Portfolio inference and portfolio overfit

Steven E. Pav
steven@gilgamath.com

Formerly of Cerebellum Capital

Sep 10, 2015

Portfolio Basics I

- Consider the case of p assets which can be held long or short.
- A portfolio, $\boldsymbol{\nu}$, is a p-vector of dollar-wise allocations.
- Let $\mathbf{x}_i$ be the p-vector of simple returns from $i-1$ to i.
- If you held $\boldsymbol{\nu}$ at time $i-1$, your portfolio's return is $\boldsymbol{\nu}^{\top}\mathbf{x}_i$. (*n.b.*, this does not hold for log returns.)
- If

$$\mathrm{E}\left[\mathbf{x}\right] = \boldsymbol{\mu}, \quad \mathrm{Var}\left(\mathbf{x}\right) = \Sigma,$$

then

$$\mathrm{E}\left[\boldsymbol{\nu}^{\top}\mathbf{x}\right] = \boldsymbol{\nu}^{\top}\boldsymbol{\mu}, \quad \mathrm{Var}\left(\boldsymbol{\nu}^{\top}\mathbf{x}\right) = \boldsymbol{\nu}^{\top}\Sigma\boldsymbol{\nu}.$$

Portfolio Basics II

- The signal-noise ratio of a portfolio $\boldsymbol{\nu}$ is

$$\zeta\left(\boldsymbol{\nu}\right) := \frac{\operatorname{E}\left[\boldsymbol{\nu}^{\top}\mathbf{x}\right] - r_0}{\sqrt{\operatorname{Var}\left(\boldsymbol{\nu}^{\top}\mathbf{x}\right)}} = \frac{\boldsymbol{\nu}^{\top}\boldsymbol{\mu} - r_0}{\sqrt{\boldsymbol{\nu}^{\top}\Sigma\boldsymbol{\nu}}}.$$

  (We will often ignore the 'risk-free' rate, $r_0 = 0$.)
- Like the Sharpe ratio, but uses population parameters.
- The Sharpe ratio is the sample mean of $\boldsymbol{\nu}^{\top}\mathbf{x}_1, \boldsymbol{\nu}^{\top}\mathbf{x}_2, \ldots, \boldsymbol{\nu}^{\top}\mathbf{x}_n$ divided by the sample standard deviation.
- For large n, the Sharpe ratio converges to the signal-noise ratio. Barring bad luck, if you could maximize signal-noise ratio, you would increase Sharpe ratio. (But $\boldsymbol{\mu}$ and Σ are unknown!)
- For a small fund, a high achieved Sharpe ratio in early trading can make a big difference!

Maximizing signal-noise ratio I

- The Markowitz portfolio maximizes the signal-noise ratio:

$$\boldsymbol{\nu}_* := \Sigma^{-1}\boldsymbol{\mu}$$

- The signal-noise ratio of the Markowitz portfolio is

$$\zeta_* := \sqrt{\boldsymbol{\mu}^\top\Sigma^{-1}\boldsymbol{\mu}}.$$

This portfolio, up to rescaling, solves many portfolio problems:

- "Maximize expected return subject to a cap on volatility."
- "Minimize volatility subject to a minimum expected return."
- "Maximize signal-noise ratio with a risk-free rate:"

$$\boldsymbol{\nu}_* \propto \boxed{\operatorname*{argmax}_{\boldsymbol{\nu}:\,\boldsymbol{\nu}^\top\Sigma\boldsymbol{\nu}\le R^2} \frac{\boldsymbol{\nu}^\top\boldsymbol{\mu} - r_0}{\sqrt{\boldsymbol{\nu}^\top\Sigma\boldsymbol{\nu}}}},$$

A weird trick

- Prepend a '1' to the vector: $\tilde{\mathbf{x}} := \left[1, \mathbf{x}^\top\right]^\top$.
- The second moment of $\tilde{\mathbf{x}}$ contains the first two moments of $\mathbf{x}$:

$$\Theta := \operatorname{E}\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\top\right] = \left[\begin{array}{cc} 1 & \boldsymbol{\mu}^\top \\ \boldsymbol{\mu} & \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^\top \end{array}\right].$$

$$\text{then:}\quad \Theta^{-1} = \left[\begin{array}{cc} 1 + \boldsymbol{\mu}^\top\Sigma^{-1}\boldsymbol{\mu} & -\boldsymbol{\mu}^\top\Sigma^{-1} \\ -\Sigma^{-1}\boldsymbol{\mu} & \Sigma^{-1} \end{array}\right]$$

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

- $\boldsymbol{\nu}_*$ is the Markowitz portfolio,

  ζ_* is the signal-noise ratio of $\boldsymbol{\nu}_*$.

  Σ^{-1} is the 'precision matrix'.

