Asymptotic Distribution of the Markowitz Portfolio

Steven E. Pav *

December 18, 2024

Abstract

The asymptotic distribution of the Markowitz portfolio, $\hat{\mathbf{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$, is derived, for the general case (assuming fourth moments of returns exist), and for the case of multivariate normal returns. The derivation allows for inference which is robust to heteroskedasticity and autocorrelation of moments up to order four. As a side effect, one can estimate the proportion of error in the Markowitz portfolio due to mis-estimation of the covariance matrix. A likelihood ratio test is given which generalizes Dempster's Covariance Selection test to allow inference on linear combinations of the precision matrix and the Markowitz portfolio. [12] Extensions of the main method to deal with hedged portfolios, conditional heteroskedasticity, and conditional expectation are given.

1 Introduction

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

$$\boldsymbol{\nu}_* =_{\mathrm{df}} \lambda \mathbf{\Sigma}^{-1}\boldsymbol{\mu} \tag{1}$$

plays a special role in modern portfolio theory. [26, 4, 7] It is known as the 'efficient portfolio', the 'tangency portfolio', and, somewhat informally, the 'Markowitz portfolio'. It appears, for various λ, in the solution to numerous portfolio optimization problems. Besides the classic mean-variance formulation, it solves the (population) Sharpe ratio maximization problem:

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

where $r_0 \ge 0$ is the risk-free, or 'disastrous', rate of return, and $R > 0$ is some given 'risk budget'. The solution to this optimization problem is $\lambda\mathbf{\Sigma}^{-1}\boldsymbol{\mu}$, where $\lambda = R/\sqrt{\boldsymbol{\mu}^{\top}\mathbf{\Sigma}^{-1}\boldsymbol{\mu}}$.

In practice, the Markowitz portfolio has a somewhat checkered history. The population parameters $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are not known and must be estimated from samples. Estimation error results in a feasible portfolio, $\hat{\boldsymbol{\nu}}_*$, of dubious value. Michaud went so far as to call mean-variance optimization, "error maximization." [30] It has been suggested that simple portfolio heuristics outperform the Markowitz portfolio in practice. [10]

*spav@alumni.cmu.edu

This paper focuses on the asymptotic distribution of the sample Markowitz portfolio. By formulating the problem as a linear regression, Britten-Jones very cleverly devised hypothesis tests on elements of $\boldsymbol{\nu}_*$, assuming multivariate Gaussian returns. [5] In a remarkable series of papers, Okhrin and Schmid, and Bodnar and Okhrin give the (univariate) density of the dot product of $\boldsymbol{\nu}_*$ and a deterministic vector, again for the case of Gaussian returns. [35, 2] Okhrin and Schmid also show that all moments of $\hat{\boldsymbol{\nu}}_*/\mathbf{1}^{\top}\hat{\boldsymbol{\nu}}_*$ of order greater than or equal to one do not exist. [35]

Here I derive asymptotic normality of $\hat{\boldsymbol{\nu}}_*$, the sample analogue of $\boldsymbol{\nu}_*$, assuming only that the first four moments exist. Feasible estimation of the variance of $\hat{\boldsymbol{\nu}}_*$ is amenable to heteroskedasticity and autocorrelation robust inference. [47] The asymptotic distribution under Gaussian returns is also derived.

After estimating the covariance of $\hat{\boldsymbol{\nu}}_*$, one can compute Wald test statistics for the elements of $\hat{\boldsymbol{\nu}}_*$, possibly leading one to drop some assets from consideration ('sparsification'). Having an estimate of the covariance can also allow portfolio shrinkage. [11, 20]

The derivations in this paper actually solve a more general problem than the distribution of the sample Markowitz portfolio. The covariance of $\hat{\boldsymbol{\nu}}_*$ and the 'precision matrix,' $\hat{\Sigma}^{-1}$ are derived. This allows one, for example, to estimate the proportion of error in the Markowitz portfolio attributable to mis-estimation of the covariance matrix. According to lore, the error in portfolio weights is mostly attributable to mis-estimation of $\boldsymbol{\mu}$, not of Σ. [6, 29]

Finally, assuming Gaussian returns, a likelihood ratio test for performing inference on linear combinations of elements of the Markowitz portfolio and the precision matrix is derived. This test generalizes a procedure by Dempster for inference on the precision matrix alone. [12]

2 The augmented second moment

Let $\boldsymbol{x}$ be an array of returns of p assets, with mean $\boldsymbol{\mu}$, and covariance Σ. Let $\tilde{\boldsymbol{x}}$ be $\boldsymbol{x}$ prepended with a 1: $\tilde{\boldsymbol{x}} = \left[1, \boldsymbol{x}^{\top}\right]^{\top}$. Consider the second moment of $\tilde{\boldsymbol{x}}$:

$$\Theta =_{\mathrm{df}} \mathrm{E}\left[\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\top}\right] = \left[\begin{array}{cc} 1 & \boldsymbol{\mu}^{\top} \\ \boldsymbol{\mu} & \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^{\top} \end{array}\right]. \tag{3}$$

By inspection one can confirm that the inverse of Θ is

$$\Theta^{-1} = \left[\begin{array}{cc} 1 + \boldsymbol{\mu}^{\top}\Sigma^{-1}\boldsymbol{\mu} & -\boldsymbol{\mu}^{\top}\Sigma^{-1} \\ -\Sigma^{-1}\boldsymbol{\mu} & \Sigma^{-1} \end{array}\right] = \left[\begin{array}{cc} 1 + \zeta_*^2 & -\boldsymbol{\nu}_*{}^{\top} \\ -\boldsymbol{\nu}_* & \Sigma^{-1} \end{array}\right], \tag{4}$$

where $\boldsymbol{\nu}_* = \Sigma^{-1}\hat{\boldsymbol{\mu}}$ is the Markowitz portfolio, and $\zeta_* = \sqrt{\boldsymbol{\mu}^{\top}\Sigma^{-1}\boldsymbol{\mu}}$ is the Sharpe ratio of that portfolio. The matrix Θ contains the first and second moment of $\boldsymbol{x}$, but is also the uncentered second moment of $\tilde{\boldsymbol{x}}$, a fact which makes it amenable to analysis via the central limit theorem.

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

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

where $\hat{\boldsymbol{\mu}}$, $\hat{\Sigma}$ are some sample estimates of $\boldsymbol{\mu}$ and Σ, and $\hat{\boldsymbol{\nu}}_* = \hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}, \hat{\zeta}_*^2 = \hat{\boldsymbol{\mu}}^\top\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}$.

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

$$\hat{\Theta} =_{\mathrm{df}} \frac{1}{n}\tilde{\mathsf{X}}^\top\tilde{\mathsf{X}} \tag{5}$$

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

2.1 Matrix Derivatives

Some notation and technical results concerning matrices are required.

Definition 2.1 (Matrix operations)**.** For matrix A, let $\mathrm{vec}\left(\mathsf{A}\right)$, and $\mathrm{vech}\left(\mathsf{A}\right)$ be the vector and half-space vector operators. The former turns an $p \times p$ matrix into an p^2 vector of its columns stacked on top of each other; the latter vectorizes a symmetric (or lower triangular) matrix into a vector of the non-redundant elements. Let L be the 'Elimination Matrix,' a matrix of zeros and ones with the property that $\mathrm{vech}\left(\mathsf{A}\right) = \mathsf{L}\,\mathrm{vec}\left(\mathsf{A}\right)$. The 'Duplication Matrix,' D, is the matrix of zeros and ones that reverses this operation: $\mathsf{D}\,\mathrm{vech}\left(\mathsf{A}\right) = \mathrm{vec}\left(\mathsf{A}\right)$. [24] Note that this implies that

$$\mathsf{L}\mathsf{D} = \mathsf{I}\left(\neq \mathsf{D}\mathsf{L}\right).$$

Let U_{-1} be the 'remove first' matrix, whose size should be inferred in context. It is a matrix of all rows but the first of the identity matrix. It exists to remove the first element of a vector.

Definition 2.2 (Derivatives)**.** For m-vector $\boldsymbol{x}$, and n-vector $\boldsymbol{y}$, let the derivative $\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}\boldsymbol{x}}$ be the $n \times m$ matrix whose first column is the partial derivative of $\boldsymbol{y}$ with respect to x_1. This follows the so-called 'numerator layout' convention. For matrices Y and X, define

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

Lemma 2.3 (Miscellaneous Derivatives)**.** *For symmetric matrices* Y *and* X*,*

$$\frac{\mathrm{d}\mathrm{vech}\left(\mathsf{Y}\right)}{\mathrm{d}\,\mathrm{vec}\left(\mathsf{X}\right)} = \mathsf{L}\frac{\mathrm{d}\mathsf{Y}}{\mathrm{d}\mathsf{X}}, \quad \frac{\mathrm{d}\mathrm{vec}\left(\mathsf{Y}\right)}{\mathrm{d}\,\mathrm{vech}\left(\mathsf{X}\right)} = \frac{\mathrm{d}\mathsf{Y}}{\mathrm{d}\mathsf{X}}\mathsf{D}, \quad \frac{\mathrm{d}\mathrm{vech}\left(\mathsf{Y}\right)}{\mathrm{d}\,\mathrm{vech}\left(\mathsf{X}\right)} = \mathsf{L}\frac{\mathrm{d}\mathsf{Y}}{\mathrm{d}\mathsf{X}}\mathsf{D}. \tag{6}$$

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

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

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

Lemma 2.4 (Derivative of matrix inverse)**.** *For invertible matrix* A*,*

$$\frac{\mathrm{d}\mathsf{A}^{-1}}{\mathrm{d}\mathsf{A}} = -\left(\mathsf{A}^{-\top} \otimes \mathsf{A}^{-1}\right) = -\left(\mathsf{A}^\top \otimes \mathsf{A}\right)^{-1}. \tag{7}$$

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

$$\frac{\mathrm{d}\mathrm{vech}\left(\mathsf{A}^{-1}\right)}{\mathrm{d}\,\mathrm{vech}\left(\mathsf{A}\right)} = -\mathsf{L}\left(\mathsf{A}^{-1} \otimes \mathsf{A}^{-1}\right)\mathsf{D}. \tag{8}$$

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

Proof. Equation 7 is a known result. [14, 25] Equation 8 then follows using Lemma 2.3. $\square$

2.2 Asymptotic distribution of the Markowitz portfolio

Collecting the mean and covariance into the second moment matrix gives the asymptotic distribution of the sample Markowitz portfolio without much work. In some sense, this computation generalizes the 'standard' asymptotic analysis of Sharpe ratio of multiple assets. [18, 23, 21, 22]

Theorem 2.5. *Let $\hat{\Theta}$ be the unbiased sample estimate of Θ, based on n i.i.d. samples of $\boldsymbol{x}$. Let Ω be the variance of $\operatorname{vech}\left(\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\top}\right)$. Then, asymptotically in n,*

$$\sqrt{n}\left(\operatorname{vech}\left(\hat{\Theta}^{-1}\right) - \operatorname{vech}\left(\Theta^{-1}\right)\right) \rightsquigarrow \mathcal{N}\left(0, \mathsf{H}\Omega\mathsf{H}^{\top}\right), \tag{9}$$

where

$$\mathsf{H} = -\mathsf{L}\left(\Theta^{-1} \otimes \Theta^{-1}\right)\mathsf{D}. \tag{10}$$

Furthermore, we may replace Ω in this equation with an asymptotically consistent estimator, $\hat{\Omega}$.

