Robust Performance Hypothesis Testing with the Sharpe Ratio

Olivier Ledoit Michael Wolf

Institute for Empirical Research in Economics
University of Zurich

Outline

Outline

General Set-Up & Notation

We use the same notation as Jobson and Korkie (1981):

- There are two investment strategies i and n
- Their excess returns are r_{ti} and r_{tn}, for $t = 1, \ldots, T$
- Bivariate return series is assumed stationary with

$$\mu = \begin{pmatrix} \mu_i \\ \mu_n \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_i^2 & \sigma_{in} \\ \sigma_{in} & \sigma_n^2 \end{pmatrix}$$

General Set-Up & Notation

We use the same notation as Jobson and Korkie (1981):

- There are two investment strategies i and n
- Their excess returns are r_{ti} and r_{tn}, for $t = 1, \ldots, T$
- Bivariate return series is assumed stationary with

$$\mu = \begin{pmatrix} \mu_i \\ \mu_n \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_i^2 & \sigma_{in} \\ \sigma_{in} & \sigma_n^2 \end{pmatrix}$$

Parameter of interest:

- Difference between the two Sharpe ratios:

$$\Delta = Sh_i - Sh_n = \frac{\mu_i}{\sigma_i} - \frac{\mu_n}{\sigma_n}$$

- Estimator is given by:

$$\hat{\Delta} = \widehat{Sh}_i - \widehat{Sh}_n = \frac{\hat{\mu}_i}{\hat{\sigma}_i} - \frac{\hat{\mu}_n}{\hat{\sigma}_n}$$

Common Approach: JK (1981) and Memmel (2003)

Approach:

- Let $u = (\mu_i, \mu_n, \sigma_i^2, \sigma_n^2)'$ and $\hat{u} = (\hat{\mu}_i, \hat{\mu}_n, \hat{\sigma}_i^2, \hat{\sigma}_n^2)'$
- If the returns are i.i.d. bivariate normal:

$$\sqrt{T}(\hat{u} - u) \xrightarrow{d} N(0; \Omega) \quad \text{with} \quad \Omega = \begin{pmatrix} \sigma_i^2 & \sigma_{in} & 0 & 0 \\ \sigma_{in} & \sigma_n^2 & 0 & 0 \\ 0 & 0 & 2\sigma_i^4 & 2\sigma_{in}^2 \\ 0 & 0 & 2\sigma_{in}^2 & 2\sigma_n^4 \end{pmatrix}$$

- Get a standard error for $\hat{\Delta}$ from $\hat{\Omega}$ and the delta method
- Allows to test H_0: $\Delta = 0$ or to compute a CI for Δ

Common Approach: JK (1981) and Memmel (2003)

Approach:

- Let $u = (\mu_i, \mu_n, \sigma_i^2, \sigma_n^2)'$ and $\hat{u} = (\hat{\mu}_i, \hat{\mu}_n, \hat{\sigma}_i^2, \hat{\sigma}_n^2)'$
- If the returns are i.i.d. bivariate normal:

$$\sqrt{T}(\hat{u} - u) \xrightarrow{d} N(0; \Omega) \quad \text{with} \quad \Omega = \begin{pmatrix} \sigma_i^2 & \sigma_{in} & 0 & 0 \\ \sigma_{in} & \sigma_n^2 & 0 & 0 \\ 0 & 0 & 2\sigma_i^4 & 2\sigma_{in}^2 \\ 0 & 0 & 2\sigma_{in}^2 & 2\sigma_n^4 \end{pmatrix}$$

- Get a standard error for $\hat{\Delta}$ from $\hat{\Omega}$ and the delta method
- Allows to test H_0: $\Delta = 0$ or to compute a CI for Δ

Pitfall:

- The above formula for Ω is not robust against heavy tails or time series effects (which are typical for financial returns)
- So the corresponding inference is generally not valid

Outline

Outline

General Idea

- Let $\gamma_i = \mathrm{E}(r_{1i}^2)$, $\gamma_n = \mathrm{E}(r_{1n}^2)$, and $v = (\mu_i, \mu_n, \gamma_i, \gamma_n)'$. Then:

$$\Delta = f(v) \quad \text{with} \quad f(a,b,c,d) = \frac{a}{\sqrt{c-a^2}} - \frac{b}{\sqrt{d-b^2}}$$