Sample estimator

- Since $\Theta = \mathrm{E}\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^{\top}\right]$ the simple estimator is unbiased:

$$\hat{\Theta} := \frac{1}{n}\sum_{1\le i\le n}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^{\top} = \left[\begin{array}{cc} 1 & \hat{\boldsymbol{\mu}}^{\top} \\ \hat{\boldsymbol{\mu}} & \hat{\Sigma} + \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^{\top} \end{array}\right].$$

- The inverse contains the sample estimates:

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

Asymptotics I

- By the Central Limit Theorem:

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

  where $\Omega := \operatorname{Var}\left(\operatorname{vech}\left(\tilde{\mathbf{x}}\tilde{\mathbf{x}}^{\top}\right)\right)$.
  We can estimate Ω from the sample, call it $\hat{\Omega}$:
  It's just sample covariance of $\operatorname{vech}\left(\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^{\top}\right)$, for $1 \le i \le n$.
- Use the delta method:

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

  Here U is some 'ugly' derivative, depending on Θ.

Asymptotics II

- Ignoring details about symmetry, *etc.*, the derivative is: [7, 14]

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

(This generalizes the scalar derivative!)

I can make a hat or a brooch or a pterodactyl...

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

What is the use for $\mathrm{Var}\left(\mathrm{vech}\left(\hat{\Theta}^{-1}\right)\right)$?

- Perform inference on elements of $\boldsymbol{\nu}_*$ via Wald statistic. (Compare elements of $\boldsymbol{\nu}_*$ to their standard errors.)
- Perform inference on the optimal signal-noise ratio, ζ_*.
- Equivalently, Hotelling's T^2 test. (tests hypothesis: $\boldsymbol{\mu}$ is all zeros)
- Portfolio shrinkage.
- Estimate the covariance of $\hat{\boldsymbol{\nu}}_*$ and $\hat{\Sigma}^{-1}$. (Attribute portfolio error to returns or covariance.) [5]

Implementation: trust but verify

```
require(MarkowitzR)
set.seed(2014)
X <- matrix(rnorm(1000 * 5), ncol = 5)  # toy data
ism <- MarkowitzR::mp_vcov(X)
walds <- function(ism) ism$W/sqrt(diag(ism$What))
print(t(walds(ism)))  # Wald stats

##              X1    X2   X3   X4    X5
## Intercept 0.89 -0.22 -1.6 -2.4 -0.49

# c.f. Britten-Jones, http://jstor.org/stable/2697722
y <- rep(1, dim(X)[1])
print(t(summary(lm(y ~ X - 1))$coefficients[, 3]))

##          X1    X2   X3   X4    X5
## [1,] 0.89 -0.22 -1.6 -2.5 -0.48
```

Selling this weird trick

Why weird trick, not Britten-Jones, or Okhrin *et al.*? [4, 2, 15]

- Fewer assumptions: fourth moments exist vs. normality of returns.
- Straightforward to use HAC estimator for Ω.
- Models covariance between return and volatility. (At a cost?)
- Solves a larger problem, *e.g.*, can use for inference on ζ_*^2.

Selling this weird trick

Why weird trick, not Britten-Jones, or Okhrin *et al.*? [4, 2, 15]

- Fewer assumptions: fourth moments exist vs. normality of returns.
- Straightforward to use HAC estimator for Ω.
- Models covariance between return and volatility. (At a cost?)
- Solves a larger problem, *e.g.*, can use for inference on ζ_*^2.

Real question: what's wrong with vanilla Markowitz?

This trick can be adapted to deal with:

- Hedged portfolios.
- Heteroskedasticity.
- Conditional expected returns.
- Perhaps more ...

Hedged portfolios I

Hedging: the goal

Returns which are statistically *independent* from some random variables.

Hedging: a more realistic goal

A portfolio with zero *covariance* to some random variables.

Hedging: an achievable goal

A portfolio with zero *sample* covariance to some other portfolios of *tradeable assets*.
(*e.g.*, you may have to hold some `Mkt` to hedge out the `Mkt`.)

Hedged portfolios II

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

where G is a $p_g \times p$ matrix of rank p_g.

- Rows of G define portfolios against which we have 0 covariance.
- Typically G consists of some rows of identity matrix.

i.e., "Maximize signal-noise ratio with risk bound and zero covariance to some other portfolios."

Solved by $c\nu_{\mathsf{G},*}$, with c to satisfy risk bound, and

$$\nu_{\mathsf{G},*} := \left(\Sigma^{-1}\mu - \mathsf{G}^{\top}\left(\mathsf{G}\Sigma\mathsf{G}^{\top}\right)^{-1}\mathsf{G}\mu\right).$$

Hedged portfolios III

- Use the weird trick! Let $\tilde{G} := \begin{bmatrix} 1 & 0 \\ 0 & G \end{bmatrix}$, then,

$$\Theta^{-1} - \tilde{G}^\top \left(\tilde{G}\Theta\tilde{G}^\top\right)^{-1} \tilde{G} =$$

$$\begin{bmatrix} \mu^\top\Sigma^{-1}\mu - \mu^\top G^\top \left(G\Sigma G^\top\right)^{-1} G\mu & -{\nu_{G,*}}^\top \\ -\nu_{G,*} & \Sigma^{-1} - G^\top \left(G\Sigma G^\top\right)^{-1} G \end{bmatrix}.$$

$-\nu_{G,*}$ is the optimal hedged portfolio.