Proof. Under the multivariate central limit theorem [45]

$$\sqrt{n}\left(\operatorname{vech}\left(\hat{\Theta}\right) - \operatorname{vech}\left(\Theta\right)\right) \rightsquigarrow \mathcal{N}\left(0, \Omega\right), \tag{11}$$

where Ω is the variance of $\operatorname{vech}\left(\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\top}\right)$, which, in general, is unknown. By the delta method [45],

$$\sqrt{n}\left(\operatorname{vech}\left(\hat{\Theta}^{-1}\right) - \operatorname{vech}\left(\Theta^{-1}\right)\right) \rightsquigarrow \mathcal{N}\left(0, \left[\frac{\mathrm{d}\operatorname{vech}\left(\Theta^{-1}\right)}{\mathrm{d}\operatorname{vech}\left(\Theta\right)}\right] \Omega \left[\frac{\mathrm{d}\operatorname{vech}\left(\Theta^{-1}\right)}{\mathrm{d}\operatorname{vech}\left(\Theta\right)}\right]^{\top}\right).$$

The derivative is given by Lemma 2.4, and the result follows. $\square$

To estimate the covariance of $\operatorname{vech}\left(\hat{\Theta}^{-1}\right)$, plug in $\hat{\Theta}$ for Θ in the covariance computation, and use some consistent estimator for Ω, call it $\hat{\Omega}$. One way to compute $\hat{\Omega}$ is to via the sample covariance of the vectors $\operatorname{vech}\left(\tilde{\boldsymbol{x}}_i\tilde{\boldsymbol{x}}_i^{\top}\right) = \left[1, \boldsymbol{x}_i^{\top}, \operatorname{vech}\left(\boldsymbol{x}_i\boldsymbol{x}_i^{\top}\right)^{\top}\right]^{\top}$. More elaborate covariance estimators can be used, for example, to deal with violations of the *i.i.d.* assumptions. [47] Note that because the first element of $\operatorname{vech}\left(\tilde{\boldsymbol{x}}_i\tilde{\boldsymbol{x}}_i^{\top}\right)$ is a deterministic 1, the first row and column of Ω is all zeros, and we need not estimate it.

2.3 The Sharpe ratio optimal portfolio

Lemma 2.6 (Sharpe ratio optimal portfolio)**.** *Assuming $\boldsymbol{\mu} \neq \mathbf{0}$, and Σ is invertible, the portfolio optimization problem*

$$\underset{\boldsymbol{\nu}:\, \boldsymbol{\nu}^\top \Sigma \boldsymbol{\nu} \leq R^2}{\operatorname{argmax}} \frac{\boldsymbol{\nu}^\top \boldsymbol{\mu} - r_0}{\sqrt{\boldsymbol{\nu}^\top \Sigma \boldsymbol{\nu}}}, \tag{12}$$

for $r_0 \geq 0, R > 0$ is solved by

$$\boldsymbol{\nu}_{R,*} =_{\text{df}} \frac{R}{\sqrt{\boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu}}} \Sigma^{-1} \boldsymbol{\mu}. \tag{13}$$

Moreover, this is the unique solution whenever $r_0 > 0$. The maximal objective achieved by this portfolio is $\sqrt{\boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu}} - r_0/R$.

Proof. By the Lagrange multiplier technique, the optimal portfolio solves the following equations:

$$\begin{aligned} 0 &= c_1 \boldsymbol{\mu} - c_2 \Sigma \boldsymbol{\nu} - \gamma \Sigma \boldsymbol{\nu}, \\ \boldsymbol{\nu}^\top \Sigma \boldsymbol{\nu} &\leq R^2, \end{aligned}$$

where γ is the Lagrange multiplier, and c_1, c_2 are scalar constants. Solving the first equation gives us

$$\boldsymbol{\nu} = c\, \Sigma^{-1} \boldsymbol{\mu}.$$

This reduces the problem to the univariate optimization

$$\max_{c:\, c^2 \leq R^2/\zeta_*^2} \operatorname{sign}(c)\, \zeta_* - \frac{r_0}{|c|\, \zeta_*}, \tag{14}$$

where $\zeta_*^2 = \boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu}$. The optimum occurs for $c = R/\zeta_*$, moreover the optimum is unique when $r_0 > 0$. □

Note that the first element of $\operatorname{vech}\left(\Theta^{-1}\right)$ is $1 + \boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu}$, and elements 2 through $p+1$ are $-\boldsymbol{\nu}_*$. Thus, $\boldsymbol{\nu}_{R,*}$, the portfolio that maximizes the Sharpe ratio, is some transformation of $\operatorname{vech}\left(\Theta^{-1}\right)$, and another application of the delta method gives its asymptotic distribution, as in the following corollary to Theorem 2.5.

Corollary 2.7. *Let*

$$\boldsymbol{\nu}_{R,*} = \frac{R}{\sqrt{\boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu}}} \Sigma^{-1} \boldsymbol{\mu}, \tag{15}$$

and similarly, let $\hat{\boldsymbol{\nu}}_{R,*}$ be the sample analogue, where R is some risk budget. Then

$$\sqrt{n}\left(\hat{\boldsymbol{\nu}}_{R,*} - \boldsymbol{\nu}_{R,*}\right) \rightsquigarrow \mathcal{N}\left(0, \mathsf{H} \Omega \mathsf{H}^\top\right), \tag{16}$$

where

$$\begin{aligned} \mathsf{H} &= \left(- \left[\frac{1}{2\zeta_*^2} \boldsymbol{\nu}_{R,*}, \frac{R}{\zeta_*} \mathsf{I}_p, 0 \right] \right) \left(-\mathsf{L} \left(\Theta^{-1} \otimes \Theta^{-1} \right) \mathsf{D} \right), \\ \zeta_*^2 &=_{\text{df}} \boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu}. \end{aligned} \tag{17}$$

Proof. By the delta method, and Theorem 2.5, it suffices to show that

$$\frac{\mathrm{d}\boldsymbol{\nu}_{R,*}}{\mathrm{d}\,\mathrm{vech}\left(\Theta^{-1}\right)} = -\left[\frac{1}{2\zeta_*^2}\boldsymbol{\nu}_{R,*}, \frac{R}{\zeta_*}\mathsf{I}_p, 0\right].$$

To show this, note that $\boldsymbol{\nu}_{R,*}$ is $-R$ times elements 2 through $p+1$ of $\mathrm{vech}\left(\Theta^{-1}\right)$ divided by $\zeta_* = \sqrt{\boldsymbol{e}_1{}^{\top}\,\mathrm{vech}\left(\Theta^{-1}\right) - 1}$, where $\boldsymbol{e}_i$ is the i^{th} column of the identity matrix. The result follows from basic calculus. ☐

3 Distribution under Gaussian returns

The goal of this section is to derive a variant of Theorem 2.5 for the case where $\boldsymbol{x}$ follow a multivariate Gaussian distribution. First, assuming $\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \Sigma\right)$, we can express the density of $\boldsymbol{x}$, and of $\hat{\Theta}$, in terms of p, n, and Θ.

Lemma 3.1 (Gaussian sample density)**.** *Suppose* $\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \Sigma\right)$. *Letting* $\tilde{\boldsymbol{x}} = \left[1, \boldsymbol{x}^{\top}\right]^{\top}$, *and* $\Theta = \mathrm{E}\left[\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\top}\right]$, *then the negative log likelihood of* $\boldsymbol{x}$ *is*

$$-\log f_{\mathcal{N}}\left(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma\right) = c_p + \frac{1}{2}\log\left|\Theta\right| + \frac{1}{2}tr\left(\Theta^{-1}\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\top}\right), \tag{18}$$

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

Proof. By the block determinant formula,

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

Note also that

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

These relationships hold without assuming a particular distribution for $\boldsymbol{x}$.

The density of $\boldsymbol{x}$ is then

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

and the result follows. ☐

Lemma 3.2 (Gaussian second moment matrix density)**.** *Let* $\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \Sigma\right)$, $\tilde{\boldsymbol{x}} = \left[1, \boldsymbol{x}^{\top}\right]^{\top}$, *and* $\Theta = \mathrm{E}\left[\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\top}\right]$. *Given* n i.i.d. *samples* $\boldsymbol{x}_i$, *let Let* $\hat{\Theta} = \frac{1}{n}\sum_i \tilde{\boldsymbol{x}}_i\tilde{\boldsymbol{x}}_i{}^{\top}$. *Then the density of* $\hat{\Theta}$ *is*

$$f\left(\hat{\Theta}; \Theta\right) = \exp\left(c'_{n,p}\right)\frac{\left|\hat{\Theta}\right|^{\frac{n-p-2}{2}}}{\left|\Theta\right|^{\frac{n}{2}}}\exp\left(-\frac{n}{2}tr\left(\Theta^{-1}\hat{\Theta}\right)\right), \tag{19}$$

for some $c'_{n,p}$.

