

# Portfolio Theory: Assignment 1

## The Statistics of Strategy Back-Testing

Nesan Naidoo : NDXNES005

2025-09-15

### Problem Specification

This assignment examined the statistics of strategy backtesting within the context of portfolio theory. Part I focused on proving that the estimated annualized Sharpe ratio(SR) converges asymptotically to a normal distribution. Furthermore, part I motivated that for a sufficiently large number of samples, the mean of the sample maximum of standard normally distributed random variables can be approximated. Lastly, part I focused on the derivation and discussion of the minimum backtest length.

Part II focused on mean–variance backtesting of the tangency portfolio under full investment and no-short-selling constraints. Sharpe ratio–maximising portfolios are computed from rolling windows of the historical data. The historical dataset is divided into in-sample (IS) and out-of-sample (OOS) datasets. Two experiments are conducted: 1) compared IS and OOS Sharpe ratios and 2) evaluated OOS backtest performance using a rolling window approach.

### Data Specification

The Tactical Asset Allocation data is from PT-DATA-ALBI-JIBAR-JSEIND-Daily-1994-2017.xlsx

1. ICB Industrial Level Indices
2. ALBI (All Bond Index (ALBI) Total Return Index (TRI) Data)
3. Money Market Data: JIBAR and STEFI TRI

#### 4. Various Indices: JSE Growth, JSE Value, JSE ALSI, JSE SRI

## Configuration control

Version control: managed with Git and GitHub.

Project structure:

Click on link to repository: [github.com/NesanNaidoo/Portfolio-Theory-Assignment-1-Backtesting](https://github.com/NesanNaidoo/Portfolio-Theory-Assignment-1-Backtesting)

## PART I : Introduction to Strategy Backtesting

### Question 1 : Sample Error when Estimating the Sharpe Ratio

Proof below is based on the asymptotic distributions of Sharpe Ratio estimators (See Appendix A. IID Returns in (Lo 2002) )

#### 1.1 Assumptions

- We assume IID excess returns  $r_1, \dots, r_n$  with  $r_t \sim \mathcal{N}(\mu, \sigma^2)$ .
- Let  $q$  be returns per year,  $y$  the number of years, and  $n = qy$ .
- The true annualised Sharpe is  $SR = \sqrt{q} \frac{\mu}{\sigma}$ .
- Estimators used in proof:

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^n r_t, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n (r_t - \hat{\mu})^2, \quad \widehat{SR} = \sqrt{q} \frac{\hat{\mu}}{\hat{\sigma}}.$$

#### 1.2 Distribution of the sample mean

Since  $r_t \sim \mathcal{N}(\mu, \sigma^2)$  and using the fact the the sum of independent Normal variables are Normal. Therefore,

$$\sum_{t=1}^n r_t \sim N(n\mu, n\sigma^2). \tag{1}$$

Therefore, the sample mean  $\hat{\mu} = \frac{1}{n} \sum_{t=1}^n r_t$  follows a Normal distribution

$$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{n}\right). \quad (2)$$

Hence, by centering and scaling by  $\sqrt{n}$ , this leads to

$$\sqrt{n}(\hat{\mu} - \mu) \sim \mathcal{N}(0, \sigma^2). \quad (3)$$

### 1.3 Distribution of the sample variance

Given  $\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n (r_t - \hat{\mu})^2$ , then the result based on Normal theory is  $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-1}^2$ .

If  $U \sim \chi_k^2$  then  $\mathbb{E}[U] = k$  and  $\text{Var}(U) = 2k$ .

Hence, to standardise  $U$  we subtract its mean  $k$  and divide by its standard deviation  $\sqrt{2k}$ . Let  $k = n - 1$  and  $U = \frac{n\hat{\sigma}^2}{\sigma^2}$ . Using the fact that  $\chi_{n-1}^2$  is the sum of  $n - 1$  independent  $Z_i^2$  terms, where each  $Z_i^2$  has mean 1 and variance 2 and by the Central Limit Theorem, the centered and scaled sum converges to  $\mathcal{N}(0, 1)$  as  $n \rightarrow \infty$ . Therefore,

$$\frac{\frac{n\hat{\sigma}^2}{\sigma^2} - (n-1)}{\sqrt{2(n-1)}} \xrightarrow{d} \mathcal{N}(0, 1), \quad (n \rightarrow \infty). \quad (4)$$