UL corner is squared signal-noise ratio of $\nu_{G,*}$.

Also used for portfolio spanning. [19, 9, 12, 13]

LR corner is loss of precision?

Hedged portfolios IV

- Delta method gives the asymptotic distribution:

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

with more ugly derivatives.

Hedged portfolios V

- Download the Fama-French 3 factor + Momentum monthly data (1927-02-01 to 2015-01-01) from *Quandl*. [20]
- Add risk-free rate back to market, compute (unhedged) Markowitz portfolio, and Wald statistics.

```
w.stats <- rbind(do.both(ff4.xts[, 1:4]), wtrick.ws(ff4.xts[,
    1:4], vcov.func = sandwich::vcovHAC))
rownames(w.stats)[3] <- c("weird trick w/ HAC")
xtable(w.stats)
```

	Mkt	SMB	HML	UMD
Britten Jones t-stat	6.31	0.60	5.01	8.18
weird trick Wald stat	5.40	0.63	4.48	6.04
weird trick w/ HAC	5.10	0.64	3.94	5.57

Hedged portfolios VI

Now hedge out Mkt:

```
walds <- function(ism) ism$W/sqrt(diag(ism$What))
Gmat <- matrix(diag(1, 4)[1, ], ncol = 4)
asymv <- MarkowitzR::mp_vcov(ff4.xts[, 1:4], fit.intercept = TRUE,
    Gmat = Gmat)
xtable(t(walds(asymv)))
```

	Mkt	SMB	HML	UMD
Intercept	2.68	0.63	4.48	6.04

And compute the spanning Wald statistic:

```
ef.stat <- function(ism) ism$mu[1]/sqrt(ism$Ohat[1, 1])
print(ef.stat(asymv))

## [1] 3.8
```

Hedged portfolios VII

Now hedge out Mkt and RF:

```
# hedge out RFR too
Gmat <- matrix(diag(1, 5)[c(1, 5), ], ncol = 5)
asymv <- MarkowitzR::mp_vcov(ff4.xts[, 1:5], fit.intercept = TRUE,
    Gmat = Gmat)
xtable(t(walds(asymv)))
```

	Mkt	SMB	HML	UMD	RF
Intercept	0.78	1.86	2.31	3.47	-1.49

And the spanning statistic:

```
print(ef.stat(asymv))

## [1] 2.1
```

Heteroskedasticity

- Prior to investment decision, observe s_i proportional to volatility.
- Two competing 'obvious' models:

$$\begin{aligned} &\text{(constant):} \quad \mathrm{E}\left[\mathbf{x}_{i+1} \,|\, s_i\right] = s_i \boldsymbol{\mu} \qquad \mathrm{Var}\left(\mathbf{x}_{i+1} \,|\, s_i\right) = {s_i}^2 \Sigma, \\ &\text{(floating):} \quad \mathrm{E}\left[\mathbf{x}_{i+1} \,|\, s_i\right] = \boldsymbol{\mu} \qquad \mathrm{Var}\left(\mathbf{x}_{i+1} \,|\, s_i\right) = {s_i}^2 \Sigma. \end{aligned}$$

For (constant), ζ_* is $\sqrt{\boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu}}$, independent of s_i.
(Volatility time vs. wall-clock time.)
For (floating), it is ${s_i}^{-1} \sqrt{\boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu}}$, higher when volatility is low.
(Volatility drinks your milkshake.)

Heteroskedasticity

- Prior to investment decision, observe s_i proportional to volatility.
- Two competing 'obvious' models:

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

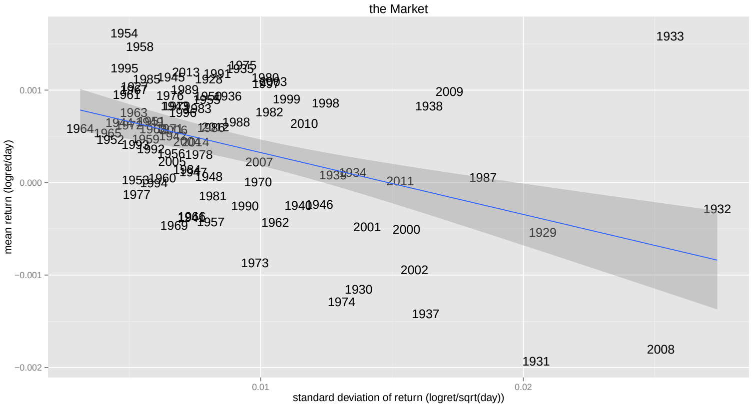
$$\text{(floating):}\quad \mathrm{E}\left[\mathbf{x}_{i+1}\,\middle|\,s_i\right] = \square\,\boldsymbol{\mu} \qquad \mathrm{Var}\left(\mathbf{x}_{i+1}\,\middle|\,s_i\right) = {s_i}^2\Sigma.$$

For (constant), ζ_* is $\sqrt{\boldsymbol{\mu}^{\top}\Sigma^{-1}\boldsymbol{\mu}}$, independent of s_i.
(Volatility time vs. wall-clock time.)
For (floating), it is ${s_i}^{-1}\sqrt{\boldsymbol{\mu}^{\top}\Sigma^{-1}\boldsymbol{\mu}}$, higher when volatility is low.
(Volatility drinks your milkshake.)

