

# Portfolio Theory A1

## MV Backtesting and Out-of-Sample Performance

Christopher Eason (ESNCHR001)

2025-09-11

### Table of contents

<b>1</b>	<b>Part I: Introduction to Strategy Backtesting</b>	<b>3</b>
1.1	Question 1: Asymptotic distribution of the estimated annualized Sharpe Ratio	3
1.1.1	Definitions and Notation . . . . .	3
1.1.2	Central Limit Theorem . . . . .	3
1.1.3	Asymptotic Joint Distribution . . . . .	4
1.1.4	Delta Method . . . . .	5
1.1.5	Annualisation . . . . .	6
1.2	Question 2: Question 2: Expected Maximum of a Sample of IID Normal Variables . . . . .	7
1.2.1	Definition of the von Mises Function . . . . .	7
1.2.2	Showing that the Normal Distribution is a von Mises Function . . . .	8
1.2.3	Proposition 3.3.25: von Mises Functions and the Max-Domain of Attraction (MDA) . . . . .	10
1.2.4	Linking the von Mises–MDA Proposition to the Standard Normal Distribution . . . . .	10
1.2.5	Proposition 3.3.28: Closure Property of MDA under Tail Equivalence	10
1.2.6	Applying the closure proposition . . . . .	11
1.2.7	Calculation $d_N$ normalising constant . . . . .	11
1.2.8	Calculation $c_N$ normalising constant . . . . .	13
1.2.9	Proposition (iii): Convergence of Normalized Moments in the Gumbel MDA . . . . .	13
1.2.10	Application to the Standard Normal Maximum . . . . .	13
1.3	Question 3: Minimum Backtest Length to Avoid Overfitting . . . . .	15
1.3.1	Null Hypothesis . . . . .	15

---

1.3.2	Relating the Maximum Sharpe Ratio to the Maximum of Normal Random Variables . . . . .	15
1.3.3	Deriving the Minimum Backtest Length $T_{min}$ . . . . .	16
1.3.4	Discussion: Implications of the Minimum Backtest Length . . . . .	18
<b>References</b>		<b>19</b>

---

## Part I: Introduction to Strategy Backtesting

### Question 1: Asymptotic distribution of the estimated annualized Sharpe Ratio

Show that the distribution of the estimated annualised Sharpe Ratio (SR) converges asymptotically as  $y \rightarrow \infty$  to:

$$\hat{SR} \underset{y \rightarrow \infty}{\overset{a}{\sim}} N \left( SR, \frac{1 + \frac{SR^2}{2q}}{y} \right)$$

#### Definitions and Notation

- Let  $q$  be the number of return observations per year (e.g.  $q = 12$  for monthly)
- Let  $y$  be the number of years of data.
- Let  $T$  be the total number of observations such that  $T = qy$
- Let  $R_f$  be the risk-free rate
- Let  $R_t$  denote the one-period simple return of a portfolio or fund between the times  $t - 1$  and  $t$ . Assume  $R_t \sim N(\mu, \sigma^2)$ .
- Let  $\mu = E(R_t) - R_f$  be the mean of the excess returns and  $\sigma^2 = Cov(R_t)$  be the variance of the excess returns.
- Let SR be the annualised Sharpe Ratio that is defined as

$$SR = \frac{\mu}{\sigma} \sqrt{q} \quad (1)$$

- Since  $\mu$  and  $\sigma$  are the population movements of the distribution of  $R_t$  however they are unobservable and must be estimated using historical data. So given a sample of historical returns  $(R_1, R_2, \dots, R_T)$ , we let  $\hat{\mu} = \frac{1}{T} \sum_{t=1}^T R_t$  and  $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (R_t - \hat{\mu})^2$  be our estimates (Lo, 2002).
- Let  $\hat{SR}$  be the annualised estimate of the Sharpe Ratio that is defined as

$$\hat{SR} = \frac{\hat{\mu}}{\hat{\sigma}} \sqrt{q} \quad (2)$$

#### Central Limit Theorem

In order to derive the distribution of the estimated Sharpe ratio, we begin by assuming that the portfolio returns  $R_t$  are independently and identically distributed (IID). Practically, this means that the distribution of returns at one period is the same as at any other period and that returns are not correlated across time.