Proof. Let $\tilde{\mathsf{X}}$ be the matrix whose rows are the vectors $\boldsymbol{x}_i{}^\top$. From Lemma 3.1, and using linearity of the trace, the negative log density of $\tilde{\mathsf{X}}$ is

$$
\begin{aligned}
-\log f_{\mathcal{N}}\left(\tilde{\mathsf{X}};\Theta\right) &= nc_p + \frac{n}{2}\log|\Theta| + \frac{1}{2}\mathrm{tr}\left(\Theta^{-1}\tilde{\mathsf{X}}^\top\tilde{\mathsf{X}}\right),\\
\therefore\ \frac{-2\log f_{\mathcal{N}}\left(\tilde{\mathsf{X}};\Theta\right)}{n} &= 2c_p + \log|\Theta| + \mathrm{tr}\left(\Theta^{-1}\hat{\Theta}\right).
\end{aligned}
$$

By Lemma (5.1.1) of Press [40], this can be expressed as a density on $\hat{\Theta}$:

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

where $c'_{n,p}$ is the term in brackets on the third line. Factoring out $-2/n$ and taking an exponent gives the result. ☐

Corollary 3.3. *The random variable $n\hat{\Theta}$ has the same density, up to a constant in p and n, as a $p+1$-dimensional Wishart random variable with n degrees of freedom and scale matrix Θ. Thus $n\hat{\Theta}$ is a* conditional *Wishart, conditional on $\hat{\Theta}_{1,1}=1$. [40, 1]*

Corollary 3.4. *The derivatives of log likelihood are given by*

$$
\begin{aligned}
\frac{\mathrm{dlog}\, f\left(\hat{\Theta};\Theta\right)}{\mathrm{d\,vec}\left(\Theta\right)} &= -\frac{n}{2}\left[\mathrm{vec}\left(\Theta^{-1}-\Theta^{-1}\hat{\Theta}\Theta^{-1}\right)\right]^\top,\\
\frac{\mathrm{dlog}\, f\left(\hat{\Theta};\Theta\right)}{\mathrm{d\,vec}\left(\Theta^{-1}\right)} &= -\frac{n}{2}\left[\mathrm{vec}\left(\Theta-\hat{\Theta}\right)\right]^\top.
\end{aligned}
\tag{20}
$$

Proof. Plugging in the log likelihood gives

$$
\frac{\mathrm{dlog}\, f\left(\hat{\Theta};\Theta\right)}{\mathrm{d\,vec}\left(\Theta\right)} = -\frac{n}{2}\left[\frac{\mathrm{dlog}\,|\Theta|}{\mathrm{d\,vec}\left(\Theta\right)} + \frac{\mathrm{dtr}\left(\Theta^{-1}\hat{\Theta}\right)}{\mathrm{d\,vec}\left(\Theta\right)}\right],
$$

and then standard matrix calculus gives the first result. [25, 39] Proceeding similarly gives the second. ☐

This immediately gives us the Maximum Likelihood Estimator.

Corollary 3.5 (MLE)**.** *$\hat{\Theta}$ is the maximum likelihood estimator of Θ.*

To compute the covariance of $\operatorname{vech}(\Theta)$, Ω, in the Gaussian case, one can compute the Fisher Information, then appeal to the fact that Θ is the MLE. However, because the first element of $\operatorname{vech}(\Theta)$ is a deterministic 1, the first row and column of Ω are all zeros. This is an unfortunate wrinkle. The solution is to compute the Fisher Information with respect to the nonredundant variables, $\mathsf{U}_{-1}\operatorname{vech}(\Theta)$, as follows.

Lemma 3.6 (Fisher Information)**.** *The Fisher Information of $\mathsf{U}_{-1}\operatorname{vech}(\Theta)$ is*

$$\mathcal{I}_n\left(\mathsf{U}_{-1}\operatorname{vech}(\Theta)\right) = \frac{n}{2}\mathsf{U}_{-1}\left[\mathsf{L}\left(\Theta^{-1}\otimes\Theta^{-1}\right)\mathsf{D}\right]^{\top}\mathsf{D}^{\top}\left(\Theta\otimes\Theta\right)\mathsf{D}\left[\mathsf{L}\left(\Theta^{-1}\otimes\Theta^{-1}\right)\mathsf{D}\right]\mathsf{U}_{-1}{}^{\top}. \quad (21)$$

Proof. First compute the Hessian of $\log f\left(\hat{\Theta};\Theta\right)$ with respect to $\operatorname{vec}\left(\Theta^{-1}\right)$. The Hessian is defined as

$$\frac{\mathrm{d}^2\log f\left(\hat{\Theta};\Theta\right)}{\mathrm{d}\left(\operatorname{vec}\left(\Theta^{-1}\right)\right)^2} =_{\mathrm{df}} \frac{\mathrm{d}\left[\frac{\mathrm{d}\log f\left(\hat{\Theta};\Theta\right)}{\mathrm{d}\operatorname{vec}\left(\Theta^{-1}\right)}\right]^{\top}}{\mathrm{d}\operatorname{vec}\left(\Theta^{-1}\right)}.$$

Then, from Equation 20,

$$\begin{aligned}\frac{\mathrm{d}^2\log f\left(\hat{\Theta};\Theta\right)}{\mathrm{d}\left(\operatorname{vec}\left(\Theta^{-1}\right)\right)^2} &= -\frac{n}{2}\frac{\mathrm{d}\left[\Theta-\hat{\Theta}\right]}{\mathrm{d}\operatorname{vec}\left(\Theta^{-1}\right)},\\ &= -\frac{n}{2}\left(\Theta\otimes\Theta\right),\end{aligned}$$

via Lemma 2.4. Perform a change of variables. Via Lemma 2.3,

$$\frac{\mathrm{d}^2\log f\left(\hat{\Theta};\Theta\right)}{\mathrm{d}\left(\operatorname{vech}\left(\Theta^{-1}\right)\right)^2} = -\frac{n}{2}\mathsf{D}^{\top}\left(\Theta\otimes\Theta\right)\mathsf{D}.$$

Using Lemma 2.4, perform another change of variables to find

$$\frac{\mathrm{d}^2\log f\left(\hat{\Theta};\Theta\right)}{\mathrm{d}\left(\operatorname{vech}\left(\Theta\right)\right)^2} = -\frac{n}{2}\left[\mathsf{L}\left(\Theta^{-1}\otimes\Theta^{-1}\right)\mathsf{D}\right]^{\top}\mathsf{D}^{\top}\left(\Theta\otimes\Theta\right)\mathsf{D}\left[\mathsf{L}\left(\Theta^{-1}\otimes\Theta^{-1}\right)\mathsf{D}\right].$$

Finally, perform the change of variables to get the Hessian with respect to $\mathsf{U}_{-1}\operatorname{vech}(\Theta)$. Since the Fisher Information is negative the expected value of this Hessian, the result follows. [37] □

Thus the analogue of Theorem 2.5 for Gaussian returns is given by the following theorem.

Theorem 3.7. *Let $\hat{\Theta}$ be the unbiased sample estimate of Θ, based on n* i.i.d. *samples of $\boldsymbol{x}$, assumed multivariate Gaussian. Then, asymptotically in n,*

$$\sqrt{n}\left(\operatorname{vech}\left(\hat{\Theta}\right)-\operatorname{vech}\left(\Theta\right)\right)\rightsquigarrow\mathcal{N}\left(0,\Omega\right), \quad (22)$$

where the first row and column of Ω *are all zero, and the lower right block part is*

$$2\left[\mathsf{U}_{-1}\left[\mathsf{L}(\Theta^{-1}\otimes\Theta^{-1})\mathsf{D}\right]^{\top}\mathsf{D}^{\top}\left(\Theta\otimes\Theta\right)\mathsf{D}\left[\mathsf{L}(\Theta^{-1}\otimes\Theta^{-1})\mathsf{D}\right]{\mathsf{U}_{-1}}^{\top}\right]^{-1}.$$

Proof. Under 'the appropriate regularity conditions,' [45, 37]

$$\left(\mathsf{U}_{-1}\,\mathrm{vech}\left(\hat{\Theta}\right)-\mathsf{U}_{-1}\,\mathrm{vech}\left(\Theta\right)\right)\rightsquigarrow\mathcal{N}\left(0,\left[\mathcal{I}_{n}\left(\mathsf{U}_{-1}\,\mathrm{vech}\left(\Theta\right)\right)\right]^{-1}\right),\tag{23}$$

and the result follows from Lemma 3.6, and the fact that the first elements of both $\mathrm{vech}\left(\hat{\Theta}\right)$ and $\mathrm{vech}\left(\Theta\right)$ are a deterministic 1. □

The 'plug-in' estimator of the covariance substitutes in $\hat{\Theta}$ for Θ in the right hand side of Equation 22. The following conjecture is true in the $p=1$ case. Use of the Sherman-Morrison-Woodbury formula might aid in a proof.

Conjecture 3.8. For the Gaussian case, asymptotically in n,

$$\sqrt{n}\left(\mathrm{vech}\left(\hat{\Theta}^{-1}\right)-\mathrm{vech}\left(\Theta^{-1}\right)\right)\rightsquigarrow\mathcal{N}\left(0,2\left[\mathsf{D}^{\top}\left(\Theta\otimes\Theta\right)\mathsf{D}\right]^{-1}-2\boldsymbol{e}_{1}{\boldsymbol{e}_{1}}^{\top}\right).\tag{24}$$

A check of Theorem 3.7 and an illustration of Conjecture 3.8 are given in the appendix.