- Why do I have to choose?

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

Market Heteroskedasticity



As an aside, both of these models are inadequate for `Mkt`. Shown are mean vs. stdev of daily log returns, 1927 through 2014.

Conditional expectation. I

- Suppose f-vector $\mathbf{f}_i$ observed prior to investment decision, and

$$\text{(conditional):}\quad \mathrm{E}\left[\mathbf{x}_{i+1} \,\middle|\, \mathbf{f}_i\right] = \mathsf{B}\mathbf{f}_i \qquad \mathrm{Var}\left(\mathbf{x}_{i+1} \,\middle|\, \mathbf{f}_i\right) = \Sigma,$$

  B is some $p \times f$ matrix. [6, 11, 3]

- Conditional on observing $\mathbf{f}_i$, solve

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

  for $r_0 \ge 0, R > 0$.
  "Maximize Sharpe, with bound on risk, conditional on $\mathbf{f}_i$."

Conditional expectation. II

- Optimal portfolio is $c\boldsymbol{\nu}_*$ with

$$\boldsymbol{\nu}_* := \boxed{\Sigma^{-1}\mathsf{B}}\,\mathbf{f}_i.$$

- $\boxed{\Sigma^{-1}\mathsf{B}}$ generalizes the Markowitz portfolio: the coefficient of the Sharpe-optimal portfolio linear in features $\mathbf{f}_i$. The 'Markowitz coefficient.'
- Conditional on $\mathbf{f}_i$, the squared signal-noise ratio of the Markowitz portfolio is

$$\zeta^2\left(\Sigma^{-1}\mathsf{B}\mathbf{f}_i\right) | \,\mathbf{f}_i = \left(\mathsf{B}\mathbf{f}_i\right)^{\top}\Sigma^{-1}\left(\mathsf{B}\mathbf{f}_i\right).$$

Typically $\mathbf{f}_i$ is random. Expected squared signal-noise ratio is

$$\mathrm{E}_{\mathbf{f}}\left[\zeta^2\left(\Sigma^{-1}\mathsf{B}\mathbf{f}_i\right)\right] = \operatorname{tr}\left(\mathsf{B}^{\top}\Sigma^{-1}\mathsf{B}\,\mathrm{E}\left[\mathbf{f}\mathbf{f}^{\top}\right]\right).$$

The Hotelling-Lawley Trace.

Conditional expectation. III

- Same weird trick works! Let $\tilde{\tilde{\mathbf{x}}}_{i+1} := \left[{\mathbf{f}_i}^{\top}, {\mathbf{x}_{i+1}}^{\top}\right]^{\top}$.
- The uncentered second moment is

$$\Theta_f := \operatorname{E}\left[\tilde{\tilde{\mathbf{x}}}\tilde{\tilde{\mathbf{x}}}^{\top}\right] = \left[\begin{array}{cc} \Gamma_f & \Gamma_f \mathsf{B}^{\top} \\ \mathsf{B}\Gamma_f & \Sigma + \mathsf{B}\Gamma_f \mathsf{B}^{\top} \end{array}\right], \quad \text{where} \quad \Gamma_f := \operatorname{E}\left[\mathbf{f}\mathbf{f}^{\top}\right].$$

- The inverse of Θ_f is

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

$\Sigma^{-1}\mathsf{B}$ appears in off diagonals.

$\mathsf{B}^{\top}\Sigma^{-1}\mathsf{B}$ related to HLT.

Conditional expectation. IV

- Again, define sample estimator,

$$\hat{\Theta}_f := \frac{1}{n} \sum_{1 \le i \le n} \tilde{\tilde{\mathbf{x}}}_i \tilde{\tilde{\mathbf{x}}}_i^{\top}.$$

- Use Central Limit theorem and delta method to get:

$$\boxed{\sqrt{n}\left(\operatorname{vech}\left(\hat{\Theta}_f^{-1}\right) - \operatorname{vech}\left({\Theta_f}^{-1}\right)\right) \rightsquigarrow \mathcal{N}\left(0, \mathsf{U}\Omega\mathsf{U}^{\top}\right)}$$

Examples. I

- Take the Fama-French 3 factor + Momentum monthly returns (1927-02-01 to 2015-01-01) from *Quandl*. [20]
- Add risk-free rate back to market.
- Use Shiller's P/E ratio as predictive state variable.

```
# Z-score the P/E data
zsc <- function(x, ...) (x - mean(x, ...))/sd(x, ...)
features.z <- zsc(features, na.rm = TRUE)
asym <- MarkowitzR::mp_vcov(ff4.xts[, 1:4], features.z,
    fit.intercept = TRUE, vcov.func = sandwich::vcovHAC)
xtable(signif(t(walds(asym)), digits = 2))
```

