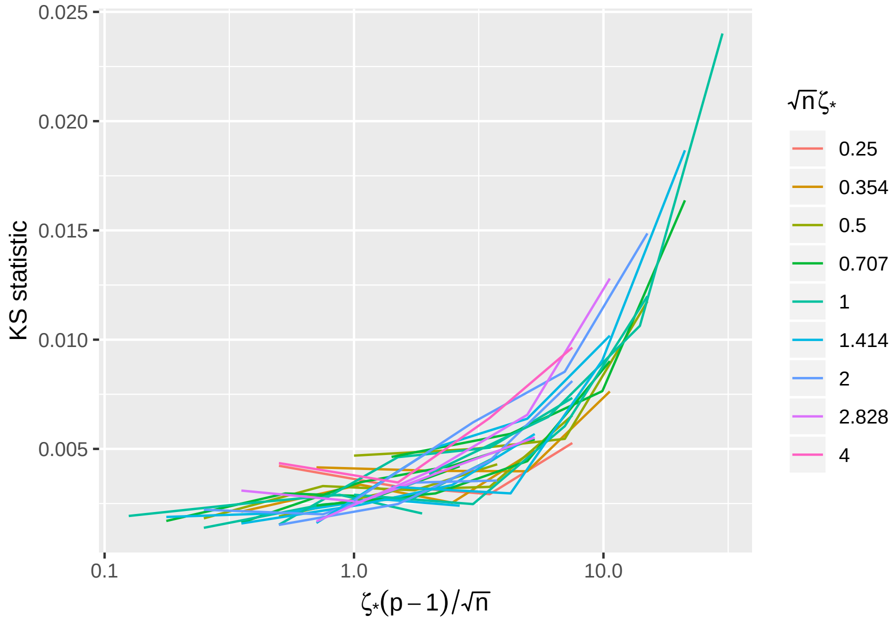

Figure 2: The Kolmogorov-Smirnov statistic for Approximation 25 over 10^5 simulations of Gaussian returns is plotted versus $\zeta_* \left(p-1\right)/\sqrt{n}$, with ζ_* in annualized terms, and n measured in years. There is one line for each combination of n and ζ_*. The line color corresponds to the 'effect size', $\sqrt{n}\zeta_*$, which is unitless.

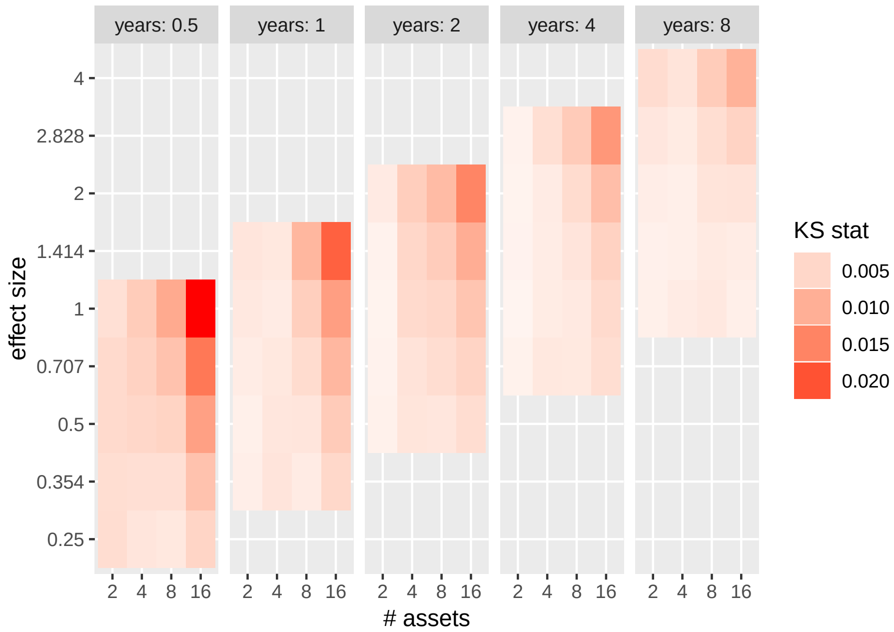

Figure 3: The Kolmogorov-Smirnov statistic for Approximation 25 over 10^5 simulations of Gaussian returns is indicated, by color, versus p, and the 'total effect size,' $\sqrt{n}\zeta_*$, which is a unitless quantity. Different facets are for different values of n (in years).

q.tile	normal	unif.	t(4)	Tukey(0.15)	Lam.W(-0.2)	approx.
0.005	-0.0443	-0.0509	-0.0443	-0.0412	-0.0453	-0.0450
0.010	0.0995	0.0911	0.0973	0.0980	0.0965	0.0996
0.025	0.2891	0.2867	0.2907	0.2915	0.2880	0.2928
0.050	0.4367	0.4345	0.4375	0.4377	0.4355	0.4397
0.250	0.7860	0.7854	0.7872	0.7862	0.7854	0.7890
0.500	0.9527	0.9525	0.9532	0.9527	0.9522	0.9550
0.750	1.0705	1.0706	1.0707	1.0705	1.0701	1.0721
0.900	1.1430	1.1431	1.1435	1.1429	1.1428	1.1442

Table 1: Empirical quantiles of portfolio signal-noise ratio from 10^6 simulations of 1012 days of 6 assets, with maximal Sharpe ratio of $1.25\mathrm{yr}^{-1/2}$ are given, along with the approximate quantiles from Approximation 25. Units of signal-noise ratio are 'annual', *i.e.*, $\mathrm{yr}^{-1/2}$.

normal	unif.	t(4)	Tukey(0.15)	Lam.W(-0.2)
0.0046	0.0055	0.0037	0.0047	0.0058

Table 2: Kolmogorov-Smirnov statistic comparing the empirical CDF to that of Approximation 25 over 10^6 simulations of 1012 days of 6 assets, with maximal Sharpe ratio of $1.25\mathrm{yr}^{-1/2}$ are given for the different returns distributions.

4 Diversification

Theorem 2.4 has implications for the diversification benefit. Consider the case of $p = 6, n = 1012, \zeta_* = 1.25\mathrm{yr}^{-1/2}$ versus some superset of this asset universe with $p = 24, n = 1012, \zeta_* = 1.6\mathrm{yr}^{-1/2}$. Since the optimum cannot decrease over a larger feasible space, we observe that the superset has a higher population signal-noise ratio, ζ_*, One should not, of course, increase the investment universe without *some* concomitant increase in ζ_*. However, in this case the bound on expected signal-noise ratio from Theorem 2.4 for the smaller asset universe is $0.93\mathrm{yr}^{-1/2}$, while for the superset it is $0.89\mathrm{yr}^{-1/2}$. Diversification has possibly caused a *decrease* in expected signal-noise ratio, even though the opportunity exists to increase signal-noise ratio by a fair amount.

By the 'Fundamental Law of Asset Management,' one vaguely expects ζ_* to increase as $\sqrt{p}$. [11] If however, ζ_* scales at a rate slower than $p^{1/4}$, then the derivative of the bound in Theorem 2.4 will be negative for sufficiently large p: adding assets to the universe causes a decrease in expected signal-noise ratio. To see why, note that $\sqrt{n}\zeta_*^2/\sqrt{p-1+n\zeta_*^2}$ has ζ_*^2 in the numerator, and $\sqrt{p}$ in the denominator; if ζ_* grows slower than $p^{1/4}$ the denominator will outpace the numerator.