3.1 Likelihood ratio test on Markowitz portfolio

Consider the null hypothesis

$$H_0:\mathrm{tr}\left(\mathsf{A}_i\Theta^{-1}\right)=a_i,\;i=1,\ldots,m.\tag{25}$$

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

Employing the Lagrange multiplier technique, the maximum likelihood estimator under the null hypothesis, call it Θ_0, solves the following equation

$$\begin{aligned}0&=\frac{\mathrm{d}\log f\left(\hat{\Theta};\Theta\right)}{\mathrm{d}\Theta^{-1}}-\sum_i\lambda_i\frac{\mathrm{d}\mathrm{tr}\left(\mathsf{A}_i\Theta^{-1}\right)}{\mathrm{d}\Theta^{-1}},\\&=-\Theta_0+\hat{\Theta}-\sum_i\lambda_i\mathsf{A}_i,.\end{aligned}$$

Thus the MLE under the null is

$$\Theta_0=\hat{\Theta}-\sum_i\lambda_i\mathsf{A}_i.\tag{26}$$

The maximum likelihood estimator under the constraints has to be found numerically by solving for the λ_i, subject to the constraints in Equation 25.

This framework slightly generalizes Dempster's "Covariance Selection," [12] which reduces to the case where each a_i is zero, and each A_i is a matrix of all zeros except two (symmetric) ones somewhere in the lower right $p \times p$ sub matrix. In all other respects, however, the solution here follows Dempster.

An iterative technique for finding the MLE based on a Newton step would proceed as follow. [34] Let $\boldsymbol{\lambda}^{(0)}$ be some initial estimate of the vector of λ_i. (A good initial estimate can likely be had by abusing the asymptotic normality result from Section 2.2.) The residual of the k^{th} estimate, $\boldsymbol{\lambda}^{(k)}$ is

$$\boldsymbol{\epsilon}_i^{(k)} =_{\text{df}} \operatorname{tr}\left(\mathsf{A}_i \left[\hat{\Theta} - \sum_j \lambda_j^{(k)} \mathsf{A}_j\right]^{-1}\right) - a_i. \tag{27}$$

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

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{\epsilon}_i^{(k)}}{\mathrm{d}\lambda_l^{(k)}} &= \operatorname{tr}\left(\mathsf{A}_i \left[\hat{\Theta} - \sum_j \lambda_j^{(k)} \mathsf{A}_j\right]^{-1} \mathsf{A}_l \left[\hat{\Theta} - \sum_j \lambda_j^{(k)} \mathsf{A}_j\right]^{-1}\right),\\ &= \operatorname{vec}\left(\mathsf{A}_i\right)^{\top}\left(\left[\hat{\Theta} - \sum_j \lambda_j^{(k)} \mathsf{A}_j\right]^{-1} \otimes \left[\hat{\Theta} - \sum_j \lambda_j^{(k)} \mathsf{A}_j\right]^{-1}\right) \operatorname{vec}\left(\mathsf{A}_l\right).\end{aligned} \tag{28}$$

Newton's method is then the iterative scheme

$$\boldsymbol{\lambda}^{(k+1)} \leftarrow \boldsymbol{\lambda}^{(k)} - \left(\frac{\mathrm{d}\boldsymbol{\epsilon}^{(k)}}{\mathrm{d}\boldsymbol{\lambda}^{(k)}}\right)^{-1} \boldsymbol{\epsilon}^{(k)}_{.} \tag{29}$$

When (if?) the iterative scheme converges on the optimum, plugging in $\boldsymbol{\lambda}^{(k)}$ into Equation 26 gives the MLE under the null. The likelihood ratio test statistic is

$$\begin{aligned}-2\log\Lambda &=_{\text{df}} -2\log\left(\frac{f\left(\Theta_0 \middle| \hat{\Theta}\right)}{f\left(\Theta_{\text{unrestricted MLE}} \middle| \hat{\Theta}\right)}\right),\\ &= n\left(\log\left|\Theta_0\hat{\Theta}^{-1}\right| + \operatorname{tr}\left(\left[{\Theta_0}^{-1} - \hat{\Theta}^{-1}\right]\hat{\Theta}\right)\right),\\ &= n\left(\log\left|\Theta_0\hat{\Theta}^{-1}\right| + \operatorname{tr}\left({\Theta_0}^{-1}\hat{\Theta}\right) - [p+1]\right),\end{aligned} \tag{30}$$

using the fact that $\hat{\Theta}$ is the unrestricted MLE, per Corollary 3.5. By Wilks' Theorem, under the null hypothesis, $-2\log\Lambda$ is, asymptotically in n, distributed as a chi-square with m degrees of freedom. [46]

4 Extensions

For large samples, Wald statistics of the elements of the Markowitz portfolio computed using the procedure outlined above tend to be very similar to the t-statistics produced by the procedure of Britten-Jones. [5] However, the technique proposed here admits a number of interesting extensions.

The script for each of these extensions is the same: define, then solve, some portfolio optimization problem; show that the solution can be defined in terms of some transformation of Θ^{-1}, giving an implicit recipe for constructing the sample portfolio based on the same transformation of $\hat{\Theta}^{-1}$; find the asymptotic distribution of the sample portfolio in terms of Ω.

4.1 Subspace Constraint

Consider the *constrained* portfolio optimization problem

$$\max_{\substack{\boldsymbol{\nu}:\mathsf{J}^{\perp}\boldsymbol{\nu}=\mathbf{0},\\ \boldsymbol{\nu}^{\top}\Sigma\boldsymbol{\nu}\le R^2}} \frac{\boldsymbol{\nu}^{\top}\boldsymbol{\mu}-r_0}{\sqrt{\boldsymbol{\nu}^{\top}\Sigma\boldsymbol{\nu}}}, \tag{31}$$

where $\mathsf{J}^{\perp}$ is a $(p-p_j)\times p$ matrix of rank $p-p_j$, r_0 is the disastrous rate, and $R>0$ is the risk budget. 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 interpret the orthogonality constraint $\mathsf{J}^{\perp}\boldsymbol{\nu}=\mathbf{0}$ as stating that $\boldsymbol{\nu}$ must be a linear combination of the columns of $\mathsf{J}^{\top}$, thus $\boldsymbol{\nu}=\mathsf{J}^{\top}\boldsymbol{\xi}$. The columns of $\mathsf{J}^{\top}$ may be considered 'baskets' of assets to which our investments are restricted.

We can rewrite the portfolio optimization problem in terms of solving for $\boldsymbol{\xi}$, but then find the asymptotic distribution of the resultant $\boldsymbol{\nu}$. Note that the expected return and covariance of the portfolio $\boldsymbol{\xi}$ are, respectively, $\boldsymbol{\xi}^{\top}\mathsf{J}\boldsymbol{\mu}$ and $\boldsymbol{\xi}^{\top}\mathsf{J}\Sigma\mathsf{J}^{\top}\boldsymbol{\xi}$. Thus we can plug in $\mathsf{J}\boldsymbol{\mu}$ and $\mathsf{J}\Sigma\mathsf{J}^{\top}$ into Lemma 2.6 to get the following analogue.

Lemma 4.1 (subspace constrained Sharpe ratio optimal portfolio)**.** *Assuming the rows of* J *span the null space of the rows of* $\mathsf{J}^{\perp}$*,* $\mathsf{J}\boldsymbol{\mu}\neq\mathbf{0}$*, and* Σ *is invertible, the portfolio optimization problem*

$$\max_{\substack{\boldsymbol{\nu}:\mathsf{J}^{\perp}\boldsymbol{\nu}=\mathbf{0},\\ \boldsymbol{\nu}^{\top}\Sigma\boldsymbol{\nu}\le R^2}} \frac{\boldsymbol{\nu}^{\top}\boldsymbol{\mu}-r_0}{\sqrt{\boldsymbol{\nu}^{\top}\Sigma\boldsymbol{\nu}}}, \tag{32}$$

for $r_0\ge 0, R>0$ *is solved by*

$$\begin{aligned}\boldsymbol{\nu}_{R,\mathsf{J},*} &=_{\mathrm{df}} c\mathsf{J}^{\top}\left(\mathsf{J}\Sigma\mathsf{J}^{\top}\right)^{-1}\mathsf{J}\boldsymbol{\mu},\\ c &= \frac{R}{\sqrt{\boldsymbol{\mu}^{\top}\mathsf{J}^{\top}\left(\mathsf{J}\Sigma\mathsf{J}^{\top}\right)^{-1}\mathsf{J}\boldsymbol{\mu}}}.\end{aligned}$$

When $r_0>0$ *the solution is unique.*

We can easily find the asymptotic distribution of $\hat{\boldsymbol{\nu}}_{R,\mathsf{J},*}$, the sample analogue of the optimal portfolio in Lemma 4.1. First define the subspace second moment.

Definition 4.2. Let $\tilde{\mathsf{J}}$ be the $(1+p_j)\times(p+1)$ matrix,

$$\tilde{\mathsf{J}} =_{\mathrm{df}} \left[\begin{array}{cc} 1 & 0\\ 0 & \mathsf{J}\end{array}\right].$$

Simple algebra proves the following lemma.

Lemma 4.3. *The elements of* $\tilde{\mathsf{J}}^\top \left(\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top\right)^{-1}\tilde{\mathsf{J}}$ *are*

$$\tilde{\mathsf{J}}^\top \left(\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top\right)^{-1}\tilde{\mathsf{J}} = \left[\begin{array}{cc} 1 + \boldsymbol{\mu}^\top\mathsf{J}^\top\left(\mathsf{J}\Sigma\mathsf{J}^\top\right)^{-1}\mathsf{J}\boldsymbol{\mu} & -\boldsymbol{\mu}^\top\mathsf{J}^\top\left(\mathsf{J}\Sigma\mathsf{J}^\top\right)^{-1}\mathsf{J} \\ -\mathsf{J}^\top\left(\mathsf{J}\Sigma\mathsf{J}^\top\right)^{-1}\mathsf{J}\boldsymbol{\mu} & \mathsf{J}^\top\left(\mathsf{J}\Sigma\mathsf{J}^\top\right)^{-1}\mathsf{J} \end{array}\right].$$

In particular, elements 2 through $p+1$ of $-\operatorname{vech}\left(\tilde{\mathsf{J}}^\top \left(\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top\right)^{-1}\tilde{\mathsf{J}}\right)$ *are the portfolio* $\hat{\boldsymbol{\nu}}_{R,\mathsf{J},*}$ *defined in Lemma 4.1, up to the scaling constant c which is the ratio of R to the square root of the first element of* $\operatorname{vech}\left(\tilde{\mathsf{J}}^\top \left(\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top\right)^{-1}\tilde{\mathsf{J}}\right)$ *minus one.*

The asymptotic distribution of $\operatorname{vech}\left(\tilde{\mathsf{J}}^\top \left(\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top\right)^{-1}\tilde{\mathsf{J}}\right)$ is given by the following theorem, which is the analogue of Theorem 2.5.