	Mkt	SMB	HML	UMD
Intercept	3.10	3.60	2.50	3.60
Cyclically Adjusted PE Ratio	-1.90	-1.20	-0.24	3.70

Examples. II

Now the same, but hedge out Mkt and RF:

```
# hedge out Mkt and RF
Gmat <- matrix(diag(1, 5)[c(1, 5), ], ncol = 5)
asym <- MarkowitzR::mp_vcov(ff4.xts[, 1:5], features.z,
    fit.intercept = TRUE, Gmat = Gmat, vcov.func = sandwich::vcovHAC)
xtable(signif(t(walds(asym)), digits = 2))
```

	Mkt	SMB	HML	UMD	RF
Intercept	0.63	2.20	2.10	2.10	-1.90
Cyclically Adjusted PE Ratio	2.60	-1.60	-0.04	5.50	-1.40

Examples. III

Since we estimate the covariance jointly of $\hat{\boldsymbol{\nu}}_*$ and $\hat{\Sigma}^{-1}$, we can estimate the amount of error in $\hat{\boldsymbol{\nu}}_*$ attributable to mis-estimation of Σ^{-1}; the rest is due to misestimation of $\boldsymbol{\mu}$. [5]
The squared coefficients of multiple correlation, in % of the vanilla Markowitz portfolio:

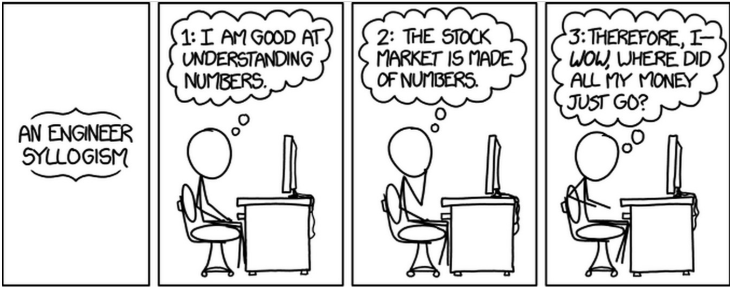
	x Mkt	x SMB	x HML	x UMD
R.squared	38 %	18 %	37 %	60 %

What else?

The same basic model can be adapted to:

- Constrained estimation of Θ. (Linear constraints; rank constraints?)
- Conditional covariance and conditional beta. [8]

Segue


AN ENGINEER SYLLOGISM
1: I AM GOOD AT UNDERSTANDING NUMBERS.
2: THE STOCK MARKET IS MADE OF NUMBERS.
3: THEREFORE, I— WOW, WHERE DID ALL MY MONEY JUST GO?

A wrinkle

- An objection against Hotelling's test is that ζ_*^2 is unlikely to be 0: Keep adding stocks and features and the ζ_*^2 cannot decrease.
- However, this seems not to work in the real world: Portfolio optimization not typically applied to 100's of free variables.

A wrinkle

- An objection against Hotelling's test is that ζ_*^2 is unlikely to be 0: Keep adding stocks and features and the ζ_*^2 cannot decrease.
- However, this seems not to work in the real world: Portfolio optimization not typically applied to 100's of free variables.

Why? "Overfitting."

- Does it suffice to correct for biased estimates of ζ_*^2?
- Does a large universe size negatively impact performance?
- Can we just knock out near-zero elements of the Markowitz coefficient?

Portfolio on a sphere?

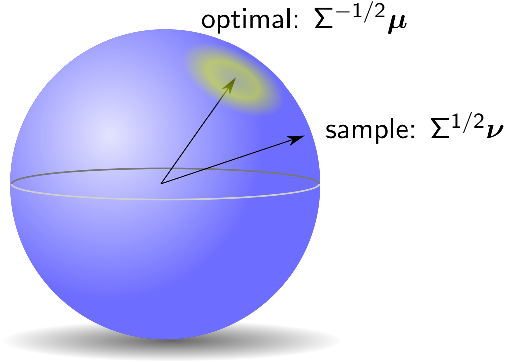
For any portfolio, $\boldsymbol{\nu}$, its signal-noise ratio can be written as:

$$\zeta\left(\boldsymbol{\nu}\right) = \frac{\boldsymbol{\nu}^{\top}\boldsymbol{\mu}}{\sqrt{\boldsymbol{\nu}^{\top}\Sigma\boldsymbol{\nu}}} = \frac{\left(\Sigma^{1/2}\boldsymbol{\nu}\right)^{\top}\Sigma^{-1/2}\boldsymbol{\mu}}{\sqrt{\left(\Sigma^{1/2}\boldsymbol{\nu}\right)^{\top}\left(\Sigma^{1/2}\boldsymbol{\nu}\right)}},$$

$$= \left(\frac{\Sigma^{1/2}\boldsymbol{\nu}}{\left\|\Sigma^{1/2}\boldsymbol{\nu}\right\|_2}\right)^{\top}\left(\frac{\Sigma^{-1/2}\boldsymbol{\mu}}{\left\|\Sigma^{-1/2}\boldsymbol{\mu}\right\|_2}\right)\zeta_*.$$

So $\zeta\left(\boldsymbol{\nu}\right)/\zeta_*$ is the dot product of two vectors on $\mathcal{S}^{p-1}$.
It is bounded from above when distance between vector tips is bounded from below, as by Cramér-Rao bound.