Next the numerator and denominator is multiplied by  $\sigma^2$ :

$$\frac{n\hat{\sigma}^2 - (n-1)\sigma^2}{\sigma^2 \sqrt{2(n-1)}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (5)$$

Then the numerator can be rewritten as  $\mathcal{N}(\hat{\sigma}^2 - \sigma^2) + \sigma^2$  and fraction can be split into 2 terms:

$$\frac{\mathcal{N}(\hat{\sigma}^2 - \sigma^2)}{\sigma^2 \sqrt{2(n-1)}} + \frac{1}{\sqrt{2(n-1)}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (6)$$

Since  $\frac{1}{\sqrt{2(n-1)}} \rightarrow 0$  as  $n \rightarrow \infty$ , therefore

$$\frac{\mathcal{N}(\hat{\sigma}^2 - \sigma^2)}{\sigma^2 \sqrt{2(n-1)}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (7)$$

Let  $A_n = \sqrt{n}(\hat{\sigma}^2 - \sigma^2)$ . Therefore,

$$\frac{\sqrt{n} A_n}{\sigma^2 \sqrt{2(n-1)}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (8)$$

Hence  $A_n \xrightarrow{d} N\left(0, \frac{\sigma^4 2(n-1)}{n}\right)$ , and letting  $n \rightarrow \infty$  leads to

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} \mathcal{N}(0, 2\sigma^4). \quad (9)$$

#### 1.4 Using Delta method to obtain $\sqrt{n}(\hat{\sigma} - \sigma)$

Let  $h(x) = \sqrt{x}$ , so  $h'(\sigma^2) = 1/(2\sigma)$ . Then by Taylor expansion,

$$\hat{\sigma} - \sigma \approx h'(\sigma^2)(\hat{\sigma}^2 - \sigma^2) = 1/(2\sigma)(\hat{\sigma}^2 - \sigma^2).$$

By multiplying by  $\sqrt{n}$  :

$$\sqrt{n}(\hat{\sigma} - \sigma) \approx \frac{\sqrt{n}}{2\sigma}(\hat{\sigma}^2 - \sigma^2)$$

Using (9):

$$\sqrt{n}(\hat{\sigma} - \sigma) \xrightarrow{d} N\left(0, \left(\frac{1}{2\sigma}\right)^2 \cdot 2\sigma^4\right) = N\left(0, \frac{\sigma^2}{2}\right).$$

#### 1.5 Combining results so far

$$\sqrt{n} \begin{pmatrix} \hat{\mu} - \mu \\ \hat{\sigma} - \sigma \end{pmatrix} \xrightarrow{d} N(0, \Sigma) \quad \text{where} \quad \Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2/2 \end{pmatrix}.$$

#### 1.6 Using multivariate delta method

Let  $g(\mu, \sigma) = \sqrt{q} \mu / \sigma = SR$ . Then the Taylor expansion leads to

$$\sqrt{n}(g(\hat{\mu}, \hat{\sigma}) - g(\mu, \sigma)) = \nabla g(\mu, \sigma)^\top \sqrt{n} \begin{pmatrix} \hat{\mu} - \mu \\ \hat{\sigma} - \sigma \end{pmatrix} + o_p(1).$$

Thus by the Central Limit Theorem,

$$\sqrt{n}(\widehat{SR} - SR) \xrightarrow{d} N(0, \nabla g^\top \Sigma \nabla g).$$

## 1.7 Calculating the gradient and variance

Partial derivatives:

$$\frac{\partial g}{\partial \mu} = \frac{\sqrt{q}}{\sigma} \quad \text{and} \quad \frac{\partial g}{\partial \sigma} = -\frac{SR}{\sigma},$$

so  $\nabla g = (\sqrt{q}/\sigma, -SR/\sigma)^\top$ . Then

$$V = \nabla g^\top \Sigma \nabla g = \left(\frac{\sqrt{q}}{\sigma}\right)^2 \sigma^2 + \left(\frac{SR}{\sigma}\right)^2 \frac{\sigma^2}{2} = q + \frac{SR^2}{2}.$$