Theorem 4.4. *Let $\hat{\Theta}$ be the unbiased sample estimate of Θ, based on n* i.i.d. *samples of $\boldsymbol{x}$. Let $\tilde{\mathsf{J}}$ be defined as in Definition 4.2. Let Ω be the variance of* $\operatorname{vech}\left(\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^\top\right)$. *Then, asymptotically in n,*

$$\sqrt{n}\left(\operatorname{vech}\left(\tilde{\mathsf{J}}^\top \left(\tilde{\mathsf{J}}\hat{\Theta}\tilde{\mathsf{J}}^\top\right)^{-1}\tilde{\mathsf{J}}\right) - \operatorname{vech}\left(\tilde{\mathsf{J}}^\top \left(\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top\right)^{-1}\tilde{\mathsf{J}}\right)\right) \rightsquigarrow \mathcal{N}\left(0, \mathsf{H}\Omega\mathsf{H}^\top\right), \tag{33}$$

where

$$\mathsf{H} = -\mathsf{L}\left(\tilde{\mathsf{J}}^\top \otimes \tilde{\mathsf{J}}^\top\right)\left(\left(\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top\right)^{-1} \otimes \left(\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top\right)^{-1}\right)\left(\tilde{\mathsf{J}} \otimes \tilde{\mathsf{J}}\right)\mathsf{D}.$$

Proof. By the multivariate delta method, it suffices to prove that

$$\mathsf{H} = \frac{\mathrm{d}\operatorname{vech}\left(\tilde{\mathsf{J}}^\top \left(\tilde{\mathsf{J}}\hat{\Theta}\tilde{\mathsf{J}}^\top\right)^{-1}\tilde{\mathsf{J}}\right)}{\mathrm{d}\operatorname{vech}\left(\Theta\right)}.$$

Via Lemma 2.3, it suffices to prove that

$$\frac{\mathrm{d}\tilde{\mathsf{J}}^\top \left(\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top\right)^{-1}\tilde{\mathsf{J}}}{\mathrm{d}\Theta} = -\left(\tilde{\mathsf{J}}^\top \otimes \tilde{\mathsf{J}}^\top\right)\left(\left(\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top\right)^{-1} \otimes \left(\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top\right)^{-1}\right)\left(\tilde{\mathsf{J}} \otimes \tilde{\mathsf{J}}\right).$$

A well-known fact regarding matrix manipulation [25] is

$$\operatorname{vec}\left(\mathsf{ABC}\right) = \left(\mathsf{A} \otimes \mathsf{C}^\top\right)\operatorname{vec}\left(\mathsf{B}\right), \quad \text{therefore,} \quad \frac{\mathrm{d}\mathsf{ABC}}{\mathrm{d}\mathsf{B}} = \mathsf{A} \otimes \mathsf{C}^\top.$$

Using this, and the chain rule, we have:

$$\begin{aligned} \frac{\mathrm{d}\tilde{\mathsf{J}}^\top \left(\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top\right)^{-1}\tilde{\mathsf{J}}}{\mathrm{d}\Theta} &= \frac{\mathrm{d}\tilde{\mathsf{J}}^\top \left(\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top\right)^{-1}\tilde{\mathsf{J}}}{\mathrm{d}\left(\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top\right)^{-1}} \frac{\mathrm{d}\left(\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top\right)^{-1}}{\mathrm{d}\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top} \frac{\mathrm{d}\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top}{\mathrm{d}\Theta} \\ &= \left(\tilde{\mathsf{J}}^\top \otimes \tilde{\mathsf{J}}^\top\right) \frac{\mathrm{d}\left(\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top\right)^{-1}}{\mathrm{d}\tilde{\mathsf{J}}\Theta\tilde{\mathsf{J}}^\top} \left(\tilde{\mathsf{J}} \otimes \tilde{\mathsf{J}}\right). \end{aligned}$$

Lemma 2.4 gives the middle term, completing the proof. ∎

An analogue of Corollary 2.7 gives the asymptotic distribution of $\boldsymbol{\nu}_{R,\mathsf{J},*}$ defined in Lemma 4.1.

4.2 Hedging Constraint

Consider, now, the constrained portfolio optimization problem,

$$\max_{\substack{\boldsymbol{\nu}:\mathsf{G}\Sigma\boldsymbol{\nu}=\mathbf{0},\\ \boldsymbol{\nu}^{\top}\Sigma\boldsymbol{\nu}\le R^2}} \frac{\boldsymbol{\nu}^{\top}\boldsymbol{\mu}-r_0}{\sqrt{\boldsymbol{\nu}^{\top}\Sigma\boldsymbol{\nu}}}, \tag{34}$$

where G is now a $p_g \times p$ matrix of rank p_g. We can interpret the G constraint as stating that the covariance of the returns of a feasible portfolio with the returns of a portfolio whose weights are in a given row of G shall equal zero. In the garden variety application of this problem, G consists of p_g rows of the identity matrix; in this case, feasible portfolios are 'hedged' with respect to the p_g assets selected by G (although they may hold some position in the hedged assets).

Lemma 4.5 (constrained Sharpe ratio optimal portfolio)**.** *Assuming $\boldsymbol{\mu} \ne \mathbf{0}$, and Σ is invertible, the portfolio optimization problem*

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

for $r_0 \ge 0, R > 0$ is solved by

$$\begin{aligned} \boldsymbol{\nu}_{R,\mathsf{G},*} &=_{\mathrm{df}} c\left(\Sigma^{-1}\boldsymbol{\mu} - \mathsf{G}^{\top}\left(\mathsf{G}\Sigma\mathsf{G}^{\top}\right)^{-1}\mathsf{G}\boldsymbol{\mu}\right),\\ c &= \frac{R}{\sqrt{\boldsymbol{\mu}^{\top}\Sigma^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^{\top}\mathsf{G}^{\top}\left(\mathsf{G}\Sigma\mathsf{G}^{\top}\right)^{-1}\mathsf{G}\boldsymbol{\mu}}}. \end{aligned}$$

When $r_0 > 0$ the solution is unique.

Proof. By the Lagrange multiplier technique, the optimal portfolio solves the following equations:

$$\begin{aligned} 0 &= c_1\boldsymbol{\mu} - c_2\Sigma\boldsymbol{\nu} - \gamma_1\Sigma\boldsymbol{\nu} - \Sigma\mathsf{G}^{\top}\boldsymbol{\gamma_2},\\ \boldsymbol{\nu}^{\top}\Sigma\boldsymbol{\nu} &\le R^2,\\ \mathsf{G}\Sigma\boldsymbol{\nu} &= \mathbf{0}, \end{aligned}$$

where γ_i are Lagrange multipliers, and c_1, c_2 are scalar constants.

Solving the first equation gives

$$\boldsymbol{\nu} = c_3\left[\Sigma^{-1}\boldsymbol{\mu} - \mathsf{G}^{\top}\boldsymbol{\gamma_2}\right].$$

Reconciling this with the hedging equation we have

$$\mathbf{0} = \mathsf{G}\Sigma\boldsymbol{\nu} = c_3\mathsf{G}\Sigma\left[\Sigma^{-1}\boldsymbol{\mu} - \mathsf{G}^{\top}\boldsymbol{\gamma_2}\right],$$

and therefore $\boldsymbol{\gamma_2} = \left(\mathsf{G}\Sigma\mathsf{G}^{\top}\right)^{-1}\mathsf{G}\boldsymbol{\mu}$. Thus

$$\boldsymbol{\nu} = c_3\left[\Sigma^{-1}\boldsymbol{\mu} - \mathsf{G}^{\top}\left(\mathsf{G}\Sigma\mathsf{G}^{\top}\right)^{-1}\mathsf{G}\boldsymbol{\mu}\right].$$

Plugging this into the objective reduces the problem to the univariate optimization

$$\max_{c_3 \colon c_3^2 \le R^2/\zeta_{*,\mathsf{G}}^2} \operatorname{sign}\left(c_3\right) \zeta_{*,\mathsf{G}} - \frac{r_0}{\left|c_3\right| \zeta_{*,\mathsf{G}}},$$

where $\zeta_{*,\mathsf{G}}^2 = \boldsymbol{\mu}^\top\Sigma^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^\top\mathsf{G}^\top\left(\mathsf{G}\Sigma\mathsf{G}^\top\right)^{-1}\mathsf{G}\boldsymbol{\mu}$. The optimum occurs for $c = R/\zeta_{*,\mathsf{G}}$, moreover the optimum is unique when $r_0 > 0$. $\square$

The optimal hedged portfolio in Lemma 4.5 is, up to scaling, the difference of the unconstrained optimal portfolio from Lemma 2.6 and the subspace constrained portfolio in Lemma 4.1. This 'delta' analogy continues for the rest of this section.

Definition 4.6 (Delta Inverse Second Moment)**.** Let $\tilde{\mathsf{G}}$ be the $(1 + p_g) \times (p + 1)$ matrix,

$$\tilde{\mathsf{G}} =_{\mathrm{df}} \left[\begin{array}{cc} 1 & 0 \\ 0 & \mathsf{G} \end{array}\right].$$

Define the 'delta inverse second moment' as

$$\Delta_{\mathsf{G}}\Theta^{-1} =_{\mathrm{df}} \Theta^{-1} - \tilde{\mathsf{G}}^\top\left(\tilde{\mathsf{G}}\Theta\tilde{\mathsf{G}}^\top\right)^{-1}\tilde{\mathsf{G}}.$$

Simple algebra proves the following lemma.

