Portfolio inference with this one weird trick

Steven E. Pav
steven@cerebellumcapital.com

Cerebellum Capital

May 16, 2014

A weird ...

- Consider p-vector of asset returns, $\mathbf{x}$. Let:

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

- 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 $\mathbf{x}$:

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

… trick

$$\text{then:}\quad \Theta^{-1} = \begin{bmatrix} 1+\boldsymbol{\mu}^\top\Sigma^{-1}\boldsymbol{\mu} & -\boldsymbol{\mu}^\top\Sigma^{-1} \\ -\Sigma^{-1}\boldsymbol{\mu} & \Sigma^{-1} \end{bmatrix}$$

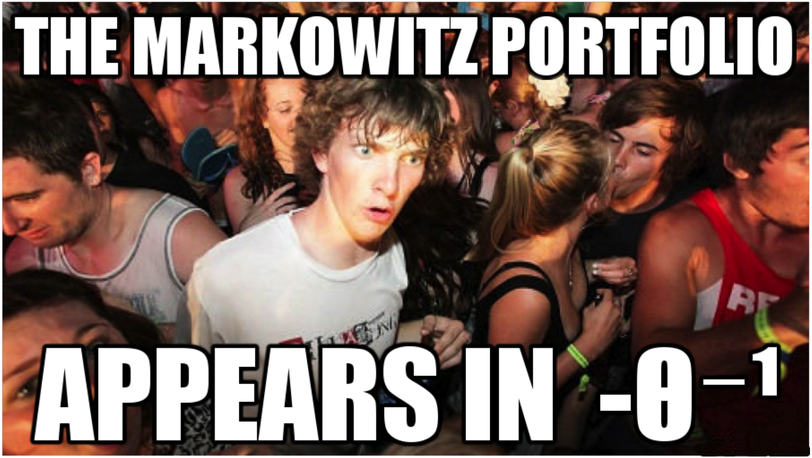
$$= \begin{bmatrix} 1+\zeta_*^2 & -\boldsymbol{\nu}_*{}^\top \\ -\boldsymbol{\nu}_* & \Sigma^{-1} \end{bmatrix},$$

- $\boldsymbol{\nu}_*$ is the Markowitz portfolio,
  ζ_* is the Sharpe ratio of $\boldsymbol{\nu}_*$ (*cf.* Hotelling's T^2),
  Σ^{-1} is the 'precision matrix'.
- The portfolio is 'optimal', solving *e.g.*, Roy's problem: [17]

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

i.e., "maximize Sharpe with a bound on risk."

But is it useful?



Sample estimator

- Since $\Theta = \operatorname{E}\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^{\top}\right]$ 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:

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

  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:

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

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

Asymptotics II

- Ignoring details about symmetry, *etc.*, the derivative is: [7, 12]

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

(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 $\operatorname{Var}\left(\operatorname{vech}\left(\hat{\Theta}^{-1}\right)\right)$?

- Perform inference on elements of $\boldsymbol{\nu}_*$ via Wald statistic. (Compare elements of $\boldsymbol{\nu}_*$ to their standard errors.)
- Perform inference on the maximal Sharpe 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
```

Game over?

Selling this weird trick

Why weird trick, not Britten-Jones, or Okhrin *et al.*? [4, 2, 14]

- 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, 14]

- 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

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

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 Sharpe ratio with risk bound and zero covariance to some other portfolios."
Solved by $c\boldsymbol{\nu}_{\mathsf{G},*}$, with c to satisfy risk bound, and

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

Hedged portfolios III

- Use the weird trick! Let $\tilde{G} := \left[\begin{array}{cc} 1 & 0 \\ 0 & G \end{array}\right]$, then,

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

$$\left[\begin{array}{cc} \boldsymbol{\mu}^{\top}\Sigma^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^{\top}G^{\top}\left(G\Sigma G^{\top}\right)^{-1}G\boldsymbol{\mu} & -\boldsymbol{\nu}_{G,*}{}^{\top} \\ -\boldsymbol{\nu}_{G,*} & \Sigma^{-1} - G^{\top}\left(G\Sigma G^{\top}\right)^{-1}G \end{array}\right].$$

$-\boldsymbol{\nu}_{G,*}$ is the optimal hedged portfolio.

UL corner is squared Sharpe ratio of $\boldsymbol{\nu}_{G,*}$.

Also used for portfolio spanning. [16, 8, 10, 11]

LR corner is loss of precision?

Hedged portfolios IV

- Delta method gives the asymptotic distribution:

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

with more ugly derivatives.

Hedged portfolios V

- Download the Fama-French 3 factor + Momentum monthly data (1927-02-01 to 2014-01-01) from *Quandl*. [13]
- 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	6.28	0.72	4.99	8.20
weird trick Wald stat	5.37	0.77	4.47	6.03
weird trick w/ HAC	5.10	0.77	3.92	5.53

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	2.69	0.77	4.47	6.03

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	0.71	2.05	2.27	3.43	-1.30

And the spanning statistic:

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

## [1] 2.1
```

Heteroskedasticity

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

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

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 models:

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

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.)

- Why do I have to choose?

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

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, 9, 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

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

- $\boxed{\mathsf{\Sigma}^{-1}\mathsf{B}}$ generalizes the Markowitz portfolio:
  the coefficient of the Sharpe-optimal portfolio linear in features $\mathbf{f}_i$.
  The 'Markowitz coefficient.'

Conditional 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

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

- 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:

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

Examples. I

- Take the Fama-French 3 factor + Momentum monthly returns (1927-02-01 to 2014-01-01) from *Quandl*. [13]
- 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	3.30	2.40	3.50
Cyclically Adjusted PE Ratio	-1.80	-1.00	-0.09	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	0.55	2.10	2.10	2.10	-1.50
Cyclically Adjusted PE Ratio	2.20	-1.20	0.12	7.50	-1.40

What's next?

- Constrained estimation of Θ. (Linear constraints; rank constraints?)
- Generalize to higher dimensions?
- Fancier hedging model?
- Conditional covariance models?
- Jak's Tap?

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 http://shr.receptidocs.ru/docs/5/4748/conv_1/file1.pdf#page=298.

[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] 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.

[9] 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.

[10] 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.

Bibliography II

[11] 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.

[12] 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.

[13] Raymond McTaggart and Gergely Daroczi. *Quandl: Quandl Data Connection*, 2014. URL http://CRAN.R-project.org/package=Quandl. R package version 2.3.2.

[14] 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.

[15] Steven E. Pav. *MarkowitzR: Statistical significance of the Markowitz portfolio*, 2014. URL http://www.r-project.org,https://github.com/shabbychef/MarkowitzR. R package version 0.1403.

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

[17] 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.

[18] 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 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 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 Sharpe 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 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 IV

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

The Cholesky factor of Θ is $\left[\begin{array}{cc} 1 & 0 \\ \mu & \Sigma^{1/2} \end{array}\right]$. 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 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. [18]

Common Questions 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. [17]

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 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 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.