Hence,

$$\sqrt{n}(\widehat{SR} - SR) \xrightarrow{d} N\left(0, q + \frac{SR^2}{2}\right).$$

## 1.8 Converting to variance per year

Therefore,

$$\text{Var}(\widehat{SR}) \approx \frac{q + \frac{SR^2}{2}}{n}.$$

By dividing numerator and denominator by  $q$ :

$$\frac{q + \frac{SR^2}{2}}{n} = \frac{1 + \frac{SR^2}{2q}}{n/q} = \frac{1 + \frac{SR^2}{2q}}{y}.$$

Since  $n/q = y$ . Thus the final result is

$$\widehat{SR} \xrightarrow{d} N\left(SR, \frac{1 + \frac{SR^2}{2q}}{y}\right).$$

## Question 2 : The Maximum of the Sample

The proof below is based on Appendix: Proof of proposition 1 in (D. H. Bailey and López de Prado 2014). Foundations for this proof and normalisation constants  $(a_N, b_N)$  were based on (Embrechts, Klüppelberg, and Mikosch 1997) (pp.138–147) and Proposition 2.1(iii) in (Resnick 2008).

### Proof

#### Definitions

- Let  $X_1, \dots, X_N$  be IID  $\mathcal{N}(0, 1)$ .
- Define the sample maximum  $M_N := \max_{1 \leq n \leq N} X_n$ .

- Let  $Z(x) = \Phi(x)$  be the standard normal CDF,
- $\varphi(x)$  to be its PDF, and  $Z^{-1}$  the inverse CDF .
- Let  $\gamma = 0.5772156649\dots$  denote the Euler–Mascheroni constant.

## 2.1 Fisher–Tippett–Gnedenko theorem (Fisher and Tippett 1928)

The Fisher–Tippett–Gnedenko theorem states that for IID samples there exists normalising sequences  $a_N > 0$  and  $b_N \in \mathbb{R}$  such that the normalised maximum converges in distribution to a non-degenerate limit which is one of three types (Gumbel, Fréchet, Weibull). This is determined by the tail of the underlying distribution.

Because the normal distribution has exponentially decaying tails, it lies in the Gumbel domain . Hence there exist sequences  $a_N > 0$  and  $b_N \in \mathbb{R}$  such that

$$\frac{M_N - b_N}{a_N} \xrightarrow{d} G, \quad (1)$$

where  $G(x) = \exp(-e^{-x})$  is the standard Gumbel CDF.

## 2.2 Normalising constant (centering): $b_N$

A standard centering is the  $(1 - 1/N)$  quantile.

Define

$$b_N := Z^{-1}\left(1 - \frac{1}{N}\right). \quad (2)$$

Therefore, by construction

$$1 - Z(b_N) = \frac{1}{N}. \quad (3)$$

This suggests in  $N$  samples, the maximum generally lies near this quantile.

## 2.3 Normalising constant (scaling factor) : $a_N$

Define  $a_N$  to the difference between the  $(1 - 1/(Ne))$  and the  $(1 - 1/N)$  quantiles. Therefore,

$$a_N := Z^{-1}\left(1 - \frac{1}{Ne}\right) - Z^{-1}\left(1 - \frac{1}{N}\right). \quad (4)$$

This choice enforces

$$1 - Z(b_N + a_N) = \frac{1}{Ne}.$$

Hence, for a fixed  $x$  such that  $x \in \mathbb{R}$ , the tail at  $b_N + a_N x$  will be approximately  $\frac{1}{N}e^{-x}$ .