More formally, let B be the bound on signal-noise ratio from Theorem 2.4:

$$B =_{\mathrm{df}} \frac{\sqrt{n}\zeta_*^2}{\sqrt{p-1+n\zeta_*^2}}.$$

normal	unif.	t(4)	Tukey(0.15)	Lam.W(-0.2)	bound
0.9	0.898	0.899	0.899	0.898	0.932

Table 3: Empirical mean portfolio signal-noise ratio from 10^6 simulations of 1012 days of 6 assets, with maximal Sharpe ratio of $1.25\mathrm{yr}^{-1/2}$ are given, along with the upper bound from Theorem 2.4. Units of signal-noise ratio are 'annual', *i.e.*, $\mathrm{yr}^{-1/2}$.

normal	unif.	t(4)	Tukey(0.15)	Lam.W(-0.2)	approx.
0.9	0.864	0.865	0.864	0.863	0.868

Table 4: Empirical mean of *squared* portfolio signal-noise ratio from 10^6 simulations of 1012 days of 6 assets, with maximal Sharpe ratio of $1.25\mathrm{yr}^{-1/2}$ are given, along with the approximate value from Equation 27. Units of squared signal-noise ratio are 'annual', *i.e.*, yr^{-1}.

By taking the derivative of $\log B$ with respect to p, a little calculus reveals that

$$\begin{aligned} \frac{\mathrm{dlog}\, B}{\mathrm{d}p} \geq 0 \Leftrightarrow \frac{\zeta_*}{2n\zeta_*^2 + 4\left(p-1\right)} &\leq \frac{\mathrm{d}\zeta_*}{\mathrm{d}p}, \\ \Leftrightarrow \frac{1}{2n\zeta_*^2 + 4\left(p-1\right)} &\leq \frac{\mathrm{dlog}\, \zeta_*}{\mathrm{d}p}, \end{aligned} \tag{29}$$

The last inequality is implied by the inequality $\frac{1}{4(p-1)} \leq \frac{\mathrm{dlog}\, \zeta_*}{\mathrm{d}p}$, with equality holding for $\zeta_* = c\left(p-1\right)^{1/4}$.

The decreasing upper bound with respect to growing universe size is illustrated in Figure 4. Under the assumption $\zeta_* = \zeta_0 p^{\gamma}$, the upper bound of Theorem 2.4 is plotted versus p for different values of γ. The value of ζ_0 is set so that $\zeta_* = 1.25\mathrm{yr}^{-1/2}$ when $p = 6$. For $\gamma < \frac{1}{4}$, one sees a local maximum in the upper bound as p increases, a behavior not seen for $\gamma > \frac{1}{4}$, where the bound on signal-noise ratio grows with p.

This relationship between signal-noise ratio and p for different values of γ appears not just in the upper bound of Theorem 2.4, but apparently also for most quantiles of the distribution given by Approximation 25, as illustrated in Figure 5. Again assuming $\zeta_* = \zeta_0 p^{\gamma}$, lines of ζ_* and the $0.25, 0.50$, and 0.75 quantiles of $\zeta_* q\left(\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)$, under Approximation 25, are plotted versus p. The panels represent γ values of $0.21, 0.25$, and 0.29. Again, the value of ζ_0 is set so that $\zeta_* = 1.25\mathrm{yr}^{-1/2}$ when $p = 6$. For $\gamma < \frac{1}{4}$, one sees a local maximum in signal-noise ratio as p increases, a behavior not seen for $\gamma > \frac{1}{4}$, where quantiles of signal-noise ratio grow with p. For the case of 'slow growth' of ζ_*, the diversification benefit is not seen by the sample Markowitz portfolio, rather its practical utility *decreases* because the estimation error outpaces the growth of ζ_*.

4.1 Diversification under CAPM

It is not clear how ζ_* 'should' scale with p. It is easy to construct a model under which ζ_* scales as $p^{\frac{1}{2}}$: assume all assets have independent returns with the same signal-noise ratio. It is also easy to accidentally construct a model under which

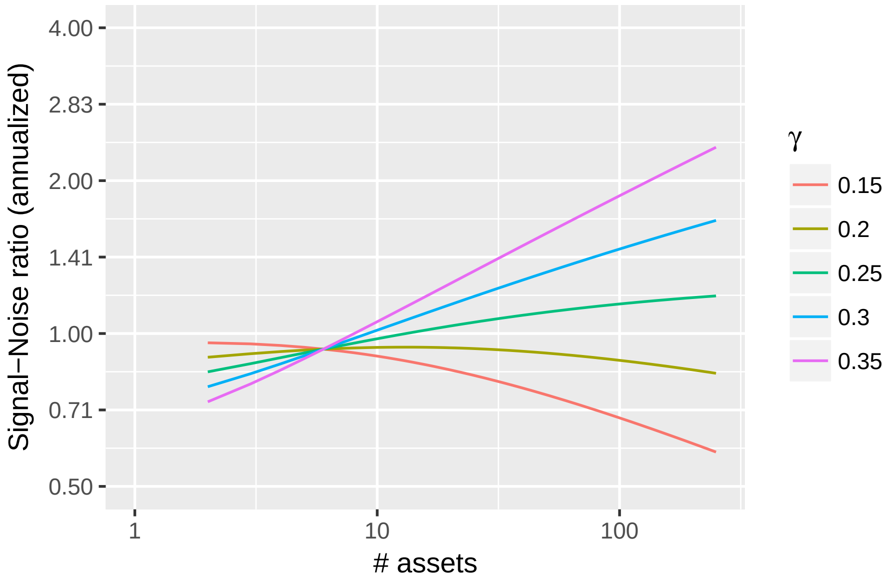

Figure 4: The upper bound of Theorem 2.4 is plotted versus p for different scaling laws for ζ_*. These scaling laws correspond to $\zeta_* = \zeta_0 p^{\gamma}$, with γ taking values between 0.15 and 0.35. The constant terms, ζ_0, are adjusted so that $\zeta_* = 1.25\text{yr}^{-1/2}$ for $p = 6$ for all the lines. The bound uses $n = 1012$, corresponding to 4 years of daily observations.

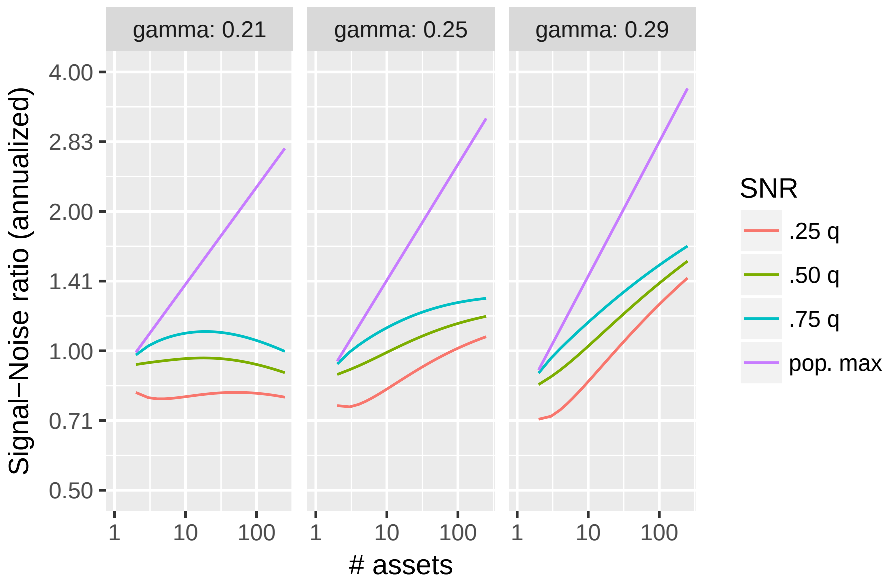

Figure 5: Some quantiles of the signal-noise ratio of the Markowitz portfolio, under Approximation 25, are plotted versus p for different scaling laws for ζ_*. The 3 panels represent different values of γ, *viz.* 0.21, 0.25, and 0.29. The bound uses $n = 1012$, corresponding to 4 years of daily observations.