- If $\sqrt{T}(\hat{v} - v) \xrightarrow{d} N(0;\Psi)$, the delta method implies:

$$\sqrt{T}(\hat{\Delta} - \Delta) \xrightarrow{d} N(0; \nabla' f(v) \Psi \nabla f(v))$$

General Idea

- Let $\gamma_i = \mathrm{E}(r_{1i}^2)$, $\gamma_n = \mathrm{E}(r_{1n}^2)$, and $v = (\mu_i, \mu_n, \gamma_i, \gamma_n)'$. Then:

$$\Delta = f(v) \quad \text{with} \quad f(a,b,c,d) = \frac{a}{\sqrt{c-a^2}} - \frac{b}{\sqrt{d-b^2}}$$

- If $\sqrt{T}(\hat{v} - v) \xrightarrow{d} N(0;\Psi)$, the delta method implies:

$$\sqrt{T}(\hat{\Delta} - \Delta) \xrightarrow{d} N(0; \nabla' f(v) \Psi \nabla f(v))$$

- So a standard error for $\hat{\Delta}$ is given by:

$$s(\hat{\Delta}) = \sqrt{\frac{\nabla' f(\hat{v}) \hat{\Psi} \nabla f(\hat{v})}{T}}$$

General Idea

- Let $\gamma_i = \mathrm{E}(r_{1i}^2)$, $\gamma_n = \mathrm{E}(r_{1n}^2)$, and $v = (\mu_i, \mu_n, \gamma_i, \gamma_n)'$. Then:

$$\Delta = f(v) \quad \text{with} \quad f(a,b,c,d) = \frac{a}{\sqrt{c-a^2}} - \frac{b}{\sqrt{d-b^2}}$$

- If $\sqrt{T}(\hat{v} - v) \xrightarrow{d} N(0;\Psi)$, the delta method implies:

$$\sqrt{T}(\hat{\Delta} - \Delta) \xrightarrow{d} N(0;\nabla' f(v)\Psi\nabla f(v))$$

- So a standard error for $\hat{\Delta}$ is given by:

$$s(\hat{\Delta}) = \sqrt{\frac{\nabla' f(\hat{v})\hat{\Psi}\nabla f(\hat{v})}{T}}$$

Remaining challenge: find a consistent estimator $\hat{\Psi}$ for Ψ.

Kernel Estimator for Ψ

$$\hat{\Psi} = \hat{\Psi}_T = \frac{T}{T-4} \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S_T}\right) \hat{\Gamma}_T(j)$$

Details:

- $\hat{\Gamma}_T(j) = \begin{cases} \frac{1}{T} \sum_{t=j+1}^{T} \hat{y}_t \hat{y}'_{t-j} & \text{for } j \geq 0 \\ \frac{1}{T} \sum_{t=-j+1}^{T} \hat{y}_{t+j} \hat{y}'_t & \text{for } j < 0 \end{cases}$
- $\hat{y}'_t = (r_{ti} - \hat{\mu}_1, r_{tn} - \hat{\mu}_n, r_{ti}^2 - \hat{\gamma}_i, r_{tn}^2 - \hat{\gamma}_n)$
- $k(\cdot)$ is a kernel and S_T is a corresponding bandwidth

Kernel Estimator for Ψ

$$\hat{\Psi} = \hat{\Psi}_T = \frac{T}{T-4} \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S_T}\right) \hat{\Gamma}_T(j)$$

Details:

- $\hat{\Gamma}_T(j) = \begin{cases} \frac{1}{T}\sum_{t=j+1}^{T} \hat{y}_t \hat{y}'_{t-j} & \text{for } j \geq 0 \\ \frac{1}{T}\sum_{t=-j+1}^{T} \hat{y}_{t+j} \hat{y}'_t & \text{for } j < 0 \end{cases}$
- $\hat{y}'_t = (r_{ti} - \hat{\mu}_1, r_{tn} - \hat{\mu}_n, r_{ti}^2 - \hat{\gamma}_i, r_{tn}^2 - \hat{\gamma}_n)$
- $k(\cdot)$ is a kernel and S_T is a corresponding bandwidth

Specific suggestions:

- QS kernel or prewhitened QS kernel
- Data-dependent choice of bandwidth
- See Andrews (1991) and Andrews & Monahan (1992)