---

Under the IID assumption, and given that  $R_t \sim N(\mu, \sigma^2)$ , the sample mean  $\hat{\mu}$  and sample variance  $\hat{\sigma}^2$  of returns are sums of IID random variables. The Normality assumption is what allows us to use the properties of sums of independent Normal random variables and the  $\chi^2$  distribution to derive the variances of these estimates.

For the sample mean:  $\hat{\mu} = \frac{1}{T} \sum_{t=1}^T R_t$  the variance of a sum of  $T$  independent random variables is  $T\sigma^2$ , and dividing by  $T^2$  (because of the  $1/T$  factor in the mean) gives

$$\text{Var}(\hat{\mu}) = \frac{\sigma^2}{T} \quad (3)$$

For the sample variance:  $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (R_t - \hat{\mu})^2$  the CLT and the properties of the  $\chi^2$  distribution imply that

$$T \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_{T-1}^2.$$

The variance of a  $\chi^2$  with  $T - 1 \approx T$  degrees of freedom is  $2T$ , so rescaling back to  $\hat{\sigma}^2$  gives

$$\text{Var}(\hat{\sigma}^2) \approx \frac{2\sigma^4}{T} \quad (4)$$

Thus, taking Equation 3 and Equation 4 and considering the total number of observations  $T = qy$ , by the Central Limit Theorem, the distributions of  $\hat{\mu}$  and  $\hat{\sigma}^2$  converge asymptotically to Normal distributions. At this first stage, the CLT applies to sums of IID random variables, allowing us to get a joint asymptotic distribution for  $\hat{\mu}$  and  $\hat{\sigma}^2$  scaled by  $T$ .

$$\sqrt{T}(\hat{\mu} - \mu) \underset{T \rightarrow \infty}{\overset{a}{\rightsquigarrow}} N(0, \sigma^2), \quad \sqrt{T}(\hat{\sigma}^2 - \sigma^2) \underset{T \rightarrow \infty}{\overset{a}{\rightsquigarrow}} N(0, 2\sigma^4). \quad (5)$$

These asymptotic distributions allow us to approximate the estimation error of  $\hat{\mu}$  and  $\hat{\sigma}^2$ , and note that as  $T$  increases, both variances shrink toward zero. This reflects the intuitive fact that the larger the dataset (i.e., the more periods per year  $q$  and/or the more years  $y$ ), the smaller the uncertainty in our estimates.

### Asymptotic Joint Distribution

We can take Equation 5 and for an asymptotic joint distribution of  $\hat{\mu}$  and  $\hat{\sigma}^2$ .

$$\sqrt{T} \begin{bmatrix} \hat{\mu} - \mu \\ \hat{\sigma}^2 - \sigma^2 \end{bmatrix} \underset{T \rightarrow \infty}{\overset{a}{\rightsquigarrow}} \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix} \right) \quad (6)$$

## Delta Method

- Let  $\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{bmatrix}$  be a column vector
- Let  $\boldsymbol{\theta} = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}$  be a column vector
- Let  $\mathbf{V}_{\boldsymbol{\theta}} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix}$  be a matrix of joint covariance-variance matrix
- Let  $g(\mu, \sigma^2) = SR$  be a function that takes  $\mu$  and  $\sigma$  as parameters, and uses Equation 1. This means that  $g(\hat{\mu}, \hat{\sigma}^2) = \hat{S}\hat{R}$  be a function that takes  $\hat{\mu}$  and  $\hat{\sigma}$  as parameters, and uses Equation 2

We apply the delta method to propagate the uncertainty from the estimators  $\mu$  and  $\sigma^2$  through the nonlinear function  $g(\mu, \sigma^2) = SR$ . This allows us to derive the asymptotic distribution of the Sharpe ratio estimator  $\hat{S}\hat{R}$  using the gradient of  $g$  and the covariance matrix of  $\mu$  and  $\sigma^2$  (Lo, 2002).