ζ_* ultimately scales as p^{ϵ} for small ϵ, as done here. Suppose the i^{th} asset has expected return α_i, exposure β_i to 'the market', and volatility σ. Assume the market return is zero mean with volatility σ_m. Then the squared signal-noise ratio is

$$\begin{aligned}\zeta_*^2 &= \frac{\boldsymbol{\alpha}^\top\boldsymbol{\alpha} + \left(\frac{\sigma_m}{\sigma}\right)^2\left[\boldsymbol{\beta}^\top\boldsymbol{\beta}\boldsymbol{\alpha}^\top\boldsymbol{\alpha} - \left(\boldsymbol{\alpha}^\top\boldsymbol{\beta}\right)^2\right]}{\sigma^2 + \sigma_m^2\boldsymbol{\beta}^\top\boldsymbol{\beta}},\\ &= \frac{\boldsymbol{\alpha}^\top\boldsymbol{\alpha}}{\sigma^2}\frac{\sigma^2 + \sigma_m^2\boldsymbol{\beta}^\top\boldsymbol{\beta}\sin^2\left(\psi\right)}{\sigma^2 + \sigma_m^2\boldsymbol{\beta}^\top\boldsymbol{\beta}},\end{aligned} \tag{30}$$

where ψ is the angle between the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Depending on how the sine of ψ grows with universe size, one observes different scaling of ζ_* with respect to p. When the assets all have the same alpha and beta, *i.e.*, $\boldsymbol{\alpha} = \alpha\mathbf{1}$ and $\boldsymbol{\beta} = \beta\mathbf{1}$, the sine is identically zero, and $\zeta_* = \sqrt{\left(p\alpha^2\right)/\left(\sigma^2 + p\sigma_m^2\beta^2\right)} < \alpha\beta^{-1}\sigma_m^{-1}$. Thus ζ_* asymptotically scales slower than p^{ϵ} for all $\epsilon > 0$.

On the other hand, when the sine is one, *i.e.*, when $\boldsymbol{\alpha}$ is orthogonal to $\boldsymbol{\beta}$, $\zeta_* = \sqrt{\boldsymbol{\alpha}^\top\boldsymbol{\alpha}}\sigma^{-1}$, which grows however the assets are ordered, presumably on the order of $p^{\frac{1}{2}}$. Thus under a CAPM model, the growth of ζ_* depends on the 'alignment' of the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

5 Generalizations

Theorem 2.4 is somewhat lacking because it ignores conditioning information which may affect the distribution of future returns, and which may inform the portfolio manager. Few active managers, it is presumed, are holding the unconditional Markowitz portfolio based on in-sample data. What is sought is a more general theorem that allows more elaborate models of returns, and more elaborate, parametrized, trading schemes, with ζ_* redefined as the maximal portfolio signal-noise ratio over the trading schemes, and p redefined as the 'degrees of freedom', perhaps the rank of some derivative at the optimal parameter, say. Towards that goal, a few generalizations can easily be made.

5.1 Conditional portfolio signal-noise ratio

The model of stationary mean returns is generalized by one where the expected return of the assets is linear in some state variables, or 'features', $\boldsymbol{f}_i$, observed prior to the investment decision. [29, 5, 14] That is, one observes the f-vector $\boldsymbol{f}_i$ at some time prior to when the investment decision is required to capture $\boldsymbol{x}_{i+1}$. The general model is now

$$\mathrm{E}\left[\boldsymbol{x}_{i+1}\,|\,\boldsymbol{f}_i\right] = \mathsf{B}\boldsymbol{f}_i, \qquad \mathrm{Var}\left(\boldsymbol{x}_{i+1}\,|\,\boldsymbol{f}_i\right) = \Sigma, \tag{31}$$

where B is some $p \times f$ matrix.

Here we bound the signal-noise ratio of portfolios which are linear in the features $\boldsymbol{f}_i$. That is, the portfolio manager allocates their assets proportional to $\hat{\mathsf{N}}\boldsymbol{f}_i$ for some matrix $\hat{\mathsf{N}}$.

Using the law of iterated expectations, the unconditional expected value of the returns of the portfolio is

$$\mathrm{E}\left[\mathrm{E}\left[\left(\hat{\mathsf{N}}\boldsymbol{f}_i\right)^\top\boldsymbol{x}_{i+1}\,|\,\boldsymbol{f}_i\right]\right] = \mathrm{tr}\left(\hat{\mathsf{N}}^\top\mathsf{B}\,\mathrm{E}\left[\boldsymbol{f}_i\boldsymbol{f}_i^\top\right]\right) = \mathrm{tr}\left(\hat{\mathsf{N}}^\top\mathsf{B}\Gamma_f\right),$$

by definition of $\mathsf{\Gamma}_f$ as the second moment of $\boldsymbol{f}_i$.

Unfortunately the unconditional variance will, in general, involve a term quadratic in the expectation. However, it can easily be shown that the unconditional *expected* variance of the portfolio's returns is

$$\mathrm{E}\left[\left(\hat{\mathsf{N}}\boldsymbol{f}_i\right)^{\top}\mathsf{\Sigma}\left(\hat{\mathsf{N}}\boldsymbol{f}_i\right)\right] = \mathrm{tr}\left(\hat{\mathsf{N}}^{\top}\mathsf{\Sigma}\hat{\mathsf{N}}\mathsf{\Gamma}_f\right).$$

We can then redefine[1] the signal-noise ratio of the portfolio as the unconditional mean divided by the unconditional expected risk:

$$Q\left(\hat{\mathsf{N}}\right) =_{\mathrm{df}} \frac{\mathrm{tr}\left(\hat{\mathsf{N}}^{\top}\mathsf{B}\mathsf{\Gamma}_f\right)}{\sqrt{\mathrm{tr}\left(\hat{\mathsf{N}}^{\top}\mathsf{\Sigma}\hat{\mathsf{N}}\mathsf{\Gamma}_f\right)}}. \tag{32}$$

When $\boldsymbol{f}_i$ is a deterministic scalar constant, this coincides with the 'usual' definition of signal-noise ratio as being like a Sharpe ratio. However, except possibly for an intercept term, one expects $\boldsymbol{f}_i$ to be random, or at least out of the control of the portfolio manager.