Lemma 4.7. *The elements of* $\Delta_{\mathsf{G}}\Theta^{-1}$ *are*

$$\Delta_{\mathsf{G}}\Theta^{-1} = \left[\begin{array}{cc} \boldsymbol{\mu}^\top\Sigma^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^\top\mathsf{G}^\top\left(\mathsf{G}\Sigma\mathsf{G}^\top\right)^{-1}\mathsf{G}\boldsymbol{\mu} & -\boldsymbol{\mu}^\top\Sigma^{-1} + \boldsymbol{\mu}^\top\mathsf{G}^\top\left(\mathsf{G}\Sigma\mathsf{G}^\top\right)^{-1}\mathsf{G} \\ -\Sigma^{-1}\boldsymbol{\mu} + \mathsf{G}^\top\left(\mathsf{G}\Sigma\mathsf{G}^\top\right)^{-1}\mathsf{G}\boldsymbol{\mu} & \Sigma^{-1} - \mathsf{G}^\top\left(\mathsf{G}\Sigma\mathsf{G}^\top\right)^{-1}\mathsf{G} \end{array}\right].$$

In particular, elements 2 through $p + 1$ of $-\operatorname{vech}\left(\Delta_{\mathsf{G}}\Theta^{-1}\right)$ *are the portfolio* $\boldsymbol{\nu}_{R,\mathsf{G},*}$ *defined in Lemma 4.5, up to the scaling constant c which is the ratio of R to the square root of the first element of* $\operatorname{vech}\left(\Delta_{\mathsf{G}}\Theta^{-1}\right)$*.*

The statistic $\hat{\boldsymbol{\mu}}^\top\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}^\top\mathsf{G}^\top\left(\mathsf{G}\hat{\Sigma}\mathsf{G}^\top\right)^{-1}\mathsf{G}\hat{\boldsymbol{\mu}}$, for the case where G is some rows of the $p \times p$ identity matrix, was first proposed by Rao, and its distribution under Gaussian returns was later found by Giri. [41, 15] This test statistic may be used for tests of portfolio *spanning* for the case where a risk-free instrument is traded. [17, 19]

The asymptotic distribution of $\Delta_{\mathsf{G}}\hat{\Theta}^{-1}$ is given by the following theorem, which is the analogue of Theorem 2.5.

Theorem 4.8. *Let $\hat{\Theta}$ be the unbiased sample estimate of Θ, based on n* i.i.d. *samples of $\boldsymbol{x}$. Let $\Delta_{\mathsf{G}}\Theta^{-1}$ be defined as in Definition 4.6, and similarly define $\Delta_{\mathsf{G}}\hat{\Theta}^{-1}$. Let Ω be the variance of* $\operatorname{vech}\left(\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^\top\right)$*. Then, asymptotically in n,*

$$\sqrt{n}\left(\operatorname{vech}\left(\Delta_{\mathsf{G}}\hat{\Theta}^{-1}\right) - \operatorname{vech}\left(\Delta_{\mathsf{G}}\Theta^{-1}\right)\right) \rightsquigarrow \mathcal{N}\left(0, \mathsf{H}\Omega\mathsf{H}^\top\right), \tag{36}$$

where

$$\mathsf{H} = -\mathsf{L}\left[\Theta^{-1} \otimes \Theta^{-1} - \left(\tilde{\mathsf{G}}^\top \otimes \tilde{\mathsf{G}}^\top\right)\left(\left(\tilde{\mathsf{G}}\Theta\tilde{\mathsf{G}}^\top\right)^{-1} \otimes \left(\tilde{\mathsf{G}}\Theta\tilde{\mathsf{G}}^\top\right)^{-1}\right)\left(\tilde{\mathsf{G}} \otimes \tilde{\mathsf{G}}\right)\right]\mathsf{D}.$$

Proof. Minor modification of proof of Theorem 4.4. □

Caution. In the hedged portfolio optimization problem considered here, the optimal portfolio will, in general, hold money in the row space of G. For example, in the garden variety application, where one is hedging out exposure to 'the market' by including a broad market ETF, and taking G to be the corresponding row of the identity matrix, the final portfolio may hold some position in that broad market ETF. This is fine for an ETF, but one may wish to hedge out exposure to an untradeable returns stream–the returns of an index, say. Combining the hedging constraint of this section with the subspace constraint of Section 4.1 is simple in the case where the rows of G are spanned by the rows of J. The more general case, however, is rather more complicated.

4.3 Conditional Heteroskedasticity

The methods described above ignore 'volatility clustering', and assume homoskedasticity. [9, 33, 3] To deal with this, consider a strictly positive scalar random variable, q_i, observable at the time the investment decision is required to capture $\boldsymbol{x}_{i+1}$. For reasons to be obvious later, it is more convenient to think of q_i as a 'quietude' indicator.

Two simple competing models for conditional heteroskedasticity are

$$\text{(constant):} \quad \mathrm{E}\left[\boldsymbol{x}_{i+1}\,|\,q_i\right] = {q_i}^{-1}\boldsymbol{\mu} \qquad \mathrm{Var}\left(\boldsymbol{x}_{i+1}\,|\,q_i\right) = {q_i}^{-2}\Sigma, \tag{37}$$

$$\text{(floating):} \quad \mathrm{E}\left[\boldsymbol{x}_{i+1}\,|\,q_i\right] = \boldsymbol{\mu} \qquad \mathrm{Var}\left(\boldsymbol{x}_{i+1}\,|\,q_i\right) = {q_i}^{-2}\Sigma. \tag{38}$$

Under the model in Equation 37, the maximal Sharpe ratio is $\sqrt{\boldsymbol{\mu}^{\top}\Sigma^{-1}\boldsymbol{\mu}}$, independent of q_i; under Equation 38, it is is $q_i\sqrt{\boldsymbol{\mu}^{\top}\Sigma^{-1}\boldsymbol{\mu}}$. The model names reflect whether or not the maximal Sharpe ratio varies conditional on q_i.

The optimal portfolio under both models is the same, as stated in the following lemma, the proof of which follows by simply using Lemma 2.6.

Lemma 4.9 (Conditional Sharpe ratio optimal portfolio)**.** *Under either the model in Equation 37 or Equation 38, conditional on observing q_i, the portfolio optimization problem*

$$\operatorname*{argmax}_{\boldsymbol{\nu}:\,\mathrm{Var}\left(\boldsymbol{\nu}^{\top}\boldsymbol{x}_{i+1}\,|\,q_i\right)\le R^2} \frac{\mathrm{E}\left[\boldsymbol{\nu}^{\top}\boldsymbol{x}_{i+1}\,|\,q_i\right] - r_0}{\sqrt{\mathrm{Var}\left(\boldsymbol{\nu}^{\top}\boldsymbol{x}_{i+1}\,|\,q_i\right)}}, \tag{39}$$

for $r_0 \ge 0, R > 0$ is solved by

$$\boldsymbol{\nu}_{*} = \frac{q_i R}{\sqrt{\boldsymbol{\mu}^{\top}\Sigma^{-1}\boldsymbol{\mu}}}\Sigma^{-1}\boldsymbol{\mu}. \tag{40}$$

Moreover, this is the unique solution whenever $r_0 > 0$.

To perform inference on the portfolio $\boldsymbol{\nu}_{*}$ from Lemma 4.9, under the 'constant' model of Equation 37, apply the unconditional techniques to the sample second moment of $q_i\tilde{\boldsymbol{x}}_{i+1}$.

For the 'floating' model of Equation 38, however, some adjustment to the technique is required. Define $\tilde{\tilde{\boldsymbol{x}}}_{i+1} =_{\mathrm{df}} q_i\tilde{\boldsymbol{x}}_{i+1}$; that is, $\tilde{\tilde{\boldsymbol{x}}}_{i+1} = \left[q_i, q_i{\boldsymbol{x}_{i+1}}^{\top}\right]^{\top}$. Consider the second moment of $\tilde{\tilde{\boldsymbol{x}}}$:

$$\Theta_q =_{\mathrm{df}} \mathrm{E}\left[\tilde{\tilde{\boldsymbol{x}}}\tilde{\tilde{\boldsymbol{x}}}^{\top}\right] = \left[\begin{array}{cc} \gamma^2 & \gamma^2\boldsymbol{\mu}^{\top} \\ \gamma^2\boldsymbol{\mu} & \Sigma + \boldsymbol{\mu}\gamma^2\boldsymbol{\mu}^{\top} \end{array}\right], \quad \text{where} \quad \gamma^2 =_{\mathrm{df}} \mathrm{E}\left[q^2\right]. \tag{41}$$

The inverse of Θ_q is

$$\Theta_q{}^{-1} = \left[\begin{array}{cc} \gamma^{-2} + \boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu} & -\boldsymbol{\mu}^\top \Sigma^{-1} \\ -\Sigma^{-1}\boldsymbol{\mu} & \Sigma^{-1} \end{array} \right] \tag{42}$$

Once again, the optimal portfolio (up to scaling and sign), appears in $\operatorname{vech}\left(\Theta_q{}^{-1}\right)$. Similarly, define the sample analogue:

$$\hat{\Theta}_q =_{\mathrm{df}} \frac{1}{n} \sum_i \tilde{\tilde{\boldsymbol{x}}}_{i+1} \tilde{\tilde{\boldsymbol{x}}}_{i+1}{}^\top. \tag{43}$$

We can find the asymptotic distribution of $\operatorname{vech}\left(\hat{\Theta}_q\right)$ using the same techniques as in the unconditional case, as in the following analogue of Theorem 2.5:

Theorem 4.10. *Let* $\hat{\Theta}_q =_{\mathrm{df}} \frac{1}{n}\sum_i \tilde{\tilde{\boldsymbol{x}}}_{i+1} \tilde{\tilde{\boldsymbol{x}}}_{i+1}{}^\top$, *based on* n i.i.d. *samples of* $\left[q, \boldsymbol{x}^\top\right]^\top$. *Let* Ω *be the variance of* $\operatorname{vech}\left(\tilde{\tilde{\boldsymbol{x}}}\tilde{\tilde{\boldsymbol{x}}}^\top\right)$. *Then, asymptotically in* n,

$$\sqrt{n}\left(\operatorname{vech}\left(\hat{\Theta}_q^{-1}\right) - \operatorname{vech}\left(\Theta_q{}^{-1}\right)\right) \rightsquigarrow \mathcal{N}\left(0, \mathsf{H}\Omega\mathsf{H}^\top\right), \tag{44}$$

where

$$\mathsf{H} = -\mathsf{L}\left(\Theta_q{}^{-1} \otimes \Theta_q{}^{-1}\right)\mathsf{D}. \tag{45}$$

Furthermore, we may replace Ω *in this equation with an asymptotically consistent estimator,* $\hat{\Omega}$.