Portfolio on a sphere?



A Theorem I

Consider portfolio construction technique as a function $\hat{\boldsymbol{\nu}}\left(\cdot\right)$:

- Takes historical data, X, a $T \times p$ matrix.
- Produces a portfolio, $\hat{\boldsymbol{\nu}} = \hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)$, a p vector.
- Assume it is not a 'stopped clock'.

If rows of X are *i.i.d.* Gaussian (hold your objections), then

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

Roughly this is

$$\operatorname{E}_{\text{historical data}}\left[\text{portfolio SNR}\right] \le \sqrt{\frac{\text{effect size}}{\#\text{ knobs} + \text{effect size}}}\text{maximal SNR}.$$

A Theorem II

Generalizes to case of conditional expectation and hedge constraints. Requires a slight redefinition of signal-noise ratio.

- For f features and p assets, the bound becomes

$$\mathrm{E}_{\mathsf{X}}\left[\zeta\left(\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)\right] \leq \sqrt{\frac{n\zeta_*^2}{(fp-1)+n\zeta_*^2}}\zeta_*.$$

- If we impose p_g hedge constraints, this becomes

$$\mathrm{E}_{\mathsf{X}}\left[\zeta\left(\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)\right] \leq \sqrt{\frac{n\zeta_*^2}{\left(f\left(p-p_g\right)-1\right)+n\zeta_*^2}}\zeta_*.$$

In summary,

$$\mathrm{E}_{\text{historical data}}\left[\text{portfolio SNR}\right] \leq \sqrt{\frac{\text{effect size}}{\#\ \text{knobs} + \text{effect size}}}\text{maximal SNR}.$$

No Stopped Clocks

- Stopped clock condition prevents *e.g.*, the 'one-over-*n* allocation' from breaking the theorem when the population Markowitz portfolio is nearly equal allocation.
- The technical condition is that $\mathrm{E}_{\mathsf{X}}\left[\zeta\left(\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right)\right)\right]$ is a function only of ζ_{*}.
- This is implied by 'rotational equivariance': if Q is non-singular, then

$$\hat{\boldsymbol{\nu}}\left(\mathsf{X}\mathsf{Q}^{\top}\right) = \mathsf{Q}\hat{\boldsymbol{\nu}}\left(\mathsf{X}\right) \quad \text{up to leverage.}$$

  (Seems reasonable if returns are images of latent factor returns.)
- (If you believe in rotational equivariance, check how you do dimensionality reduction and regularization!)

Some depressing math

The Cramér-Rao bound explains why portfolio optimization is not performed on 100's of unknowns:
If $\zeta_* = 1.1\text{yr}^{-1/2}$, observing $5\text{yr}^{-1/2}$ of data:

- for 10 stocks, the bound is $0.7\text{yr}^{-1/2}$.
- for 40 stocks, the bound is $0.4\text{yr}^{-1/2}$.
- for 160 stocks, the bound is $0.21\text{yr}^{-1/2}$.

Some depressing math

The Cramér-Rao bound explains why portfolio optimization is not performed on 100's of unknowns:
If $\zeta_* = 1.1\text{yr}^{-1/2}$, observing $5\text{yr}^{-1/2}$ of data:

- for 10 stocks, the bound is $0.7\text{yr}^{-1/2}$.
- for 40 stocks, the bound is $0.4\text{yr}^{-1/2}$.
- for 160 stocks, the bound is $0.21\text{yr}^{-1/2}$.

But maximal signal-noise ratio should *grow* with universe size.
Can it grow fast enough?
The 'fundamental law of Active Management' [10] suggests

$$\zeta_* = \zeta_0 p^{1/2}.$$

So explore power law relationships.

Power law bound I

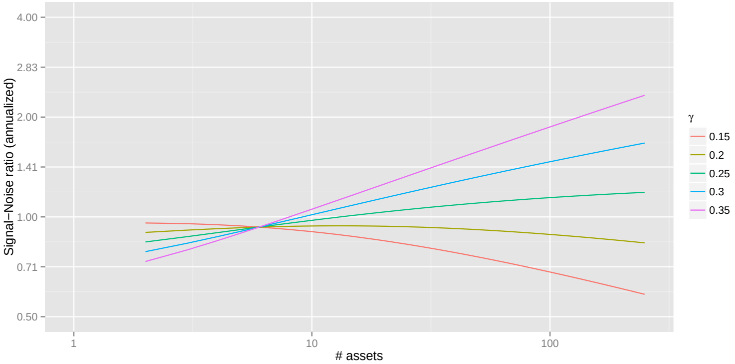


Figure : Bound vs. p for $\zeta_* = \zeta_0 p^\gamma$, with γ between 0.15 and 0.35. ζ_* must grow at rate faster than $\gamma = 1/4$, otherwise bound will decrease.