Once again, a risk transform can be injected to express portfolio optimization as an estimation problem on a sphere:

$$\begin{aligned} Q\left(\hat{\mathsf{N}}\right) &= \frac{\mathrm{tr}\left(\left(\mathsf{\Sigma}^{\top/2}\hat{\mathsf{N}}{\mathsf{\Gamma}_f}^{1/2}\right)^{\top}\left(\mathsf{\Sigma}^{-1/2}\mathsf{B}{\mathsf{\Gamma}_f}^{1/2}\right)\right)}{\sqrt{\mathrm{tr}\left(\left(\mathsf{\Sigma}^{\top/2}\hat{\mathsf{N}}{\mathsf{\Gamma}_f}^{1/2}\right)^{\top}\left(\mathsf{\Sigma}^{\top/2}\hat{\mathsf{N}}{\mathsf{\Gamma}_f}^{1/2}\right)\right)}}, \\ &= \frac{\mathrm{vec}\left(\mathsf{\Sigma}^{\top/2}\hat{\mathsf{N}}{\mathsf{\Gamma}_f}^{1/2}\right)^{\top}\mathrm{vec}\left(\mathsf{\Sigma}^{-1/2}\mathsf{B}{\mathsf{\Gamma}_f}^{1/2}\right)}{\sqrt{\mathrm{vec}\left(\mathsf{\Sigma}^{\top/2}\hat{\mathsf{N}}{\mathsf{\Gamma}_f}^{1/2}\right)^{\top}\mathrm{vec}\left(\mathsf{\Sigma}^{\top/2}\hat{\mathsf{N}}{\mathsf{\Gamma}_f}^{1/2}\right)}}. \end{aligned} \tag{33}$$

This function is maximized by taking

$$\hat{\mathsf{N}} = \mathsf{N}_* =_{\mathrm{df}} \mathsf{\Sigma}^{-1}\mathsf{B}, \tag{34}$$

which has signal-noise ratio

$$\zeta_* =_{\mathrm{df}} Q\left(\mathsf{N}_*\right) = \sqrt{\mathrm{tr}\left(\mathsf{B}^{\top}\mathsf{\Sigma}^{-1}\mathsf{B}\mathsf{\Gamma}_f\right)}. \tag{35}$$

The square of this quantity, ζ_*^2, is the 'population analogue' of the Hotelling-Lawley trace. [30, 26]

Again we can write

$$\frac{Q\left(\hat{\mathsf{N}}\right)}{\zeta_*} = f_{\mathcal{S}}\left(\mathrm{vec}\left(\mathsf{\Sigma}^{\top/2}\hat{\mathsf{N}}{\mathsf{\Gamma}_f}^{1/2}\right)\right)^{\top} f_{\mathcal{S}}\left(\mathrm{vec}\left(\mathsf{\Sigma}^{-1/2}\mathsf{B}{\mathsf{\Gamma}_f}^{1/2}\right)\right).$$

[1] If an analysis of the conditional expected return divided by risk is required, it is possible one could define $Q\left(\cdot\right)$ as the expected return divided by square root of the unconditional second moment. The signal-noise ratio would then be $\tan\left(\arcsin\left(Q\left(\cdot\right)\right)\right)$. One could possibly find a Cramér-Rao bound on the expected value of this $Q\left(\cdot\right)$. This 'Pillai-Bartlett' form of $Q\left(\cdot\right)$ is likely unrequired for low frequency settings.

Thus finding a 'good' $\hat{\mathsf{N}}$ becomes an estimation problem on the sphere $\mathcal{S}^{fp-1}$. An analogue to Theorem 2.4 can be proved with fp replacing p, by assuming a particular form to the likelihood. We must generalize the assumption of Directional Independence, after which the theorem proceeds easily.

Assumption 5.1 (Conditional Directional Independence)**.** Assume that

$$\mathrm{E}\left[f_{\mathcal{S}}\left(\mathrm{vec}\left(\Sigma^{\top/2}\hat{\mathsf{N}}\left(\mathsf{X},\mathsf{F}\right)\right)\right)\right] = c_n\left(\zeta_*^2\right) f_{\mathcal{S}}\left(\mathrm{vec}\left(\Sigma^{\top/2}\mathsf{N}_*{\Gamma_f}^{1/2}\right)\right)+\boldsymbol{B}_n\left(\boldsymbol{\mu},\Sigma,\Gamma_f\right), \tag{36}$$

where $\boldsymbol{B}_n\left(\boldsymbol{\mu},\Sigma,\Gamma_f\right)$ is the bias term, orthogonal to $f_{\mathcal{S}}\left(\mathrm{vec}\left(\Sigma^{\top/2}\mathsf{N}_*{\Gamma_f}^{1/2}\right)\right)$.

Theorem 5.2. *Let one element of $\boldsymbol{f}_i$ be a deterministic* 1*. Suppose the vector of the remaining $f-1$ elements of $\boldsymbol{f}_i$ stacked on top of $\boldsymbol{x}_{i+1}$ are multivariate Gaussian. Let X, F be $n \times p$ and $n \times f$ matrices of* i.i.d. *observations of the features and returns. Let $\hat{\mathsf{N}}\left(\mathsf{X},\mathsf{F}\right)$ be an estimator satisfying the assumptions of Conditional Directional Independence and Residual Independence. Then*

$$\mathrm{E}\left[Q\left(\hat{\mathsf{N}}\left(\mathsf{X},\mathsf{F}\right)\right)\right] \le \frac{\sqrt{n}\zeta_*^2}{\sqrt{fp-1+n\zeta_*^2}}. \tag{37}$$

Proof. We can proceed as in Section 2. Let X be the $n \times p$ matrix of portfolio returns, and let F be the corresponding $n \times f$ matrix of features. View the portfolio coefficient $\hat{\mathsf{N}}$ as an estimator, a function of the random data, *i.e.*, $\hat{\mathsf{N}}\left(\mathsf{X},\mathsf{F}\right)$. Define

$$\mathsf{H} =_{\mathrm{df}} \Sigma^{-1/2}\mathsf{B}{\Gamma_f}^{1/2}. \tag{38}$$

Then

$$\zeta_*^2 = \mathrm{tr}\left(\mathsf{H}^\top\mathsf{H}\right).$$

We get, analogously to Equation 12,

$$\frac{1}{n}\mathrm{tr}\left(\mathsf{D}\mathcal{I}_{\mathrm{vec}(\mathsf{H})}^{-1}\mathsf{D}^\top\right) \le 1 - c_n^2\left(\mathrm{tr}\left(\mathsf{H}^\top\mathsf{H}\right)\right), \tag{39}$$

where

$$\mathsf{D} =_{\mathrm{df}} \frac{\mathrm{d}c_n\left(\mathrm{tr}\left(\mathsf{H}^\top\mathsf{H}\right)\right)\frac{\mathsf{H}}{\sqrt{\mathrm{tr}(\mathsf{H}^\top\mathsf{H})}}}{\mathrm{d}\,\mathrm{vec}\left(\mathsf{H}\right)}. \tag{40}$$

Without loss of generality, we assume it is the first element of $\boldsymbol{f}_i$ that is a deterministic 1. Then, the log likelihood of the vector of $\boldsymbol{f}_i$ stacked on top of $\boldsymbol{x}_{i+1}$ is: [29]

$$\log\mathcal{L}\left(\left[\begin{array}{c}\boldsymbol{f}_i\\ \boldsymbol{x}_{i+1}\end{array}\right]|\Theta\right) = c_{f+p} - \frac{1}{2}\log\left|\Theta\right| - \frac{1}{2}\mathrm{tr}\left(\Theta^{-1}\left[\begin{array}{c}\boldsymbol{f}_i\\ \boldsymbol{x}_{i+1}\end{array}\right]\left[\begin{array}{c}\boldsymbol{f}_i\\ \boldsymbol{x}_{i+1}\end{array}\right]^\top\right), \tag{41}$$

where Θ is the second moment matrix:

$$\Theta =_{\mathrm{df}} \mathrm{E}\left[\left[\begin{array}{c}\boldsymbol{f}_i\\ \boldsymbol{x}_{i+1}\end{array}\right]\left[\begin{array}{c}\boldsymbol{f}_i\\ \boldsymbol{x}_{i+1}\end{array}\right]^\top\right] = \left[\begin{array}{cc}\Gamma_f & \Gamma_f\mathsf{B}^\top\\ \mathsf{B}\Gamma_f & \Sigma+\mathsf{B}\Gamma_f\mathsf{B}^\top\end{array}\right]. \tag{42}$$