Resulting Inference

- Standard error for $\hat{\Delta}$:

$$s(\hat{\Delta}) = \sqrt{\frac{\nabla' f(\hat{v}) \hat{\Psi} \nabla f(\hat{v})}{T}}$$

- A two-sided p-value for H: $\Delta = 0$ is given by:

$$\hat{p} = 2\Phi\left(-\frac{|\hat{\Delta}|}{s(\hat{\Delta})}\right)$$

where $\Phi(\cdot)$ denotes the c.d.f. of $N(0,1)$

- A two-sided confidence interval for Δ is given by

$$\hat{\Delta} \pm z_{1-\alpha/2}\, s(\hat{\Delta})$$

where z_λ denotes the λ quantile of $N(0,1)$

Two Remarks

Remark:

- Lo (2002) discusses inference for a single Sharpe Ratio Sh
- Section "IID Returns" corresponds to Memmel (2003)
- Section "Non-IID Returns" corresponds to HAC inference (though he uses an 'inferior' kernel and does not deal with the choice of the bandwidth)

Two Remarks

Remark:

- Lo (2002) discusses inference for a single Sharpe Ratio *Sh*
- Section "IID Returns" corresponds to Memmel (2003)
- Section "Non-IID Returns" corresponds to HAC inference (though he uses an 'inferior' kernel and does not deal with the choice of the bandwidth)

Remark:

- Opdyke (2007) discusses inference for both *Sh* and Δ
- First, he considers general i.i.d. data
- Then, he considers general time series data
- But his formulas for the time series case are actually equivalent to those for the i.i.d. case . . .

Outline

Basic Idea

Construct a two-sided symmetric confidence interval for Δ:

- Approximate the 'absolute' sampling distribution of the studentized statistic using the bootstrap:

$$\mathcal{L}\left(\frac{|\hat{\Delta} - \Delta|}{s(\hat{\Delta})}\right) \approx \mathcal{L}\left(\frac{|\hat{\Delta}^* - \hat{\Delta}|}{s(\hat{\Delta}^*)}\right)$$

where $\mathcal{L}(\cdot)$ denotes the law of a random variable

Basic Idea

Construct a two-sided symmetric confidence interval for Δ:

- Approximate the 'absolute' sampling distribution of the studentized statistic using the bootstrap:

$$\mathcal{L}\left(\frac{|\hat{\Delta}-\Delta|}{s(\hat{\Delta})}\right) \approx \mathcal{L}\left(\frac{|\hat{\Delta}^*-\hat{\Delta}|}{s(\hat{\Delta}^*)}\right)$$

  where $\mathcal{L}(\cdot)$ denotes the law of a random variable
- Let $z^*_{|\cdot|,\lambda}$ be a λ quantile of the r.h.s. distribution
- A two-sided symmetric confidence interval is given by:

$$\hat{\Delta} \pm z^*_{|\cdot|,1-\alpha}\, s(\hat{\Delta})$$

Basic Idea

Construct a two-sided symmetric confidence interval for Δ:

- Approximate the 'absolute' sampling distribution of the studentized statistic using the bootstrap:

$$\mathcal{L}\left(\frac{|\hat{\Delta}-\Delta|}{s(\hat{\Delta})}\right) \approx \mathcal{L}\left(\frac{|\hat{\Delta}^{*}-\hat{\Delta}|}{s(\hat{\Delta}^{*})}\right)$$

  where $\mathcal{L}(\cdot)$ denotes the law of a random variable
- Let $z^{*}_{|\cdot|,\lambda}$ be a λ quantile of the r.h.s. distribution
- A two-sided symmetric confidence interval is given by:

$$\hat{\Delta} \pm z^{*}_{|\cdot|,1-\alpha}\, s(\hat{\Delta})$$

Reject H_0: $\Delta = 0$ at level α if 0 is not contained in the CI.

