Maximizing Sharpe and re-inventing the wheel

Steven E. Pav
Cerebellum Capital
steven@cerebellumcapital.com
(@shabbychef)

May 11, 2012

- The "Reward-to-Variability" ratio introduced in 1966.[10]
- Inherently a sample statistic:

$$\hat{\zeta} =_{\text{df}} \frac{\hat{\mu}}{\hat{\sigma}},$$

  where $\hat{\mu}$ is sample mean, $\hat{\sigma}$ sample standard deviation of strategy returns. (Ignore risk-free rate for simplicity.)
- Call it 'SR'. Contrast with population analogue, $\zeta =_{\text{df}} \mu/\sigma$ the 'Signal to Noise Ratio' (SNR).
- Units of SR and SNR are "one over root time." So "annualized" SR units are $\text{yr}^{-1/2}$.

- Actually the same as Student's ratio.[3]

  of $z = \frac{x}{s}$.

  Later converted to 'modern' form by Fisher.[2]
- Sharpe ratio is, up to scaling, the t-statistic:

$$t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{n}} = \sqrt{n}\hat{\zeta}.$$

- $\hat{\zeta}/\sqrt{n}$ asymptotically takes (non-central) t-distribution.
- Can test not just mean, but the SNR $\zeta = \mu/\sigma$.

```
p.sr <- function(sr,df,snr=0,a.fact=252) {
  #H_0: SNR=given snr vs H_1: SNR>given snr
  p.val <- pt(sr*sqrt(df/a.fact),df-1,
              snr*sqrt(df/a.fact),lower.tail=FALSE)
}
```

Why it Matters: the *t*-stat is well studied

- Standard error estimation: for SR by Lo (2002), for *t*-stat by Johnson and Welch (1940). [7, 5]

$$\mathrm{SE}(\widehat{\mathrm{SR}}) \stackrel{a}{=} \sqrt{\left(1+\tfrac{1}{2}\mathrm{SR}^2\right)/T},$$

$$\left(1+\frac{\delta^2}{2f}\right)^{\frac{1}{2}}.$$

- Confidence intervals via *t*-distribution:

```
qco.sr <- function(sr,df,alpha,a.fact=252) {
  #q'tile of S.R. 'Confidence Distribution'
  adj.c <- sqrt(df/a.fact)
  find.ncp <- uniroot(function(ncp)
        (pt(sr*adj.c,df-1,ncp) - (1-alpha)),
        interval=c(-36,36)) # a hack!
  return(find.ncp$root / adj.c)
}
sr.ci <- function(sr,df,alpha,...) {  # CI
  return(c(qco.sr(sr,df,alpha/2,...),
           qco.sr(sr,df,1-(alpha/2),...)))
}
```

Relation between sample size, effect size (SNR), and rates of false positives and false negatives.

- Hypothetical Vendor: "I have three years of historical data."
- Hypothetical Strategist: "This strategy has SNR $0.6\mathrm{yr}^{-1/2}$."
- Hypothetical Investor: "You have one year to prove yourself."

Good sample size approximation of form

$$n \approx \frac{\kappa}{\zeta^2},$$

with κ a function of type I and type II rates, 1-sided or 2-sided test, *etc.* "Lehr's rule." [14, 5]

	one.sided	two.sided
power $= 0.50$	2.72	3.86
power $= 0.80$	6.20	7.87

Table: Value of κ to achieve given power in t-test, $\alpha = 0.05$.

remember this rule

For test that SNR is *zero*, with 0.05 type I rate, and 50 percent power, let

$$n \approx \frac{2.72}{\zeta^2}. \qquad \text{mnemonic form: } e \approx n\zeta^2$$

Using the rule:

- Hypothetical Vendor: "I have three years of historical data."
  Answer: Strategy must have SNR $\geq 0.95\text{yr}^{-1/2}$.
- Hypothetical Strategist: "This strategy has SNR $0.6\text{yr}^{-1/2}$."
  Answer: Need 7.6 years of data to backtest. (!)
- Hypothetical Investor: "You have one year to prove yourself."
  Answer: Strategy must have SNR $\geq 1.6\text{yr}^{-1/2}$.

Generalizing to Hotelling

- Hotelling T^2 is multivariate generalization of Student's t:[4]

$$T^2 =_{\text{df}} n\hat{\boldsymbol{\mu}}^{\top}\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}.$$

- An Instance of Roy's Union-Intersection Test:[9]

$$T^2 = n \max_{\mathbf{w}} \left(\frac{\mathbf{w}^{\top}\hat{\boldsymbol{\mu}}}{\sqrt{\mathbf{w}^{\top}\hat{\Sigma}\mathbf{w}}}\right)^2 = n \max_{\mathbf{w}} \hat{\zeta}^2\left(\mathbf{w}\right) = n\hat{\zeta}_*^2,$$

  where $\hat{\zeta}_*$ is SR of Markowitz portfolio $\mathbf{w}_* \propto \hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}$.
- Via Hotelling T^2, maximal Sharpe squared has F-distribution:

$$\frac{(n-p)}{p(n-1)} n\hat{\zeta}_*^2 \sim F\left(n\zeta_*^2, p, n-p\right),$$

  where ζ_* is SNR of *population* optimal portfolio, $\boldsymbol{\nu}_* \propto \Sigma^{-1}\boldsymbol{\mu}$.
- *n.b.* SNR of $\mathbf{w}_*$ is no greater than ζ_*.