This leads to the Gumbel limit:

$$Z(b_N + a_N x)^N \longrightarrow \exp(-e^{-x}). \quad (5)$$

## 2.4 Convergence of the normalized maximum and expectation

From 2.3 and the Fisher–Tippett–Gnedenko theorem we obtain

$$\frac{M_N - b_N}{a_N} \xrightarrow{d} G.$$

Results for the Gumbel domain (see (Resnick 2008) Prop. 2.1(iii)) imply convergence of expectations under the integrability conditions satisfied here:

$$\lim_{N \rightarrow \infty} \mathbb{E}\left[\frac{M_N - b_N}{a_N}\right] = \mathbb{E}[G] = \gamma. \quad (6)$$

Hence for large  $N$ ,

$$\mathbb{E}[M_N] \approx b_N + \gamma a_N. \quad (7)$$

## 2.5 Final substitution

Substituting the definitions of  $b_N$  and  $a_N$  into  $\mathbb{E}[M_N] \approx b_N + \gamma a_N$  results in the following approximation:

$$\mathbb{E}[M_N] \approx Z^{-1}\left(1 - \frac{1}{N}\right) + \gamma \left( Z^{-1}\left(1 - \frac{1}{Ne}\right) - Z^{-1}\left(1 - \frac{1}{N}\right) \right).$$

Distributing  $\gamma$  over the parentheses,

$$\mathbb{E}[M_N] \approx Z^{-1}\left(1 - \frac{1}{N}\right) + \gamma Z^{-1}\left(1 - \frac{1}{Ne}\right) - \gamma Z^{-1}\left(1 - \frac{1}{N}\right).$$

Collecting the two terms that multiply  $Z^{-1}(1 - \frac{1}{N})$  leads to the final result,

$$\mathbb{E}[M_N] \approx [1 - \gamma] Z^{-1}\left(1 - \frac{1}{N}\right) + \gamma Z^{-1}\left(1 - \frac{1}{Ne}\right),$$

for some constant  $\gamma$  (Euler–Mascheroni constant)

### Question 3 : Minimum Backtest Length

The proof below is based on Theorem 2 in (D. H. Bailey and López de Prado 2014). Furthermore the proof uses results from the asymptotic distribution of Sharpe ratio estimators (see Question 1) and the theory of Normal maxima (see Question 2).

#### 3.1 Assumptions and initial setup

Using the result from Question 1 (section 1.8), the asymptotic distribution of the estimated annualised Sharpe ratio is

$$\widehat{SR} \xrightarrow{d} \mathcal{N}\left(SR, \frac{1 + \frac{SR^2}{2q}}{y}\right),$$

where  $q$  is returns per year,  $y$  the number of years and  $n = qy$ , and  $SR$  denotes the true annualised Sharpe ratio.

As we are testing  $N$  independent strategies, all with true Sharpe ratio  $SR = 0$  (out-of-sample Sharpe ratio = 0). Therefore, we substitute  $SR = 0$  into the mean and the variance:

- Mean: the asymptotic mean is  $SR$ . Hence,  $SR = 0$  so mean becomes 0.
- Variance:

$$\frac{1 + \frac{SR^2}{2q}}{y} \longrightarrow \frac{1 + \frac{0^2}{2q}}{y} = \frac{1}{y}.$$

Therefore, when the true Sharpe ratio is zero,

$$\widehat{SR} \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{y}\right).$$

We can also write this as,

$$\widehat{SR} = \frac{1}{\sqrt{y}} Z, \quad \text{where} \quad Z \sim \mathcal{N}(0, 1),$$

which will be used to derive the minimum backtest length.

#### 3.2 Maximum across $N$ strategies

Suppose  $N$  independent strategies are tested.

Let  $Z_1, \dots, Z_N \sim \mathcal{N}(0, 1)$  be iid, and we can define the sample maximum

$$M_N := \max_{1 \leq i \leq N} Z_i.$$



Then the maximum observed Sharpe ratio is

$$\widehat{SR}_{(N)} = \max_i \widehat{SR}_i = \frac{1}{\sqrt{y}} M_N. \quad (1)$$

### 3.3 Minimum backtest length

Taking expectations in (1):

$$\mathbb{E}[\widehat{SR}_{(N)}] = \frac{1}{\sqrt{y}} \mathbb{E}[M_N]. \quad (2)$$

Even if the true Sharpe ratio is zero, selecting the strategy with the highest in-sample Sharpe ratio introduces a positive bias which means  $\mathbb{E}[\widehat{SR}_{(N)}] > 0$ . Therefore, (2) can be rearranged to obtain an expression for  $y$  in terms of  $\mathbb{E}[M_N]$ :

$$y = \frac{(\mathbb{E}[M_N])^2}{(\mathbb{E}[\widehat{SR}_{(N)}])^2}. \quad (3)$$

Here,  $y$  is the backtest length. So by definition, the minimum backtest length  $T_{\min}$  is the value of  $y$  needed to ensure that the expected observed maximum Sharpe ratio  $\mathbb{E}[\widehat{SR}_{(N)}]$  does not exceed a target .