The only real difference from the unconditional case is that we cannot automatically assume that the first row and column of Ω is zero (unless q is actually constant, which misses the point). Moreover, the shortcut for estimating Ω under Gaussian returns is not valid without some patching, an exercise left for the reader.

Dependence or independence of maximal Sharpe ratio from volatility is an assumption which, ideally, one could test with data. A mixed model containing both characteristics can be written as follows:

$$\text{(mixed):} \quad \mathrm{E}\left[\boldsymbol{x}_{i+1} \,|\, q_i\right] = q_i{}^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 \qquad \mathrm{Var}\left(\boldsymbol{x}_{i+1} \,|\, q_i\right) = q_i{}^{-2}\Sigma. \tag{46}$$

One could then test whether elements of $\boldsymbol{\mu}_0$ or of $\boldsymbol{\mu}_1$ are zero. Analyzing this model is somewhat complicated without moving to a more general framework, as in the sequel.

4.4 Conditional Expectation and Heteroskedasticity

Suppose you observe random variables $q_i > 0$, and f-vector $\boldsymbol{f}_i$ at some time prior to when the investment decision is required to capture $\boldsymbol{x}_{i+1}$. It need not be the case that q and $\boldsymbol{f}$ are independent. The general model is now

$$\text{(bi-conditional):} \quad \mathrm{E}\left[\boldsymbol{x}_{i+1} \,|\, q_i, \boldsymbol{f}_i\right] = \mathsf{B}\boldsymbol{f}_i \quad \mathrm{Var}\left(\boldsymbol{x}_{i+1} \,|\, q_i, \boldsymbol{f}_i\right) = q_i{}^{-2}\Sigma, \tag{47}$$

where B is some $p \times f$ matrix. Without the q_i term, these are the 'predictive regression' equations commonly used in Tactical Asset Allocation. [8, 16, 4]

By letting $\boldsymbol{f}_i = \left[q_i^{-1}, 1\right]^\top$ we recover the mixed model in Equation 46; the bi-conditional model is considerably more general, however. The conditionally-optimal portfolio is given by the following lemma. Once again, the proof proceeds simply by plugging in the conditional expected return and volatility into Lemma 2.6.

Lemma 4.11 (Conditional Sharpe ratio optimal portfolio)**.** *Under the model in Equation 47, conditional on observing q_i and $\boldsymbol{f}_i$, the portfolio optimization problem*

$$\underset{\boldsymbol{\nu}:\,\mathrm{Var}\left(\boldsymbol{\nu}^\top \boldsymbol{x}_{i+1}\,|\,q_i,\boldsymbol{f}_i\right)\le R^2}{\operatorname{argmax}} \frac{\mathrm{E}\left[\boldsymbol{\nu}^\top \boldsymbol{x}_{i+1}\,|\,q_i,\boldsymbol{f}_i\right] - r_0}{\sqrt{\mathrm{Var}\left(\boldsymbol{\nu}^\top \boldsymbol{x}_{i+1}\,|\,q_i,\boldsymbol{f}_i\right)}}, \tag{48}$$

for $r_0 \ge 0, R > 0$ is solved by

$$\boldsymbol{\nu}_* = \frac{q_i R}{\sqrt{\boldsymbol{f}_i^\top \mathsf{B}^\top \Sigma^{-1} \mathsf{B} \boldsymbol{f}_i}} \Sigma^{-1} \mathsf{B} \boldsymbol{f}_i. \tag{49}$$

Moreover, this is the unique solution whenever $r_0 > 0$.

Caution. It is emphatically *not* the case that investing in the portfolio $\boldsymbol{\nu}_*$ from Lemma 4.11 at every time step is long-term Sharpe ratio optimal. One may possibly achieve a higher long-term Sharpe ratio by down-levering at times when the conditional Sharpe ratio is low. The optimal long term investment strategy falls under the rubric of 'multiperiod portfolio choice', and is an area of active research. [32, 13, 4]

The matrix $\Sigma^{-1}\mathsf{B}$ is the generalization of the Markowitz portfolio: it is the multiplier for a model under which the optimal portfolio is linear in the features $\boldsymbol{f}_i$ (up to scaling to satisfy the risk budget). We can think of this matrix as the 'Markowitz coefficient'. If an entire column of $\Sigma^{-1}\mathsf{B}$ is zero, it suggests that the corresponding element of $\boldsymbol{f}$ can be ignored in investment decisions; if an entire row of $\Sigma^{-1}\mathsf{B}$ is zero, it suggests the corresponding instrument delivers no return or hedging benefit.

Tests on $\Sigma^{-1}\mathsf{B}$ should be contrasted with the so-called Multivariate General Linear Hypothesis (MGLH), which tests the matrix equation $\mathsf{ABC} = \mathsf{T}$, for conformable $\mathsf{A}, \mathsf{C}, \mathsf{T}$. [42, 31]

To perform inference on the Markowitz coefficient, we can proceed exactly as above. Let

$$\tilde{\tilde{\boldsymbol{x}}}_{i+1} =_{\mathrm{df}} \left[q_i \boldsymbol{f}_i^\top, q_i \boldsymbol{x}_{i+1}^\top\right]^\top. \tag{50}$$

Consider the second moment of $\tilde{\tilde{\boldsymbol{x}}}$:

$$\Theta_f =_{\mathrm{df}} \mathrm{E}\left[\tilde{\tilde{\boldsymbol{x}}}\tilde{\tilde{\boldsymbol{x}}}^\top\right] = \begin{bmatrix} \Gamma_f & \Gamma_f \mathsf{B}^\top \\ \mathsf{B}\Gamma_f & \Sigma + \mathsf{B}\Gamma_f \mathsf{B}^\top \end{bmatrix}, \quad \text{where} \quad \Gamma_f =_{\mathrm{df}} \mathrm{E}\left[q^2 \boldsymbol{f}\boldsymbol{f}^\top\right]. \tag{51}$$

The inverse of Θ_f is

$$\Theta_f^{-1} = \begin{bmatrix} \Gamma_f^{-1} + \mathsf{B}^\top \Sigma^{-1} \mathsf{B} & -\mathsf{B}^\top \Sigma^{-1} \\ -\Sigma^{-1}\mathsf{B} & \Sigma^{-1} \end{bmatrix} \tag{52}$$

Once again, the Markowitz coefficient (up to scaling and sign), appears in $\operatorname{vech}\left(\Theta_f^{-1}\right)$.

The following theorem is an analogue of, and shares a proof with, Theorem 2.5.

Theorem 4.12. *Let* $\hat{\Theta}_f =_{\mathrm{df}} \frac{1}{n}\sum_i \tilde{\tilde{\boldsymbol{x}}}_{i+1}{\tilde{\tilde{\boldsymbol{x}}}_{i+1}}^\top$*, based on* n i.i.d. *samples of* $\left[q, \boldsymbol{f}^\top, \boldsymbol{x}^\top\right]^\top$*, where*

$$\tilde{\tilde{\boldsymbol{x}}}_{i+1} =_{\mathrm{df}} \left[q_i {\boldsymbol{f}_i}^\top, q_i {\boldsymbol{x}_{i+1}}^\top\right]^\top.$$

Let Ω *be the variance of* $\mathrm{vech}\left(\tilde{\tilde{\boldsymbol{x}}}\tilde{\tilde{\boldsymbol{x}}}^\top\right)$*. Then, asymptotically in* n*,*

$$\sqrt{n}\left(\mathrm{vech}\left(\hat{\Theta}_f^{-1}\right) - \mathrm{vech}\left({\Theta_f}^{-1}\right)\right) \rightsquigarrow \mathcal{N}\left(0, \mathsf{H}\Omega\mathsf{H}^\top\right), \tag{53}$$

where

$$\mathsf{H} = -\mathsf{L}\left({\Theta_f}^{-1} \otimes {\Theta_f}^{-1}\right)\mathsf{D}. \tag{54}$$

Furthermore, we may replace Ω *in this equation with an asymptotically consistent estimator,* $\hat{\Omega}$*.*

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] Tim Bollerslev. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69(3):pp. 542–547, 1987. ISSN 00346535. URL http://www.jstor.org/stable/1925546.

[4] Michael W Brandt. Portfolio choice problems. *Handbook of financial econometrics*, 1:269–336, 2009. URL http://shr.receptidocs.ru/docs/5/4748/conv_1/file1.pdf#page=298.

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

[6] 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.

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

[8] 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.

[9] 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.

[10] 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://docs.edhec-risk.com/mrk/120503_Princeton/Research_papers/DeMiguel-Garlappi-Uppal-RFS-2009-OptimalVersusNaiveDiversification.pdf.

[11] Victor DeMiguel, Alberto Martin-Utrera, and Francisco J Nogales. Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking & Finance*, 2013. URL http://faculty.london.edu/avmiguel/DMN-2011-07-21.pdf.

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

[13] F.J. Fabozzi, P.N. Kolm, D. Pachamanova, and S.M. Focardi. *Robust Portfolio Optimization and Management*. Frank J. Fabozzi series. Wiley, 2007. ISBN 9780470164891. URL http://books.google.com/books?id=PUnRxEBIFb4C.

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

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

[16] 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.

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

[18] 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.

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

[20] Takuya Kinkawa. Estimation of optimal portfolio weights using shrinkage technique. 2010. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1576052.

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

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

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

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

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

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

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

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

[29] Robert C. Merton. On estimating the expected return on the market: An exploratory investigation. Working Paper 444, National Bureau of Economic Research, February 1980. URL http://www.nber.org/papers/w0444.

[30] 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.

[31] 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.