Inference on ζ_*

- Received wisdom: "never look at $\hat{\zeta}_*$ when computing $\mathbf{w}_*$."
- Via F-distribution we have

$$\mathrm{E}\left[\hat{\zeta}_*^2\right] = \frac{\zeta_*^2 + c}{1 - c}, \qquad \text{where } c = p/n,$$
$$n \,\mathrm{var}\left(\hat{\zeta}_*^2\right) \approx \frac{2\left(c + 2\zeta_*^2\right)}{(1 - c)^3}.$$

  Gives unbiased estimator: $\mathrm{E}\left[(1 - c)\hat{\zeta}_*^2 - c\right] = \zeta_*^2$.
- CI on ζ_* by inverting F-CDF; MLE by maximizing F-PDF. (Use `pf` and `df`.)
- Via F-distribution, if $\hat{\zeta}_*^2 \leq \frac{c}{1-c}$ then MLE of ζ_* is 0. [13]

rule of thumb

"If $\hat{\zeta}_*^2 < c$, don't trade it!"

Index data `BASI`, `INDU`, `CONG`, `HLTH`, `CONS`, `TELE`, `UTIL`, `FINA`, `TECH` from `fPortfolio::SPISECTOR`, from 2000-01-04 to 2008-10-17. ($n = 2198$ days, $p = 9$ stocks)

- Optimal in-sample Sharpe ratio is $1\mathrm{yr}^{-1/2}$.
- MLE for ζ_* is $0.1\mathrm{yr}^{-1/2}$.
  Note $\hat{\zeta}_*^2 = 0.00415\mathrm{day}^{-1}$ and $c = 0.00409\mathrm{day}^{-1}$, close to the rule of thumb cutoff.
- 95% CI for ζ_* is $\left[0\mathrm{yr}^{-1/2}, 1.3\mathrm{yr}^{-1/2}\right]$

Caricature of quant work:

- Construct strategy parametrized by θ;
- Backtest strategy for $\theta_1, \theta_2, \ldots, \theta_m$.
- Pick θ_i that maximizes SR of backtest, call it θ_*.
- Profit! (or not)

Toy Example: Moving Average Crossover:

- θ is vector of 2 window lengths; Long the instrument exactly when one moving average exceeds the other.
- Brute-force backtest for allowable window lengths.

Q: How to estimate the SNR of θ_*?

Approximation of Strategy Overfit II

Q: How to estimate the SNR of θ_*?
A(?): Make PCA-like linear approximation of returns vectors:

$$\{\mathbf{x_1}, \ldots, \mathbf{x_m}\} \approx \mathcal{L} \subset \{\mathsf{Y}\mathbf{w} \,|\, \mathbf{w} \in \mathbb{R}^p\}$$

Then use Hotelling to make rough inference on ζ_*.
Only need to observe

$$\hat{\zeta}_* = \hat{\zeta}\left(\theta_*\right) =_{\mathrm{df}} \max_{1 \le i \le m} \hat{\zeta}\left(\theta_i\right),$$

and estimate p.
Estimate p by Monte Carlo under null, PCA, or SWAG method.

Overfit of Simple 2-Window MAC

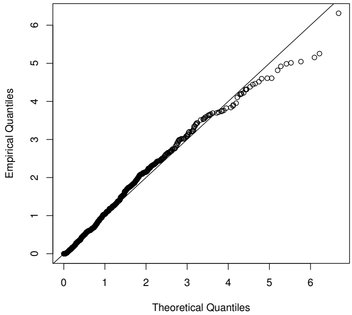

Figure: 1024 Monte Carlo sims of model selection in 2-window MAC over 2500 days, under the null (no population drift or autocorrelation). In-sample $\hat{\zeta}_*$ values transformed to F-statistics with $p = 2.075$.

Use on `GSPC` adjusted returns (2000-01-03 to 2009-12-31, 2515 days), MLE of ζ_* is $0.65\text{yr}^{-1/2}$; 95% CI is $\left(0\text{yr}^{-1/2}, 1.3\text{yr}^{-1/2}\right)$.

- Can generalize Hotelling statistic to get 'Spanning Tests' [1, 6]
- Let T^2_{p+q}, T^2_{p} be Hotelling stats on full set of $p+q$ assets and subset of p assets. Do the q marginal assets 'add any value'.
- Let

$$\Delta T^2 = (n-p-1)\frac{T^2_{p+q} - T^2_{p}}{n-1+T^2_{p}} = (n-p-1)\frac{\hat{\zeta}^2_{*,p+q} - \hat{\zeta}^2_{*,p}}{1-(1/n)+\hat{\zeta}^2_{*,p}}.$$

- ΔT^2 also takes a (non-central) Hotelling distribution:

$$\frac{n-(p+q)}{q(n-p-1)}\Delta T^2 \sim F\left(q, n-(p+q), \delta\right),$$

$$\delta = (n-p-1)\frac{\zeta^2_{*,p+q} - \zeta^2_{*,p}}{1-(1/n)+\zeta^2_{*,p}}.$$

- Same inference on δ can be applied (MLE, CI).