The inverse of Θ has the following, somewhat surprising, form [29]:

$$\Theta^{-1} = \left[\begin{array}{cc}{\Gamma_f}^{-1}+\mathsf{B}^\top\Sigma^{-1}\mathsf{B} & -\mathsf{B}^\top\Sigma^{-1}\\ -\Sigma^{-1}\mathsf{B} & \Sigma^{-1}\end{array}\right]. \tag{43}$$

A square root of this matrix (a Cholesky factor, up to permutation) is:

$$\begin{aligned}\Theta^{-1} &= \left[\begin{array}{cc}\Gamma_f{}^{-1/2} & -\mathsf{B}^\top\Sigma^{-1/2}\\ 0 & \Sigma^{-1/2}\end{array}\right]\left[\begin{array}{cc}\Gamma_f{}^{-1/2} & -\mathsf{B}^\top\Sigma^{-1/2}\\ 0 & \Sigma^{-1/2}\end{array}\right]^\top,\\ &= \left[\begin{array}{cc}\Gamma_f{}^{-1/2} & 0\\ 0 & \mathsf{I}\end{array}\right]\left[\begin{array}{cc}\mathsf{I} & -\mathsf{H}^\top\\ 0 & \Sigma^{-1/2}\end{array}\right]\left[\begin{array}{cc}\mathsf{I} & -\mathsf{H}^\top\\ 0 & \Sigma^{-1/2}\end{array}\right]^\top\left[\begin{array}{cc}\Gamma_f{}^{-\top/2} & 0\\ 0 & \mathsf{I}\end{array}\right].\end{aligned} \tag{44}$$

By the block determinant formula,

$$\left|\Theta\right| = \left|\Gamma_f\right|\left|\Sigma + \mathsf{B}\Gamma_f\mathsf{B}^\top - \mathsf{B}\Gamma_f\Gamma_f{}^{-1}\Gamma_f\mathsf{B}^\top\right| = \left|\Gamma_f\right|\left|\Sigma\right|. \tag{45}$$

Thus, conditional on Γ_f and Σ, the negative log likelihood takes the form:

$$\begin{aligned}-\log\mathcal{L}\left(\left[\begin{array}{c}\boldsymbol{f}_i\\ \boldsymbol{x}_{i+1}\end{array}\right]|\mathsf{H},\Gamma_f,\Sigma\right) &= -c_{f+p} + \frac{1}{2}\log\left|\Gamma_f\right| + \frac{1}{2}\log\left|\Sigma\right|\\ +\frac{1}{2}\operatorname{tr}\left(\left[\begin{array}{cc}\mathsf{I} & -\mathsf{H}^\top\\ 0 & \Sigma^{-1/2}\end{array}\right]\left[\begin{array}{cc}\mathsf{I} & -\mathsf{H}^\top\\ 0 & \Sigma^{-1/2}\end{array}\right]^\top\right.&\left.\left[\begin{array}{c}\Gamma_f{}^{-\top/2}\boldsymbol{f}_i\\ \boldsymbol{x}_{i+1}\end{array}\right]\left[\begin{array}{c}\Gamma_f{}^{-\top/2}\boldsymbol{f}_i\\ \boldsymbol{x}_{i+1}\end{array}\right]^\top\right).\end{aligned} \tag{46}$$

Sweeping the nuisance parameter terms into the constant, as well as terms in the trace which are not quadratic in H, we have

$$-\log\mathcal{L}\left(\left[\begin{array}{c}\boldsymbol{f}_i\\ \boldsymbol{x}_{i+1}\end{array}\right]|\mathsf{H},\Gamma_f,\Sigma\right) = -c' + \frac{1}{2}\operatorname{tr}\left(\mathsf{H}^\top\mathsf{H}\left(\Gamma_f{}^{-\top/2}\boldsymbol{f}_i\right)\left(\Gamma_f{}^{-\top/2}\boldsymbol{f}_i\right)^\top\right), \tag{47}$$

$$= -c' + \frac{1}{2}\operatorname{vec}\left(\mathsf{H}\right)^\top\operatorname{vec}\left(\mathsf{H}\Gamma_f{}^{-\top/2}\boldsymbol{f}_i\boldsymbol{f}_i{}^\top\Gamma_f{}^{-1/2}\right). \tag{48}$$

$$= -c' + \frac{1}{2}\operatorname{vec}\left(\mathsf{H}\right)^\top\left(\left[\Gamma_f{}^{-\top/2}\boldsymbol{f}_i\boldsymbol{f}_i{}^\top\Gamma_f{}^{-1/2}\right]\otimes\mathsf{I}\right)\operatorname{vec}\left(\mathsf{H}\right). \tag{49}$$

The Fisher Information, then, is

$$\mathcal{I}_{\operatorname{vec}(\mathsf{H})} = \operatorname{E}\left[\left(\left[\Gamma_f{}^{-\top/2}\boldsymbol{f}_i\boldsymbol{f}_i{}^\top\Gamma_f{}^{-1/2}\right]\otimes\mathsf{I}\right)\right] = \mathsf{I}_{fp}. \tag{50}$$

The remainder of the proof proceeds exactly as in Section 2. ☐

5.2 Subspace constraints

Consider, now, the case of conditional expectation, as presented in Section 5.1, but where the portfolio is constrained to be in some lower dimensional subspace. That is, by design,

$$\mathsf{J}^\perp\hat{\mathsf{N}}\left(\mathsf{X},\mathsf{F}\right) = \mathbf{0}, \tag{51}$$

where $\mathsf{J}^\perp$ is a $(p - p_j)\times p$ matrix of rank $p - p_j$, that is chosen indpendently of the observations of X and F. Let the rows of J span the null space of the rows of $\mathsf{J}^\perp$; that is, $\mathsf{J}^\perp\mathsf{J}^\top = 0$, and $\mathsf{J}\mathsf{J}^\top = \mathsf{I}$.

We can simply use the results of Section 5.1, but replacing the assets with the p_j assets spanned by the rows of J. That is, we can replace the $\boldsymbol{x}_{i+1}$ with $\mathsf{J}\boldsymbol{x}_{i+1}$, and replace $\hat{\mathsf{N}}\left(\mathsf{X},\mathsf{F}\right)$ with $\mathsf{J}^\top\left(\mathsf{J}\mathsf{J}^\top\right)^{-1}\hat{\mathsf{N}}\left(\mathsf{X},\mathsf{F}\right)$ to arrive at the following analogue of Theorem 5.2:

Theorem 5.3. *Let one element of $\boldsymbol{f}_i$ be a deterministic* 1*. Suppose the vector of the remaining $f-1$ elements of $\boldsymbol{f}_i$ stacked on top of $\boldsymbol{x}_{i+1}$ are multivariate Gaussian. Let X, F be $n \times p$ and $n \times f$ matrices of* i.i.d. *observations of the features and returns. Let $\hat{\mathsf{N}}\left(\mathsf{X},\mathsf{F}\right)$ be an estimator satisfying the assumptions of directional independence and residual independence, with the constraint*

$$\mathsf{J}^{\perp}\hat{\mathsf{N}}\left(\mathsf{X},\mathsf{F}\right) = \mathbf{0}, \tag{52}$$

for $(p-p_j) \times p$ matrix $\mathsf{J}^{\perp}$, which is chosen independently of the observed X and F. Let the rows of J span the null space of the rows of $\mathsf{J}^{\perp}$.