First, we can re-write Equation 6 as

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \underset{T \rightarrow \infty}{\overset{a}{\rightsquigarrow}} \text{N}(0, \mathbf{V}_{\boldsymbol{\theta}})$$

Employing the delta method:

$$\sqrt{T}(g(\hat{\boldsymbol{\theta}}) - g(\boldsymbol{\theta})) \underset{T \rightarrow \infty}{\overset{a}{\rightsquigarrow}} \text{N}\left(0, \left(\frac{\partial g}{\partial \boldsymbol{\theta}}\right)' \mathbf{V}_{\boldsymbol{\theta}} \frac{\partial g}{\partial \boldsymbol{\theta}}\right) \quad (7)$$

Looking at just the variance term we can compute the gradient:

$$\frac{\partial g}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial g}{\partial \mu} \\ \frac{\partial g}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial \mu} \frac{\mu}{\sigma} \sqrt{q} \\ \frac{\partial}{\partial \sigma^2} \frac{\mu}{\sigma} \sqrt{q} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial \mu} \frac{\mu}{\sigma} \sqrt{q} \\ \frac{\partial}{\partial \sigma^2} \frac{\mu}{\sqrt{\sigma^2}} \sqrt{q} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{q}}{-2\mu(\sigma^2)^{-\frac{3}{2}}} \sqrt{q} \\ -\frac{\mu}{2\sigma^3} \sqrt{q} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{q}}{-2\sigma^3} \sqrt{q} \\ -\frac{\mu}{2\sigma^3} \sqrt{q} \end{bmatrix}.$$

Taking this partial derivative and calculating the variance term for Equation 7.

$$\begin{aligned} \left(\frac{\partial g}{\partial \boldsymbol{\theta}}\right)' \mathbf{V}_{\boldsymbol{\theta}} \frac{\partial g}{\partial \boldsymbol{\theta}} &= \begin{bmatrix} \frac{\sqrt{q}}{\sigma} & -\frac{\mu}{2\sigma^3} \sqrt{q} \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{q}}{-2\sigma^3} \sqrt{q} \\ -\frac{\mu}{2\sigma^3} \sqrt{q} \end{bmatrix} = [\sqrt{q}\sigma \quad -\mu\sigma\sqrt{q}] \begin{bmatrix} \frac{\sqrt{q}}{-2\sigma^3} \sqrt{q} \\ -\frac{\mu}{2\sigma^3} \sqrt{q} \end{bmatrix} \\ \left(\frac{\partial g}{\partial \boldsymbol{\theta}}\right)' \mathbf{V}_{\boldsymbol{\theta}} \frac{\partial g}{\partial \boldsymbol{\theta}} &= q + \frac{q\mu^2}{2\sigma^2} = q + \frac{SR^2}{2} = q \left(1 + \frac{SR^2}{2q}\right) \end{aligned}$$

---

Using Equation 1 we see that  $SR = \frac{\mu}{\sigma}\sqrt{q}$  so  $SR^2 = \frac{\mu^2}{\sigma^2}q$ . We can substitute this back into Equation 7 and find the distribution for  $\hat{SR}$ , remember that  $g(\hat{\theta}) = \hat{SR}$  and  $g(\theta) = SR$

$$\sqrt{T}(\hat{SR} - SR) \underset{T \rightarrow \infty}{\overset{a}{\sim}} N\left(0, q\left(1 + \frac{SR^2}{2q}\right)\right)$$

$$\hat{SR} \underset{T \rightarrow \infty}{\overset{a}{\sim}} N\left(SR, \frac{q\left(1 + \frac{SR^2}{2q}\right)}{T}\right)$$

### Annualisation

Since  $T = yq$ , we express the asymptotic variance per year by switching the limiting argument from  $T \rightarrow \infty$  to  $y \rightarrow \infty$  to reflect the annualized Sharpe ratio. Writing the variance in terms of years makes it explicit that the uncertainty in the estimate decreases as the number of years of data grows, which is the meaningful timescale for investors.

$$\hat{SR} \underset{y \rightarrow \infty}{\overset{a}{\sim}} N\left(SR, \frac{q\left(1 + \frac{SR^2}{2q}\right)}{qy}\right)$$

$$\hat{SR} \underset{y \rightarrow \infty}{\overset{a}{\sim}} N\left(SR, \frac{1 + \frac{SR^2}{2q}}{y}\right) \tag{8}$$

The final asymptotic variance  $\frac{1 + \frac{SR^2}{2q}}{y}$  shows two effects: (1) the variance shrinks with more years of data, and (2) higher Sharpe ratios increase estimation error slightly due to their dependence on both  $\mu$  and  $\sigma^2$ .