Details

Bootstrap world:

- Bootstrap data are generated using the circular block bootstrap
- This avoids the 'edge effects' of the moving blocks bootstrap, and so one can recenter simply using $\hat{\Delta}$
- There is a natural and easy way to compute $s(\hat{\Delta}^*)$, since the blocks are chosen in an i.i.d. fashion

Details

Bootstrap world:

- Bootstrap data are generated using the circular block bootstrap
- This avoids the 'edge effects' of the moving blocks bootstrap, and so one can recenter simply using $\hat{\Delta}$
- There is a natural and easy way to compute $s(\hat{\Delta}^*)$, since the blocks are chosen in an i.i.d. fashion

Real world:

- Götze & Künsch (1996) propose the rectangular kernel for $s(\hat{\Delta})$ with bandwidth equal to the bootstrap block size
- However, we use the prewhitened QS kernel with a data-dependent choice of block size instead

More Details

Circular block bootstrap:

- Resample size-b blocks of observed excess return pairs
- Typical block: $\{(r_{ti}, r_{tn})', \ldots, (r_{(t+b-1)i}, r_{(t+b-1)n})'\}$
- Last block: $\{(r_{Ti}, r_{Tn})', (r_{1i}, r_{2n})' \ldots, (r_{(b-1)i}, r_{(b-1)n})'\}$

More Details

Circular block bootstrap:

- Resample size-b blocks of observed excess return pairs
- Typical block: $\{(r_{ti}, r_{tn})', \ldots, (r_{(t+b-1)i}, r_{(t+b-1)n})'\}$
- Last block: $\{(r_{Ti}, r_{Tn})', (r_{1i}, r_{2n})' \ldots, (r_{(b-1)i}, r_{(b-1)n})'\}$

Corresponding natural standard error, where $l = \lfloor n/b \rfloor$:

$$y_t^* = (r_{ti}^* - \hat{\mu}_i^*, r_{tn}^* - \hat{\mu}_n^*, r_{ti}^{*\,2} - \hat{\gamma}_i^*, r_{tn}^{*\,2} - \hat{\gamma}_n^*)' \quad t = 1, \ldots, T$$

$$\zeta_j = \frac{1}{\sqrt{b}} \sum_{t=1}^{b} y_{(j-1)b+t}^* \quad t = 1, \ldots, l$$

$$\hat{\Psi}^* = \frac{1}{l} \sum_{j=1}^{l} \zeta_j \zeta_j' \quad \text{and} \quad s(\hat{\Delta}^*) = \sqrt{\frac{\nabla' f(\hat{v}^*) \hat{\Psi}^* \nabla f(\hat{v}^*)}{T}}$$

Choice of the Block Size

Problem:

- A block bootstrap method uses a block size b
- This choice can be rather crucial for applications

Choice of the Block Size

Problem:

- A block bootstrap method uses a block size b
- This choice can be rather crucial for applications

Solution via a *Calibration Method*:

- Fit an unrestricted DGP to the observed data
- Suggestion: VAR model with bootstrapping the residuals (use the stationary bootstrap here)
- Simulating from this DGP, construct a CI for observed $\hat{\Delta}$
- Do this for a variety of candidate block sizes b
- Find the block size $\tilde{b}$ that yields simulated confidence level closest to the nominal level $1 - \alpha$

Two Remarks

Remark:

- Vinod & Morey (1999) discuss bootstrap inference for Δ
- However, they use the i.i.d. bootstrap of Efron
- Also, they studentize in the wrong way

Two Remarks

Remark:

- Vinod & Morey (1999) discuss bootstrap inference for Δ
- However, they use the i.i.d. bootstrap of Efron
- Also, they studentize in the wrong way

Remark:

- Scherer (2004) discusses bootstrap inference for Sh
- He does not use a studentized bootstrap
- He proposes a double bootstrap for i.i.d. data
- But not for time series data (where an AR model is used)

Outline

1. The Problem
2. Solutions
   - HAC Inference
   - Bootstrap Inference
3. Simulations
4. Empirical Applications
5. Conclusions

Set-Up

DGPs:

1. Bivariate i.i.d. with $\rho = 0.5$
2. Bivariate GARCH(1,1)
3. Bivariate VAR(1) with $\rho = 0.5$ and $\phi = 0.2$

- Innovations are either $N(0,1)$ oder standardized t_6
- $T = 120$

Set-Up

DGPs:

1. Bivariate i.i.d. with $\rho = 0.5$
2. Bivariate GARCH(1,1)
3. Bivariate VAR(1) with $\rho = 0.5$ and $\phi = 0.2$

- Innovations are either $N(0,1)$ oder standardized t_6
- $T = 120$

Methods:

- JK (1981) and Memmel (2003)
- HAC with QS kernel
- HAC with prewhitened QS kernel
- Studentized bootstrap

Set-Up

DGPs:

1. Bivariate i.i.d. with $\rho = 0.5$
2. Bivariate GARCH(1,1)
3. Bivariate VAR(1) with $\rho = 0.5$ and $\phi = 0.2$

- Innovations are either $N(0, 1)$ oder standardized t_6
- $T = 120$

Methods:

- JK (1981) and Memmel (2003)
- HAC with QS kernel
- HAC with prewhitened QS kernel
- Studentized bootstrap

Common goal:

- Test H_0: $\Delta = 0$ at nominal level $\alpha = 5\%$.