Delta Hotelling: what do `BASI`, `INDU`, `CONG`, `HLTH`, `CONS`, `TELE` add to `UTIL`, `FINA`, `TECH`?

- MLE for $\left(\zeta^2_{*,p+q} - \zeta^2_{*,p}\right)$ is $0\mathrm{yr}^{-1}$.
- 95% CI for $\left(\zeta^2_{*,p+q} - \zeta^2_{*,p}\right)$ is $\left[0\mathrm{yr}^{-1}, 0.26\mathrm{yr}^{-1}\right]$.

- Problems under constrained portfolios; T^2_+ is *not* a 'similar' statistic; distribution depends on nuisance parameter Σ. [11, 8, 12]
- Inference on $\hat{\zeta}_*$ not the same as inference on $\mathbf{w}_*$; Model the SNR of $\mathbf{w}_*$.
- Version 0.01 of `ratarb` package.

Bibliography I

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.

Ronald Aylmer Fisher.
Applications of "Student's" distribution.
Metron, 5:90–104, 1925.
URL http://hdl.handle.net/2440/15187.

William Sealy Gosset.
The probable error of a mean.
Biometrika, 6(1):1–25, March 1908.
URL http://dx.doi.org/10.2307/2331554.
Originally published under the pseudonym "Student".

Harold Hotelling.
The generalization of Student's ratio.
Annals of Mathematical Statistics, 2(3):360–378, 1931.
URL http://projecteuclid.org/euclid.aoms/1177732979.

N. L. Johnson and B. L. Welch.
Applications of the non-central t-distribution.
Biometrika, 31(3-4):362–389, March 1940.
URL http://dx.doi.org/10.1093/biomet/31.3-4.362.

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.

Bibliography II

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

Pranab K. Sen and Ming-Tien Tsai.
Two-stage likelihood ratio and union-intersection tests for one-sided alternatives multivariate mean with nuisance dispersion matrix.
Journal of Multivariate Analysis, 68(2):264 – 282, 1999.
ISSN 0047-259X.
URL `http://www.sciencedirect.com/science/article/pii/S0047259X98917912`.

Pranab Kumar Sen.
Union-intersection principle and constrained statistical inference.
Journal of Statistical Planning and Inference, 137(11):3741 – 3752, 2007.
ISSN 0378-3758.
URL `http://www.sciencedirect.com/science/article/pii/S0378375807001310`.
<ce:title>Special Issue: In Celebration of the Centennial of The Birth of Samarendra Nath Roy (1906-1964)</ce:title>.

William F. Sharpe.
Mutual fund performance.
Journal of Business, 39:119, 1965.
URL `http://ideas.repec.org/a/ucp/jnlbus/v39y1965p119.html`.

Mervyn J. Silvapulle.
On an F-type statistic for testing one-sided hypotheses and computation of chi-bar-squared weights.
Statistics and Probability Letters, 28(2):137 – 141, 1996.
ISSN 0167-7152.
URL `http://www.sciencedirect.com/science/article/pii/0167715295001085`.

Mervyn J. Silvapulle and Pranab Kumar Sen.
Constrained statistical inference : inequality, order, and shape restrictions.
Wiley-Interscience, Hoboken, N.J. :, 2005.
ISBN 0471208272.

M. C. Spruill.
Computation of the maximum likelihood estimate of a noncentrality parameter.
Journal of Multivariate Analysis, 18(2):216 – 224, 1986.
ISSN 0047-259X.
URL http://www.sciencedirect.com/science/article/pii/0047259X86900709.

Gerald van Belle.
Statistical Rules of Thumb (Wiley Series in Probability and Statistics).
Wiley-Interscience, 1 edition, March 2002.
ISBN 0471402273.
URL http://www.vanbelle.org/.

distribution	param	skew	ex.kurtosis	typeI
Gaussian		0	0	0.048
Student's t	df = 10	0	1	0.039
SP500		-1	26	0.077
symmetric SP500		0	25	0.063
Tukey h	h = 0.1	0	5.5	0.057
Tukey h	h = 0.24	0	1.3e+03	0.056
Tukey h	h = 0.4		Inf	0.15
Lambert W x Gaussian	delta = -0.2	-1.2	5.7	0.05
Lambert W x Gaussian	delta = -0.4	-2.7	18	0.08
Lambert W x Gaussian	delta = -1.2	-30	5.2e+03	0.28

Table: Empirical type I rates of the test for $\zeta = 1.0$ via distribution of the Sharpe ratio are given for various distributions of returns. The empirical rates are based on 1024 simulations of three years of daily returns, with a nominal rate of $\alpha = 0.05$. Skew appears to have a much more adverse effect than kurtosis alone.