## Question 2: Question 2: Expected Maximum of a Sample of IID Normal Variables

Motivate and justify the following approximation for large  $N$ :

**Theorem 1.1.** Given a sample of  $N$  IID Normal random variables  $X_n, n = 1, 2, \dots, N$ , where  $Z$  is the CDF of the standard normal distribution, the expected maximum of the sample is:

$$E[\max_N] := E[\max\{X_n\}].$$

The expected maximum can be approximated as:

$$E[\max_N] \approx (1 - \gamma)Z^{-1}\left(1 - \frac{1}{N}\right) + \gamma Z^{-1}\left(1 - \frac{1}{N}e^{-1}\right)$$

for some constant  $\gamma$

To approximate the expected maximum of  $N$  i.i.d. Normal random variables, we proceed in three steps.

Step 1: Show the Normal is von Mises. Using Example 3.3.29 (Embrecchts et al., 1997), we first verify that the standard Normal distribution is a von Mises function with auxiliary function  $a(x)$ .

Step 2: Connect to the Gumbel MDA. By Proposition 3.3.25 (Embrecchts et al., 1997), any von Mises function belongs to the maximum domain of attraction of the Gumbel distribution ( $\text{MDA}(\Lambda)$ ). Moreover, Proposition 3.3.28 (Embrecchts et al., 1997) shows that if two distributions are tail equivalent, they share the same MDA and norming constants. Together, these results guarantee that the maxima of a Normal sample, once properly normalized, converge in distribution to the Gumbel law, which is the Gumbel case of the Fisher–Tippett–Gnedenko theorem.

Step 3: Convergence of moments. We apply Resnick's Proposition (iii) on moment convergence, we obtain that the expectation of the normalised maximum converges to the Euler–Mascheroni constant  $\gamma$  (Resnick, 1987). Together, these results yield the approximation  $\mathbb{E}(x) \approx \alpha + \gamma\beta$  with  $\alpha, \beta$  being norming constants derived from the Normal distribution.

### Definition of the von Mises Function

Let  $F$  be a cumulative distribution function (CDF) with right endpoint  $x_F$  is the largest possible value that the random variable  $X$  can take (if it exists) or  $+\infty$  if  $X$  is unbounded.

$$x_F = \sup\{x \in \mathbb{R} : F(x) < 1\} \in (-\infty, \infty]$$

We denote the survival function by

$$\bar{F}(x) = 1 - F(x)$$

We say  $F$  is a von Mises Function if there exists a scalar  $z < x_F$  and functions  $a(x)$ ,  $c(x)$  satisfying the following conditions:

- 
- $a : (z, x_f) \rightarrow (0, \infty)$  is a positive, absolutely continuous function (called the auxiliary function). It is a positive function that controls the rate of decay of the tail of F.
  - $c : (z, x_F) \rightarrow (0, \infty)$  is a positive function such that  $\lim_{x \rightarrow x_f} c(x) = c > 0$  i.e. This means that as  $x$  gets arbitrarily close to  $x_F$  from below, the function  $c(x)$  approaches a finite positive constant  $c$ . It serves as a normalizing factor to make the representation exact.

Then for all  $x \in (z, x_F)$  the survival function admits the representation

$$\bar{F}(x) = c(x) \exp \left( - \int_z^x \frac{1}{a(t)} dt \right) \quad z < x < x_F \quad (9)$$

### Showing that the Normal Distribution is a von Mises Function

Let  $X \sim N(0, 1)$  with the cumulative distribution function  $Z(x)$  and the probability density function  $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ . Denote the survival function (tail) by

$$\bar{Z}(x) = 1 - Z(x)$$

We are going to check the von Mises conditions to show that  $\bar{F}(x) = \bar{Z}(x)$ .

- The standard normal is unbounded above, so  $x_F = +\infty$ . We can choose arbitrary  $z = 0$  where  $z < x_F$ .
- We can define the auxilliary function as

$$a(x) = \frac{\bar{Z}(x)}{\phi(x)}$$

where  $\phi(x) > 0$  and  $\bar{Z}(x) > 0$  for all  $x$ , so  $a(x) > 0$ .