Results

Empirical rejection probabilities in percent:

DGP	JKM	HAC	HAC_{PW}	Boot-TS
Normal-IID	5.0	5.3	5.4	4.8
t_6-IID	10.7	6.7	6.9	5.0
Normal-GARCH	7.2	7.1	7.2	5.5
t_6-GARCH	7.4	7.7	7.5	5.7
Normal-VAR	9.5	6.9	6.1	5.0
t_6-VAR	14.5	7.9	7.3	5.1

Outline

Data

(1) Mutual funds:

- Fidelity, a 'large blend' fund
- Fidelity Aggressive Growth, a 'mid-cap growth' fund
- Data obtained from Yahoo! Finance

(2) Hedge funds:

- Coast Enhanced Income
- JMG Capital Partners
- Data obtained from CISDM database

In both cases:

- Return period: 01/94–12/03 $\Longrightarrow T = 120$
- Use log returns in excess of the risk-free rate
- Compute p-values for H_0: $\Delta = 0$

Summary Sample Statistics

Excess returns are in percentages;
sample statistics are not annualized:

- $\bar{r}$ = average excess return
- s = standard deviation of excess returns
- $\widehat{Sh}$ = Sharpe ratio
- $\hat{\phi}$ = first-oder autocorrelation

Fund	$\bar{r}$	s	$\widehat{Sh}$	$\hat{\phi}$
Fidelity	0.511	4.760	0.108	−0.010
Fidelity Agressive Growth	0.098	9.161	0.011	0.090
Coast Enhanced Income	0.245	0.168	1.461	0.152
JMG Capital Partners	1.228	1.211	1.014	0.435

p-Values

Data	JKM	HAC	HAC_{PW}	Boot
Mutual funds	3.9	6.3	6.7	10.2
Hedge funds	1.0	14.7	25.4	29.4

Outline

Conclusions

Findings:

- The test of JK (1981) and Memmel (2003) shouldn't be used
- HAC inference is consistent, but can be somewhat liberal in finite samples (of sizes typical in applications)
- Bootstrap inference is the most reliable one

Conclusions

Findings:

- The test of JK (1981) and Memmel (2003) shouldn't be used
- HAC inference is consistent, but can be somewhat liberal in finite samples (of sizes typical in applications)
- Bootstrap inference is the most reliable one

Furthermore:

- Free programming code for HAC and bootstrap inference is available on the website of the second author
- HAC and bootstrap inference can be modified for alternative performance measures, such as:
  - Refined Sharpe ratios
  - Jensen's alpha
  - Treynor ratio
  - Etc.

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858.

Andrews, D. W. K. and Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60:953–966.

Götze, F. and Künsch, H. R. (1996). Second order correctness of the blockwise bootstrap for stationary observations. *Annals of Statistics*, 24:1914–1933.

Jobson, J. D. and Korkie, B. M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance*, 36:889–908.

Lo, A. W. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal*, 58(4):36–52.

References (continued)

Memmel, C. (2003). Performance hypothesis testing with the Sharpe Ratio. *Finance Letters*, 1:21–23.

Opdyke, J. D. (2007). Comparing Sharpe ratios: So where are the *p*-values? *Journal of Asset Management*, 8(5):308–336.

Scherer, B. (2004). An alternative route to hypothesis testing. *Journal of Asset Management*, 5(1):5–12.

Vinod, H. D. and Morey, M. R. (1999). Confidence intervals and hypothesis testing for the Sharpe and Treynor performance measures: A bootstrap approach. In Abu-Mostafa, Y. S., LeBaron, B., Lo., A., and Weigend, A. S., editors, *Computational Finance 1999*, pages 25–39. The MIT Press, Cambridge.