Power law bound II

Idea: estimate γ from data. On the S&P 100 universe, it looks like $\gamma = 0$:

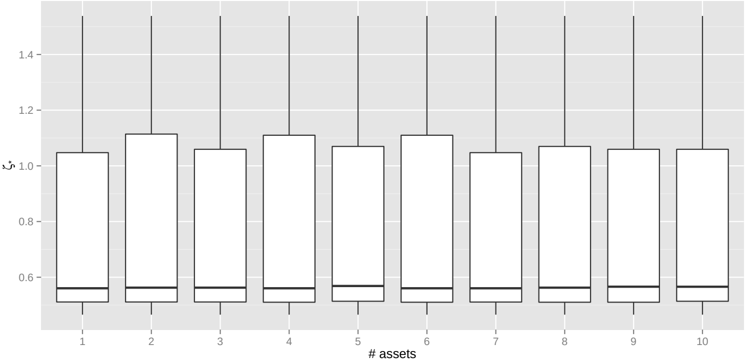


Figure : Estimated ζ_* vs. p for S&P 100 names, over 1000 order permutations.

Philosophical Q&A

"What does this say about *my* portfolio?"

Nothing. It is a frequentist argument about your *method* of constructing portfolios. It does not condition on *e.g.*, $\hat{\zeta}_*^2$. If $\hat{\zeta}_*^2$ is 'large' compared to degrees of freedom, the bound may not be an issue.

"Can I use historical data to reduce the degrees of freedom, and escape the bound?"

Probably not. By using historical data, you subject yourself to the bound or your meta-method is a stopped clock.

"A rational agent cannot be harmed by more data, opportunities."

This is a bad definition, or rational agents cannot exist, or they hold only the market portfolio.

Future work

- Compute confidence intervals on signal-noise ratio of a portfolio?
- Is there a sensible Bayesian version of this result?
- Can we sensibly perform dimensionality reduction using historical data and avoid this ‘overfitting’?
- Is there a more general result which really captures ‘effect size’ and ‘number of knobs’?
- Get a bound on variance of signal-noise ratio of portfolios?
- Prove the bound is worse for returns with fatter tails?

Thank You.

Bibliography I

[1] Clifford S. Asness, Andrea Frazzini, Ronen Israel, and Tobias J. Moskowitz. Fact, fiction and momentum investing. Privately Published, May 2014. URL http://ssrn.com/abstract=2435323.

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

[6] 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.

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

[8] Wayne E. Ferson and Campbell R. Harvey. Conditioning variables and the cross-section of stock returns. Technical report, Fuqua School of Business Working Paper No. 9902, February 1999. URL http://ssrn.com/abstract=152910.

[9] 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.

[10] 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.

Bibliography II

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

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

[17] Steven E. Pav. Bounds on portfolio quality. Privately Published, 2014. URL http://arxiv.org/abs/1409.5936.

[18] Steven E. Pav. *MarkowitzR: Statistical Significance of the Markowitz Portfolio*, 2015. URL https://github.com/shabbychef/MarkowitzR. R package version 0.1502.

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

[20] Raymond McTaggart and Gergely Daroczi. *Quandl: API Wrapper for Quandl.com*, 2015. URL http://CRAN.R-project.org/package=Quandl. R package version 2.6.0.

[21] A. D. Roy. Safety first and the holding of assets. *Econometrica*, 20(3):pp. 431–449, 1952. ISSN 00129682. URL http://www.jstor.org/stable/1907413.

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

Common Questions (Inference) I

Doesn't this require fourth order moments?

I always use relative (or 'percent') returns. These are *bounded*. All moments exist. Identical distribution is a *much* more questionable assumption.

Isn't the complexity $\Omega\left(p^4\right)$?

Portfolio optimization for large p (bigger than 20?) is not typically recommended.

Won't estimating a large number of parameters hurt performance?

The covariance $\operatorname{Var}\left(\operatorname{vech}\left(\tilde{\mathbf{x}}\tilde{\mathbf{x}}^{\top}\right)\right)$ has $\Omega\left(p^4\right)$ elements, but the portfolio is constructed only from $\Omega\left(p^2\right)$ elements, as with vanilla Markowitz.

Common Questions (Inference) II

I want to hedge out exposure to a non-asset.

I want that as well. It does not appear to be a simple modification of the weird trick, but it may be one discovery away.

I want to maximize signal-noise ratio with a time-dependent risk-free rate.

I suspect that the 'right' way to do this is to include the RFR as an asset, then hedge out exposure to it. This effectively allows each asset to have a non-unit 'beta' to the risk-free, which seems like a higher bar than just hedging a constant unit of the risk-free.

What was the quote about the pterodactyl?

It was from the movie, *Airplane*.

Common Questions (Inference) III

I want to hedge out an asset, but I do not want the mean of that asset to be estimated.