Then

$$\mathrm{E}\left[Q\left(\hat{\mathsf{N}}\left(\mathsf{X},\mathsf{F}\right)\right)\right] \le \frac{\sqrt{n}\zeta_{*,\mathsf{J}}^2}{\sqrt{fp_j - 1 + n\zeta_{*,\mathsf{J}}^2}}, \tag{53}$$

where

$$\zeta_{*,\mathsf{J}}^2 =_{\mathrm{df}} \operatorname{tr}\left(\mathsf{B}^{\top}\mathsf{J}^{\top}\left(\mathsf{J}\mathsf{\Sigma}\mathsf{J}^{\top}\right)^{-1}\mathsf{J}\mathsf{B}\mathsf{\Gamma}_f\right).$$

5.3 Hedging constraints

Consider, now, the case where one seeks a portfolio whose returns are independent, in the probabilistic sense, of the returns of some traded instruments in the investment universe. Independence is a difficult property to check or enforce; however, independence implies zero covariation, which can be easily formulated and checked.

Since the portfolio estimator may not deliver a perfectly hedged portolio due to misestimation of the covariance matrix, we will, with perfect knowledge of $\mathsf{\Sigma}$, consider the signal-noise ratio of the hedged part of the portfolio. The hedged part is defined in terms of a risk projection. If $\hat{\boldsymbol{\nu}}_1$ is a feasible portfolio based on the sample, then the hedged version of this portfolio is the solution to the optimization problem

$$\min_{\hat{\boldsymbol{\nu}}:\mathsf{G}\mathsf{\Sigma}\hat{\boldsymbol{\nu}}=\mathbf{0}} \operatorname{Var}\left(\left(\hat{\boldsymbol{\nu}} - \hat{\boldsymbol{\nu}}_1\right)^{\top}\boldsymbol{x}_{i+1}\right), \tag{54}$$

where G is a $p_g \times p$ matrix of rank p_g, the rows of which we wish to 'hedge out.'

Using the Lagrange multiplier technique, this can easily be found to be solved by

$$\hat{\boldsymbol{\nu}} = \hat{\boldsymbol{\nu}}_1 - \mathsf{G}^{\top}\left(\mathsf{G}\mathsf{\Sigma}\mathsf{G}^{\top}\right)^{-1}\mathsf{G}\mathsf{\Sigma}\hat{\boldsymbol{\nu}}_1. \tag{55}$$

Thus we will consider the signal-noise ratio of the portfolio estimator

$$\left(\mathsf{I}_p - \mathsf{G}^{\top}\left(\mathsf{G}\mathsf{\Sigma}\mathsf{G}^{\top}\right)^{-1}\mathsf{G}\mathsf{\Sigma}\right)\hat{\mathsf{N}}\left(\mathsf{X},\mathsf{F}\right).$$

Note, however, that the row rank of $\left(\mathsf{I}_p - \mathsf{G}^{\top}\left(\mathsf{G}\mathsf{\Sigma}\mathsf{G}^{\top}\right)^{-1}\mathsf{G}\mathsf{\Sigma}\right)$ is $p-p_g$. Thus hedging is an instance of a subspace constraint and we can apply Theorem 5.3 outright.

Theorem 5.4. *Let one element of $\boldsymbol{f}_i$ be a deterministic* 1*. Suppose the vector of the remaining $f-1$ elements of $\boldsymbol{f}_i$ stacked on top of $\boldsymbol{x}_{i+1}$ are multivariate Gaussian. Let X, F be $n \times p$ and $n \times f$ matrices of* i.i.d. *observations of the features and returns. Let $\hat{\mathsf{N}}\left(\mathsf{X},\mathsf{F}\right)$ be an estimator satisfying the assumptions* *of directional independence and residual independence. Let $p_g \times p$ matrix G be chosen independently of X and F.*

Define

$$\Delta_{\mathsf{I},\mathsf{G}}\zeta_*^2 =_{\mathrm{df}} \operatorname{tr}\left(\mathsf{B}^\top\hat{\Sigma}^{-1}\mathsf{B}\Gamma_f\right) - \operatorname{tr}\left(\mathsf{B}^\top\mathsf{G}^\top\left(\mathsf{G}\Sigma\mathsf{G}^\top\right)^{-1}\mathsf{G}\mathsf{B}\Gamma_f\right). \tag{56}$$

Then

$$\operatorname{E}\left[Q\left(\left[\mathsf{I}_p - \mathsf{G}^\top\left(\mathsf{G}\Sigma\mathsf{G}^\top\right)^{-1}\mathsf{G}\Sigma\right]\hat{\mathsf{N}}\left(\mathsf{X},\mathsf{F}\right)\right)\right] \leq \frac{\sqrt{n}\Delta_{\mathsf{I},\mathsf{G}}\zeta_*^2}{\sqrt{f\left(p-p_g\right)-1+n\Delta_{\mathsf{I},\mathsf{G}}\zeta_*^2}}. \tag{57}$$

6 Examples

6.1 The equal weight puzzle

Theorem 2.4 can help us make sense of puzzling findings in the literature. For example, in the "$1/N$" paper, DeMiguel *et al.* find that the equal-weighting portfolio outperforms, in terms of out-of-sample Sharpe ratio (and other measures), the Markowitz portfolio and numerous other portfolio estimators. [7] This finding is supported on a number of real world data sets, and a few synthetic ones. One data set used was the returns of the 10 industry portfolios and the US equity market portfolio, computed by Ken French.

The monthly returns, from 1927-01-01 to 2016-12-01, for these 10 assets were downloaded from Kenneth French's data library. [8] The Sharpe ratio of the equal weighted portfolio on the assets, over the 1080 months, is around $0.66\mathrm{yr}^{-1/2}$. The Sharpe ratio of the sample Markowitz portfolio over the 10 assets over the same period is around $0.89\mathrm{yr}^{-1/2}$. [28] Now consider a portfolio estimator given 5 years of observations, as in DeMiguel *et al.* [7], assuming $\zeta_* = 0.89\mathrm{yr}^{-1/2}$. The bound on expected value of $q\left(\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)$ from Theorem 2.4 is only $0.49\mathrm{yr}^{-1/2}$. Under Approximation 25, the probability that $q\left(\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)$ exceeds $0.66\mathrm{yr}^{-1/2}$ in this case is only 0.16. It is not surprising that DeMiguel *et al.* drew the conclusions they did, nor that they would be refuted by looking at a longer sample, as by Kritzman *et al.* [18]

One could also use Theorem 5.4 here. However, the upper bound of that theorem is non-negative, and zero only if the quantity $\Delta_{\mathsf{I},\mathsf{G}}\ \zeta_*^2$ is zero. This is a statement regarding unknown population parameters, but we can perform inference on this quantity. For example, based on the 1080 months of data on these 10, the 95% confidence interval on $\Delta_{\mathsf{I},\mathsf{G}}\ \zeta_*^2$, where G is the 1×10 matrix of all ones, is $[0.07, 0.54]\,\mathrm{yr}^{-1}$, under the assumption of Gaussian returns. [28]

6.2 Empirical diversification in the S&P 100

To check how ζ_* *might* scale with p, the weekly log returns of the adjusted close prices of the stocks in the S&P 100 Index, as of March 21, 2014, were downloaded from *Quandl.* [23] Adjustments for splits and dividends were made in some unspecified way by the upstream source of the data, Yahoo Finance. Stocks without a full 5 years of history were discarded, leaving 96 stocks. Note that selection based on membership in the index at the end of the period adds no small amount of selection bias, which we shall ignore here.