- Defining the normalising factor  $c(x)$ :
  - Using the von Mises representation, it suggests we can find

$$c(x) = \bar{Z}(x) \exp \left( - \int_z^x \frac{1}{a(t)} dt \right)$$

- To obtain the asymptotic form of the standard normal tail we apply L'Hôpital's rule, this will give us Mills' Ratio. Firstly we check that the limits of the numerator and denominator of our proposed ratio

$$\lim_{x \rightarrow \infty} \frac{\bar{Z}(x)}{\phi(x)} \rightarrow 0$$



Secondly, we find the first derivatives for the numerator and the denominator

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{\bar{Z}(x)}{\frac{\phi(x)}{x}} &= \lim_{x \rightarrow \infty} \frac{\frac{d}{dx}(\bar{Z}(x))}{\frac{d}{dx} \frac{\phi(x)}{x}} = \lim_{x \rightarrow \infty} \frac{\frac{d}{dx}(1 - Z(x))}{\frac{\phi'(x)x - \phi(x)}{x^2}} = \lim_{x \rightarrow \infty} \frac{-Z'(x)}{\frac{-x\phi(x) \cdot x - \phi(x)}{x^2}} = \\ &= \lim_{x \rightarrow \infty} \frac{-\phi(x)}{-\phi(x) \left(1 + \frac{1}{x^2}\right)} = \lim_{x \rightarrow \infty} \frac{-\phi(x)}{-\phi(x)} = 1 \end{aligned}$$

Therefore, we can use the following rates of change form instead of the raw values (i.e. Mills' Ratio)

$$\bar{Z}(x) \sim \frac{\phi(x)}{x} \quad x \rightarrow \infty \quad (10)$$

- We can apply Mills' Ratio  $\bar{Z}(x) \sim \frac{\phi(x)}{x}$  as  $x \rightarrow \infty$  to our auxiliary function becomes

$$a(x) = \frac{\bar{Z}(x)}{\phi(x)} = \frac{\frac{\phi(x)}{x}}{\phi(x)} \sim \frac{1}{x} \quad (11)$$

- Therefore solving for  $c(x)$

$$c(x) \sim \bar{Z}(x) \exp \left( \int_z^x \frac{1}{a(t)} dt \right) \sim \bar{Z}(x) \exp \left( \int_z^x \frac{1}{\frac{1}{x}} dt \right) \sim \frac{\phi(x)}{x} \cdot \exp \left( \frac{x^2}{2} \right).$$

$$c(x) \sim \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right) \frac{1}{x} \exp \left( \frac{x^2}{2} \right) \sim \frac{1}{x\sqrt{2\pi}}$$

Note that  $c(x) > 0$  can go to zero as  $x \rightarrow \infty$  for unbounded distributions like the normal, however as long as it varies slower than the exponential decay i.e. the exponential term on the right-hand side of Equation 9.

Therefore, with  $a(x) \sim \frac{1}{x}$  and  $c(x) \sim \frac{1}{x\sqrt{2\pi}}$ , using Equation 9 we have

$$\bar{Z}(x) \sim c(x) \exp \left( - \int_z^x \frac{1}{a(t)} dt \right) \sim \frac{1}{x\sqrt{2\pi}} \exp \left( - \frac{x^2}{2} \right)$$

which is exactly the standard tail approximation (Mills' ratio Equation 10 i.e.  $\bar{Z}(x) \sim \frac{\phi(x)}{x} \sim$

$\frac{\frac{1}{x\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right)}{x}$ ) for the standard normal distribution and we further show that  $c(x) \sim \frac{1}{x\sqrt{2\pi}}$

decays much slower than the exponential decay i.e.  $\exp \left( -\frac{x^2}{2} \right)$  as  $x \rightarrow \infty$ . Since the standard normal distribution satisfies all the von Mises conditions we conclude that the standard normal distribution is a von Mises function.

---

**Proposition 3.3.25: von Mises Functions and the Max-Domain of Attraction (MDA)**

Suppose the distribution function  $F$  is a von Mises function with auxiliary function  $a(x)$ . Then  $F$  belongs to the max-domain of attraction (MDA) of the Gumbel distribution.