I believe this can be done with constrained estimation of $\hat{\Theta}$. Briefly, if there are linear constraints one believes Θ satisfies, you can solve a least-squares problem to get a sample estimate which satisfies the constraints and is not too 'far' from the unconstrained estimator. I have not done the analysis, but believe it is another simple application of the delta method.

The conditional expectation model is many-to-many. How do I sparseify it?

Similar to the above, but I believe one would want to specify linear constraints on the *Cholesky factor* of Θ. This might be more complicated. Or maybe not.

Common Questions (Inference) IV

I don't want to deal with the headaches of symmetry!

The Cholesky factor of Θ is $\left[\begin{array}{cc} 1 & 0 \\ \boldsymbol{\mu} & \Sigma^{1/2} \end{array}\right]$. This is a lower triangular matrix and completely determines Θ. I suspect much of the analysis can be re-couched in terms of this square root, but I do not know the matrix derivative of the Cholesky factorization.

What about a mashup with Kalman Filters?

Sure! This should probably be expressed as an update on the Cholesky factor, $\Theta^{1/2}$.

Which portfolio managers are using the weird trick?

All of them except you!

Common Questions (Inference) V

I am not comforted by the fact that $\hat{\zeta}_*^2 \rightsquigarrow \zeta_*^2$, since the portfolio $\hat{\boldsymbol{\nu}}_*$ may achieve a much lower Sharpe ratio than optimal.

Because $\boldsymbol{\nu}_*$ is the optimal population Sharpe ratio of *any* portfolio, it is an upper bound on the Sharpe ratio of $\hat{\boldsymbol{\nu}}_*$. To estimate the 'gap' requires, I believe, the second-order multivariate delta method. I have not done the analysis.

Can you shoehorn a short-sale constraint into the model?

I doubt it is feasible. It is known, for example, that Hotelling's statistic under a positivity constraint is not a *similar* statistic, indicating Sharpe ratio is an imperfect yardstick for sign-constrained portfolio problems. [22]

Common Questions (Inference) VI

Why maximize Sharpe ratio? Everyone else maximizes 'utility'.

No investor has ever told us their 'risk aversion parameter,' but they ask about our Sharpe ratio all the time. Also, read Roy for the connection between Sharpe ratio and probability of a loss. [21]

How do you deal with trade costs?

It is not clear. One hack would be to assume trade costs *quadratic* in the target portfolio. I believe this merely leads to an inflation of the $\hat{\Sigma}$, but there are likely complications.

Isn't independence of $\tilde{\tilde{\mathbf{x}}}_i$ suspicious?

If the state variables w_i depend on the previous period returns, $\mathbf{x}_i$, independence will be violated. However, the CLT may apply if the sequence is weakly dependent, or 'strongly mixing'.

Common Questions (Inference) VII

How do you detect outliers?

This probably requires one to impose a likelihood on $\tilde{\tilde{\mathbf{x}}}_i$.

Does the math simplify if you assume normal returns?

In this case $n\hat{\Theta}$ takes a *conditional* Wishart distribution.

But does it do big data?

Computation of $\hat{\Theta}$ is very simple, since it is just an uncentered moment...

How should a Bayesian approach estimation of Θ?

I don't know. Ask one. I suspect they would assume normal returns, then assume some kind of conditional Wishart prior.

Common Questions (Inference) VIII

Does the hedged portfolio involve a projection?

It does! The hedged portfolio is the optimal portfolio minus a projection under the metric induced by Σ.

It seems that when I hedge out a single asset, only the holdings in that asset change in the portfolio.

If you look at the projection operation, the change can only occur in the column space of $\tilde{G}^{\top}$, which in this case means only the holdings in the single asset will change. (This is all modulo adjustments to overall gross leverage to meet the risk budget.)

Can you back out the traditional significance tests from the asymptotic distribution of $\hat{\Theta}$?

Possibly, but probably a bit uglier than I can stomach.

Common Questions (Overfit) I

The proof assumes *normal* returns?

Indeed, however I suspect that a stronger upper bound holds for the case of more fat-tailed distributions, though I do not have a proof yet.

This bound uses an unknown population parameter. Can you do better?

Not at the moment. This is a particularly interesting question: how to construct confidence intervals on the Sharpe of the Markowitz portfolio. It is different than the typical statistical analysis, which performs inference on the maximal population Sharpe ratio.

What does this bound say about *my* portfolio?

Very little. It only gives a bound on the *expected* Sharpe based on repeated draws of the historical data. You only got one draw of that historical data.

Common Questions (Overfit) II

I don't like this result: it seems I can be harmed by performing more backtests.

It is commonly stipulated that a perfectly rational agent cannot be harmed by the addition of new information or new optional courses of action. Barring the fact that humans are not perfectly rational agents, a quantitative trading scheme that can only improve with the addition of new information sounds like the Holy Grail. Like the Holy Grail, it is unlikely to exist.

This Cramér-Rao bound feels very Frequentist.

I suppose it does.

How did you estimate ζ_* in the S&P 100 study?

I used the 'KRS' estimator from the SharpeR package.