Based on the weekly returns from 2009-03-27 to 2014-04-04, estimates of ζ_* were computed, using the 'KRS' estimator. [19, 28] This was performed on the

Figure 6: Growth of estimated ζ_* versus p for the S&P 100 Index names, in alphabetical order, showing the 'Apple effect.'

first p assets, with p ranging from 1 to 96. The estimate of ζ_* versus p is plotted in Figure 6, with assets added in alphabetical order. Because Apple appears at the beginning of this list, it appears that ζ_* starts reasonably large, but then actually *decreases* when adding assets. This is an artifact of the estimator, since the true ζ_* can only increase when adding assets.

Since the ordering of assets here is arbitrary, the experiment was repeated 1000 times, with the stocks randomly permuted, and ζ_* estimated as a function of p. Boxplots, over the 1000 simulations, of the KRS statistic versus p are given in Figure 7. There is effectively no diversification benefit observed here beyond the mean effect, which is equivalent to holding an equal weight portfolio. Given the conditions under which signal-noise ratio grows with p outlined in Section 4, one expects poor performance of directionally independent portfolio estimators over even a small subset of the S&P 100.

7 Discussion

Care should be taken in the interpretation of Theorem 2.4, or its generalizations from Section 5. It does not claim that the sample Markowitz portfolio is somehow 'optimal,' nor does it make comparative claims about different portfolio estimators when presented with the same data. The theorem does not imply that somehow 'overfitting' to the observed data can be mitigated by selecting a less desireable portfolio. It does not claim that sample estimates of the signal-noise ratio of a portfolio are useless. It is trivially the case, for example, that if $q\left(\hat{\boldsymbol{\nu}}_1\right) > q\left(\hat{\boldsymbol{\nu}}_2\right)$, then, with probability greater than half, ${\hat{\boldsymbol{\nu}}_1}^{\top}\hat{\boldsymbol{\mu}}/\sqrt{{\hat{\boldsymbol{\nu}}_1}^{\top}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\nu}}_1} > {\hat{\boldsymbol{\nu}}_2}^{\top}\hat{\boldsymbol{\mu}}/\sqrt{{\hat{\boldsymbol{\nu}}_2}^{\top}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\nu}}_2}$, where the probability is over draws of $\hat{\boldsymbol{\mu}}$

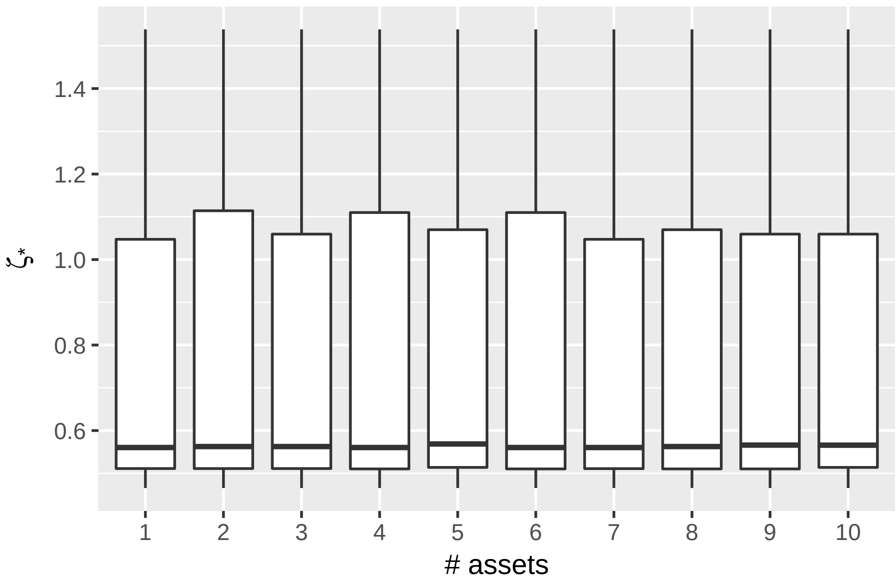

Figure 7: Growth of estimated ζ_* versus p for the S&P 100 Index names is shown over 1000 permutations of the stocks. There is effectively *no* diversification benefit here beyond an equal weight portfolio.

and $\hat{\Sigma}$. The theorem does not claim that the expected signal-noise ratio of a portfolio estimator is negative. (Indeed, it can not, since the portfolio estimator which generates a random portfolio, ignoring the data, has zero expected signal-noise ratio). The theorem makes no claims (*e.g.*, providing a Bayesian posterior) about any particular portfolio based on a single sample of the data: it is a statement about the expectation of the *estimator* under replication of draws of the sample.

One should recognize, moreover, there are situations where the assumptions of the theorem are violated. For example, in some cases a prior bias for positive expected returns, *i.e.*, $\boldsymbol{\mu} \geq 0$, is warranted, and thus a portfolio estimator with a long bias is chosen. This can happen when the underlying assets are equities, and the eligible universe is based on some minimum longevity, as this introduces a 'good' survivorship bias: companies with negative expected return should founder and perish, leaving behind those with more positive μ. Effectively this acts to boost n somewhat, although the effect is likely small.

There are other reasonable portfolio estimators which violate the assumption of Directional Independence. For example, an estimator which performs some dimensionality reduction based on the observed data, X and F will not be covered by Theorem 5.3 since the subspace is chosen based on the sample. However, it might not be covered by Theorem 5.2 because the expected signal-noise ratio might depend on how B aligns with the leading eigenvectors of Σ, say.

7.1 Future work

These findings perhaps raise more questions than they answer:

1. Foremost, the bounds of Theorem 2.4 and Theorem 5.2 depend on the unknown quantity, ζ_*^2. How can we perform inference, Frequentist or Bayesian, on $q(\hat{\boldsymbol{\nu}}_*)$, where $\hat{\boldsymbol{\nu}}_*$ is the Markowitz portfolio, given the observed information (*viz.* $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$)? This is a problem of enormous practical concern to hundreds of quantitative portfolio managers.
   Contrast inference on the portfolio signal-noise ratio with inference on the population signal-noise ratio: under Gaussian returns, the distribution of $\hat{\zeta}_*^2$ in terms of n, p and ζ_*^2 is known. [1, Theorem 5.2.2] Thus, for example, the quantity $(1 - p/n)\,\hat{\zeta}_*^2 - p/n$ is an unbiased estimator for ζ_*^2, *etc.* Performing inverence on $q(\hat{\boldsymbol{\nu}}_*)$ is tricky because ζ_*^2 is unknown and the error $\hat{\boldsymbol{\nu}}_* - \boldsymbol{\nu}_*$ is likely not independent of the error in the estimate $\hat{\zeta}_*$. It may be the case, however, that inference on the portfolio signal-noise ratio qualifies as an 'impossible' estimation-after-selection problem. [20]
2. While Theorem 2.4 requires Gaussian returns, one expects that the result holds for returns distributions whose likelihood is "more concave" than the Gaussian at the MLE. Exact conditions for this to hold should be established.
3. Theorem 5.2 applies to the case of trading strategies where the portfolio is linear in the observable features, $\boldsymbol{f}_i$. Can it be used as an approximate bound for trading strategies which are nonlinear, complex functions of the features?
4. What can be said about scaling of ζ_* with respect to p for different models of market returns? Can one establish sane sufficient conditions for which ζ_* grows slower than $p^{\frac{1}{4}}$? What is the analogue of Equation 30 for a multi-factor model of returns?
5. Can we find a lower bound, or a non-trivial upper bound on the variance of $q(\hat{\boldsymbol{\nu}}(\mathsf{X}))$? Together these could be used to give guarantees about the quantiles of $q(\hat{\boldsymbol{\nu}}(\mathsf{X}))$. A lower bound on the variance can likely be had via a result of Kakarala and Watson. [16] Together with Cantelli's Inequality, these would give rough (perhaps useless) upper bounds on the s^{th} quantile of portfolio signal-noise ratio, for $\frac{1}{2} < s < 1$.
6. How tight is the bound of Theorem 2.4, and can it be much improved by directly analyzing the differential inequality of Equation 19, rather than discarding the derivative term? Or perhaps the bound can be improved by using an 'intrinsic' Cramér-Rao bound. [33]
7. How good is Approximation 25? Can we find the expected value of the distribution in Approximation 25, and what is the gap between it and the bound of Theorem 2.4? Can we find the *exact* distribution of signal-noise ratio of the sample Markowitz portfolio under Gaussian returns, perhaps leveraging the work of Bodnar and Okhrin, or of Britton-Jones. [2, 4]
8. Can the assumption of Directional Independence be weakened? Can the Theorem 5.2 be generalized to deal with omitted variable bias in $\boldsymbol{f}_i$?
9. The analysis of signal-noise ratio ignores the 'risk-free' or 'disastrous' rate of return, and all trading costs. Can the expected bounds be generalized to include these costs?

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] 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.

[3] Michael W Brandt. Portfolio choice problems. *Handbook of financial econometrics*, 1:269–336, 2009. URL https://faculty.fuqua.duke.edu/~mbrandt/papers/published/portreview.pdf.

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

[5] Gregory Connor. Sensible return forecasting for portfolio management. *Financial Analysts Journal*, 53(5):pp. 44–51, 1997. ISSN 0015198X. URL https://faculty.fuqua.duke.edu/~charvey/Teaching/BA453_2006/Connor_Sensible_Return_Forecasting_1997.pdf.

[6] 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.

[7] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22(5):1915–1953, 2009. URL http://faculty.london.edu/avmiguel/DeMiguel-Garlappi-Uppal-RFS.pdf.

[8] Kenneth R. French. 10 industry portfolios. Privately Published, 2013. URL http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_10_ind_port.html.

[9] 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.

[10] 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.

[11] 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.

[12] H. W. M. Hendriks, J. H. M. Janssen, and F. H. Ruymgaart. A Cramér-Rao type inequality for random variables in Euclidean manifolds. *Sankhyâ: The Indian Journal of Statistics, Series A (1961-2002)*, 54(3):pp. 387–401, 1992. ISSN 0581572X. URL http://www.jstor.org/stable/25050892.

[13] Harrie Hendriks. A Cramér-Rao type lower bound for estimators with values in a manifold. *Journal of Multivariate Analysis*, 38(2):245 – 261, 1991. ISSN 0047-259X. doi: http://dx.doi.org/10.1016/0047-259X(91)90044-3. URL http://www.sciencedirect.com/science/article/pii/0047259X91900443.

[14] Ulf Herold and Raimond Maurer. Tactical asset allocation and estimation risk. *Financial Markets and Portfolio Management*, 18(1):39–57, 2004. ISSN 1555-4961. doi: 10.1007/s11408-004-0104-2. URL http://dx.doi.org/10.1007/s11408-004-0104-2.

[15] 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.

[16] Ramakrishna Kakarala and Geoffrey S. Watson. Lower bounds for the divergence of directional and axial estimators. *Australian Journal of Statistics*, 39(3):253–260, 1997. ISSN 1467-842X. doi: 10.1111/j.1467-842X.1997.tb00690.x. URL http://dx.doi.org/10.1111/j.1467-842X.1997.tb00690.x.

[17] Helena Chmura Kraemer and Minja Paik. A central t approximation to the noncentral t distribution. *Technometrics*, 21(3):pp. 357–360, 1979. ISSN 00401706. URL http://www.jstor.org/stable/1267759.

[18] Mark Kritzman, Sébastien Page, and David Turkington. In defense of optimization: The fallacy of 1/ N. *Financial Analysts Journal*, 66(2):pp. 31–39, 2010. ISSN 0015198X. URL http://www.jstor.org/stable/27809177.

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

[20] Hannes Leeb and Benedikt M. Pötscher. Can one estimate the conditional distribution of post-model-selection estimators? *The Annals of Statistics*, 34(5):2554–2591, 10 2006. doi: 10.1214/009053606000000821. URL http://projecteuclid.org/euclid.aos/1169571807.

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

[22] 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.

[23] Raymond McTaggart and Gergely Daroczi. *Quandl: Quandl Data Connection*, 2014. URL http://CRAN.R-project.org/package=Quandl. R package version 2.4.0.

[24] Richard O. Michaud. The Markowitz optimization enigma: is 'optimized' optimal? *Financial Analysts Journal*, pages 31–42, 1989. URL http://newfrontieradvisors.com/Research/Articles/documents/markowitz-optimization-enigma-010189.pdf.

[25] Terrence Joseph Moore, Jr. *A theory of Cramér-Rao bounds for constrained parametric models*. PhD thesis, University of Maryland, College Park, 2010. URL http://drum.lib.umd.edu/bitstream/1903/10290/1/Moore_umd_0117E_11120.pdf.

[26] Keith E. Muller and Bercedis L. Peterson. Practical methods for computing power in testing the multivariate general linear hypothesis. *Computational Statistics & Data Analysis*, 2(2):143–158, 1984. ISSN 0167-9473. doi: 10.1016/0167-9473(84)90002-1. URL http://www.sciencedirect.com/science/article/pii/0167947384900021.

[27] Frank W. Olver, Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark. *NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, NY, USA, 1st edition, 2010. ISBN 0521140633, 9780521140638. URL http://dlmf.nist.gov/.

[28] Steven E. Pav. *SharpeR: Statistical significance of Sharpe ratio*, 2013. URL https://github.com/shabbychef/SharpeR. R package version 0.1401.

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

[30] Alvin C. Rencher. *Methods of Multivariate Analysis*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. J. Wiley, 2002. ISBN 9780471418894. URL http://books.google.com/books?id=SpvBd7IUCxkC.

[31] Jun Tu and Guofu Zhou. Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, 99(1):204–215, 2011. URL http://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=2104&context=lkcsb_research.

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

[33] João Manuel Freitas Xavier and Victor A. N. Barroso. Intrinsic variance lower bound (IVLB): an extension of the Cramér-Rao bound to Riemannian manifolds. In *ICASSP (5)*, pages 1033–1036. IEEE, 2005. ISBN 0-7803-8874-7. URL http://users.isr.ist.utl.pt/~jxavier/icassp2005.pdf.