Hence we can write:

$$T_{\min} = y = \frac{(\mathbb{E}[M_N])^2}{(\mathbb{E}[\widehat{SR}_{(N)}])^2}.$$

To control for the bias mentioned earlier, the minimum backtest length  $y$  needs to be long enough so that the expected maximum observed Sharpe ratio is no greater than a chosen threshold.

From Question 2, we know that  $\mathbb{E}[M_N]$  grows with  $N$ , and we derived an explicit approximation:

$$\mathbb{E}[M_N] \approx [1 - \gamma] Z^{-1} \left( 1 - \frac{1}{N} \right) + \gamma Z^{-1} \left( 1 - \frac{1}{Ne} \right). \quad (4)$$

Also from Question 2, the expected maximum of  $N$  i.i.d. standard Normals grows logarithmically for large  $N$ . Therefore,

$$\mathbb{E}[M_N]^2 \lesssim 2 \ln N. \quad (5)$$

### 3.4 Final result

Substituting (5) into (3) gives the bound:

$$T_{\min} < \frac{2 \ln(N)}{(\mathbb{E}[\mathbb{E}[\widehat{SR}_{(N)}]])^2} = \frac{2 \ln(N)}{(\mathbb{E}[\widehat{SR}_{(N)}])^2}. \quad (6)$$

where the denominator  $(\mathbb{E}[\mathbb{E}[\widehat{SR}_{(N)}]])^2$  becomes  $(\mathbb{E}[\widehat{SR}_{(N)}])^2$  since the expectation of a constant is the constant.

Finally, (6) gives the minimum number of years needed to avoid overfitting from selecting the maximum in-sample Sharpe ratio among  $N$  strategies.

### 3.5 Discussion

As the number of strategies  $N$  increases, the minimum backtest length grows logarithmically with  $N$ . This shows that the minimum backtest length grows as the analyst tries more independent configurations of the model so as to keep the Sharpe Ratio at a given level.

## PART II : Backtest Performance of the Tangency Portfolio

### Experiment 1 : In-Sample and Out-Of-Sample Sharpe Ratios

### Experiment 2 : Out-Of-Sample Backtesting using a Rolling Window

## References

- Bailey, David H., and Marcos López de Prado. 2014. “The Deflated Sharpe Ratio:correcting for Selection Bias, BacktestOverfitting, and Non-Normality.” *The Journal of Portfolio Management* 40 (September): 94–107. <https://doi.org/10.3905/jpm.2014.40.5.094>.
- Bailey, David, Jonathan Borwein, Marcos López de Prado, and Qiji Jim Zhu. 2016. “The Probability of Backtest Overfitting.” *The Journal of Computational Finance*, September. <https://doi.org/10.21314/jcf.2016.322>.
- Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch. 1997. *Modelling Extremal Events for Insurance and Finance*. Springer.
- Fisher, Ronald Aylmer, and Leonard Henry Caleb Tippett. 1928. *Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample*. Vol. 24. 2. Cambridge University Press.
- Gnedenko, BV. 2006. *On Limit Theorems for a Random Number of Random Variables*. Springer.
- Lee, Wai. 2000. *Theory and Methodology of Tactical Asset Allocation*. John Wiley & Sons.
- Liu, Ying, Marie Rekkas, and Augustine Wong. 2012. “Inference for the Sharpe Ratio Using a Likelihood-Based Approach.” *Journal of Probability and Statistics* 2012: 1–24. <https://doi.org/10.1155/2012/878561>.
- Lo, Andrew W. 2002. “The Statistics of Sharpe Ratios.” *Financial Analysts Journal* 58 (July): 36–52. <https://doi.org/10.2469/faj.v58.n4.2453>.
- Resnick, Sidney I. 2008. *Extreme Values, Regular Variation and Point Processes*. Springer, Cop.

**Appendix A : Code**

**Appendix B : Session Information**