[32] John M Mulvey, William R Pauling, and Ronald E Madey. Advantages of multiperiod portfolio models. *The Journal of Portfolio Management*, 29(2):35–45, 2003. doi: 10.3905/jpm.2003.319871. URL http://dx.doi.org/10.3905/jpm.2003.319871#sthash.oKQ9cHFy.jsYuZ7C2.dpuf.

[33] Daniel B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):pp. 347–370, 1991. ISSN 00129682. URL http://www.samsi.info/sites/default/files/Nelson_1991.pdf.

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

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

[36] Steven E. Pav. Scalar Gaussian example via Sympy. Privately Published, 2013. URL http://nbviewer.ipython.org/gist/anonymous/8116771.

[37] Yudi Pawitan. *In all likelihood: statistical modelling and inference using likelihood*. Oxford science publications. Clarendon press, Oxford, 2001. ISBN 978-0-19-850765-9. URL http://books.google.com/books?id=8T8fAQAAQBAJ.

[38] Fernando Pérez and Brian E. Granger. IPython: a System for Interactive Scientific Computing. *Comput. Sci. Eng.*, 9(3):21–29, May 2007. URL http://ipython.org.

[39] Kaare Brandt Petersen and Michael Syskind Pedersen. The matrix cookbook, nov 2012. URL http://www2.imm.dtu.dk/pubdb/p.php?3274. Version 20121115.

[40] 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.

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

[42] 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.

[43] M.R. Spiegel and L.J. Stephens. *Schaum's Outline of Statistics*. Schaum's Outline Series. Mcgraw-hill, 2007. ISBN 9780071594462. URL http://books.google.com/books?id=qdcBmgs3N3AC.

[44] SymPy Development Team. *SymPy: Python library for symbolic mathematics*, 2011. URL http://www.sympy.org.

[45] 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.

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

[47] Achim Zeileis. Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, 11(10):1–17, 11 2004. ISSN 1548-7660. URL http://www.jstatsoft.org/v11/i10.

A Confirming the scalar Gaussian case

Example A.1. To sanity check Theorem 3.7, consider the $p = 1$ Gaussian case. In this case,

$$\operatorname{vech}\left(\Theta\right) = \left[1, \mu, \sigma^2 + \mu^2\right]^\top, \quad \text{and} \quad \operatorname{vech}\left(\Theta^{-1}\right) = \left[1 + \frac{\mu^2}{\sigma^2}, -\frac{\mu}{\sigma^2}, \frac{1}{\sigma^2}\right]^\top.$$

Let $\hat{\mu}, \hat{\sigma}^2$ be the unbiased sample estimates. By well known results [43], $\hat{\mu}$ and $\hat{\sigma}^2$ are independent, and have asymptotic variances of σ^2/n and $2\sigma^4/n$ respectively. By the delta method, the asymptotic variance of $\mathsf{U}_{-1} \operatorname{vech}\left(\hat{\Theta}\right)$ and $\operatorname{vech}\left(\hat{\Theta}^{-1}\right)$ can be computed as

$$\begin{aligned}
\operatorname{Var}\left(\mathsf{U}_{-1} \operatorname{vech}\left(\hat{\Theta}\right)\right) &\rightsquigarrow \frac{1}{n}\left[\begin{array}{cc} 1 & 2\mu \\ 0 & 1 \end{array}\right]^\top \left[\begin{array}{cc} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{array}\right] \left[\begin{array}{cc} 1 & 2\mu \\ 0 & 1 \end{array}\right], \\
&= \frac{1}{n}\left[\begin{array}{cc} \sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 4\mu^2\sigma^2 + 2\sigma^4 \end{array}\right].
\end{aligned} \tag{55}$$

$$\begin{aligned}
\operatorname{Var}\left(\operatorname{vech}\left(\hat{\Theta}^{-1}\right)\right) &\rightsquigarrow \frac{1}{n}\left[\begin{array}{ccc} \frac{2\mu}{\sigma^2} & -\frac{1}{\sigma^2} & 0 \\ -\frac{\mu^2}{\sigma^4} & \frac{\mu}{\sigma^4} & -\frac{1}{\sigma^4} \end{array}\right]^\top \left[\begin{array}{cc} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{array}\right] \left[\begin{array}{ccc} \frac{2\mu}{\sigma^2} & -\frac{1}{\sigma^2} & 0 \\ -\frac{\mu^2}{\sigma^4} & \frac{\mu}{\sigma^4} & -\frac{1}{\sigma^4} \end{array}\right], \\
&= \frac{1}{n}\left[\begin{array}{ccc} 2\zeta & -\frac{1}{\sigma} & 0 \\ -\sqrt{2}\zeta^2 & \sqrt{2}\frac{\zeta}{\sigma} & -\frac{\sqrt{2}}{\sigma^2} \end{array}\right]^\top \left[\begin{array}{ccc} 2\zeta & -\frac{1}{\sigma} & 0 \\ -\sqrt{2}\zeta^2 & \sqrt{2}\frac{\zeta}{\sigma} & -\frac{\sqrt{2}}{\sigma^2} \end{array}\right], \\
&= \frac{1}{n}\left[\begin{array}{ccc} 2\zeta^2\left(2 + \zeta^2\right) & -\frac{2\zeta}{\sigma}\left(1 + \zeta^2\right) & 2\frac{\zeta^2}{\sigma^2} \\ -\frac{2\zeta}{\sigma}\left(1 + \zeta^2\right) & \frac{1 + 2\zeta^2}{\sigma^2} & -\frac{2\zeta}{\sigma^3} \\ 2\frac{\zeta^2}{\sigma^2} & -\frac{2\zeta}{\sigma^3} & \frac{2}{\sigma^4} \end{array}\right].
\end{aligned} \tag{56}$$

Now it remains to compute $\operatorname{Var}\left(\mathsf{U}_{-1} \operatorname{vech}\left(\hat{\Theta}\right)\right)$ via Theorem 3.7, and then $\operatorname{Var}\left(\operatorname{vech}\left(\hat{\Theta}^{-1}\right)\right)$ via Theorem 2.5, and confirm they match the values above. This is a rather tedious computation best left to a computer. Below is an excerpt of an iPython notebook using Sympy [38, 44] which performs this computation. This notebook is available online. [36]

In [1]:
```python
# confirm the asymptotic distribution of Theta
# for scalar Gaussian case.
from __future__ import division
from sympy import *
from sympy.physics.quantum import TensorProduct
init_printing(use_unicode=False, wrap_line=False, \
              no_global=True)
mu = symbols('\mu')
sg = symbols('\sigma')
# the elimination, duplication and U_{-1} matrices:
Elim = Matrix(3,4,[1,0,0,0,  0,1,0,0,   0,0,0,1])
Dupp = Matrix(4,3,[1,0,0, 0,1,0, 0,1,0, 0,0,1])
Unun = Matrix(2,3,[0,1,0,  0,0,1])
def Qform(A,x):
    """compute the quadratic form x'Ax"""
    return x.transpose() * A * x
```

In [2]:
```python
Theta = Matrix(2,2,[1,mu,mu,mu**2 + sg**2])
Theta
```

Out[2]:

$$\begin{bmatrix} 1 & \mu \\ \mu & \mu^2 + \sigma^2 \end{bmatrix}$$

In [3]:
```python
# compute tensor products and
# the derivative d vech(Theta^-1) / d vech(Theta)
# see also Theorem 2.5
Theta_Theta = TensorProduct(Theta,Theta)
iTheta_iTheta = TensorProduct(Theta.inv(),Theta.inv())
theta_i_deriv = Elim * (iTheta_iTheta) * Dupp
```

In [4]:
```python
# towards Theorem 3.7
DTTD = Qform(Theta_Theta,Dupp)
D_DTTD_D = Qform(DTTD,theta_i_deriv)
iOmega = Qform(D_DTTD_D,Unun.transpose())
Omega = 2 * iOmega.inv()
simplify(Omega)
```

Out[4]:

$$\begin{bmatrix} \sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 2\sigma^2\left(2\mu^2 + \sigma^2\right) \end{bmatrix}$$

In [5]:
```python
# this matches the computation in Equation 55
# on to the inverse:
# actually use Theorem 2.5
theta_i_deriv_t = theta_i_deriv.transpose()
theta_inv_var = Qform(Qform(Omega,Unun),theta_i_deriv_t)
simplify(theta_inv_var)
```

Out[5]:

$$\begin{bmatrix}\frac{2\mu^2}{\sigma^4}\left(\mu^2+2\sigma^2\right) & -\frac{2\mu}{\sigma^4}\left(\mu^2+\sigma^2\right) & \frac{2\mu^2}{\sigma^4}\\ -\frac{2\mu}{\sigma^4}\left(\mu^2+\sigma^2\right) & \frac{1}{\sigma^4}\left(2\mu^2+\sigma^2\right) & -\frac{2\mu}{\sigma^4}\\ \frac{2\mu^2}{\sigma^4} & -\frac{2\mu}{\sigma^4} & \frac{2}{\sigma^4}\end{bmatrix}$$

In [6]:
```
# this matches the computation in Equation 56
# now check Conjecture 3.8
conjec = Qform(Theta_Theta,Dupp)
e1 = Matrix(3,1,[1,0,0])
convar = 2 * (conjec.inv() - e1 * e1.transpose())
simplify(convar)
```

Out[6]:

$$\begin{bmatrix}\frac{2\mu^2}{\sigma^4}\left(\mu^2+2\sigma^2\right) & -\frac{2\mu}{\sigma^4}\left(\mu^2+\sigma^2\right) & \frac{2\mu^2}{\sigma^4}\\ -\frac{2\mu}{\sigma^4}\left(\mu^2+\sigma^2\right) & \frac{1}{\sigma^4}\left(2\mu^2+\sigma^2\right) & -\frac{2\mu}{\sigma^4}\\ \frac{2\mu^2}{\sigma^4} & -\frac{2\mu}{\sigma^4} & \frac{2}{\sigma^4}\end{bmatrix}$$

In [7]:
```
# are they the same?
simplify(theta_inv_var - convar)
```

Out[7]:

$$\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$$