- Max-domain of attraction (MDA): This means that if  $X_1, X_2, \dots, X_N$ , where  $n = 1, 2, \dots, N$ , are IID random variables with distribution  $F$ , then the properly normalized maximum

$$M_N := \max\{X_1, X_2, \dots, X_N\} \underset{N \rightarrow \infty}{\overset{a}{\sim}} G \quad (12)$$

Where  $G$  is the Gumbel distribution i.e.  $F \in \text{MDA}(\Lambda)$

- A possible choice of norming constants for continuous and strictly increasing distribution function like the normal distribution is:

$$d_N := F^{-1}\left(1 - \frac{1}{N}\right) \quad c_N := a(d_N) \quad (13)$$

where  $a(x)$  is the auxiliary function of  $F$ . Then the normalized maximum converges to

$$\frac{M_N - d_N}{c_N} \underset{N \rightarrow \infty}{\overset{d}{\rightarrow}} G \quad (14)$$

where  $G$  is the standard Gumbel distribution with cdf  $G(x) = \exp(-e^{-x})$

**Linking the von Mises–MDA Proposition to the Standard Normal Distribution**

We now apply Proposition 3.3.25 to the standard normal distribution  $X \sim N(0, 1)$ . Since we have already shown from Equation 11 that  $Z(x)$  is a von Mises function with auxiliary function  $a(x) \sim \frac{1}{x}$  as  $x \rightarrow \infty$ , the standard normal belongs to the max-domain of attraction (MDA) of the Gumbel distribution i.e.  $Z(x) \in \text{MDA}(\Lambda)$  Equation 12.

**Proposition 3.3.28: Closure Property of MDA under Tail Equivalence**

Let  $F$  and  $H$  be two distribution functions with the same right endpoint  $x_F = x_H$ . Suppose  $F$  belongs to the max-domain of attraction (MDA)  $F \in \text{MDA}(\Lambda)$  with norming constants  $(c_N > 0, d_N \in \mathbb{R})$ .

Then  $G$  also belongs to the same MDA with the same norming constants  $(c_N, d_N)$ , i.e.

$$\lim_{n \rightarrow \infty} F^n(c_N x + d_N) = \Lambda(x) \quad x \in \mathbb{R}$$

then

$$\lim_{N \rightarrow \infty} H^N(c_N x + d_N) = \Lambda(x + b) \quad x \in \mathbb{R}$$

if and only if  $F$  and  $H$  are tail equivalent with

---


$$\lim_{x \rightarrow x_F} \frac{\bar{F}(x)}{\bar{H}(x)} = e^b \quad (15)$$

for some finite constant  $b \in \mathbb{R}$

### Applying the closure proposition

Now that we have established the standard normal belongs to the max-domain of attraction of the Gumbel Distribution, we need to compute the norming constants  $d_N$  and  $c_N$  to properly normalise the maximum  $M_N$  i.e. Equation 14 establishing the Gumbel limiting distribution for the standard normal maximum.

First we acknowledge the usage of the L'Hôpital's rule to give us Equation 10 showing that  $\bar{Z}(x) \sim \frac{\phi(x)}{x}$  as  $x \rightarrow \infty$ . We apply Proposition 3.3.28 to justify replacing the standard normal tail  $\bar{Z}(x) = 1 - Z(x)$ , a more complicated function, with the simpler Mills' ratio  $\bar{H}(x) = \frac{\pi(x)}{x}$ , which is tail equivalent. This simplifies the calculation of the norming constants  $d_N$  and  $c_N$  for the maximum.

So taking  $\bar{Z}(x) \sim \frac{\phi(x)}{x}$  as  $x \rightarrow \infty$  and denoting  $\bar{H}(x) := \frac{\phi(x)}{x}$ , we can see that  $\bar{Z}$  and  $\bar{H}$  are tail equivalent with

$$\lim_{x \rightarrow \infty} \frac{\bar{Z}(x)}{\bar{H}(x)} = \lim_{x \rightarrow \infty} \frac{\frac{\phi(x)}{x}}{\frac{\phi(x)}{x}} = e^0 = 1$$

where  $b = 0$ , therefore showing Equation 15 and allowing us to use the norming constants computed for  $\bar{H}$  directly for  $\bar{Z}$  i.e.

$$\bar{Z}(x) \sim \bar{H}(x) \quad (16)$$

where  $x \rightarrow \infty$ . This simplifies the calculation for  $d_N$  and  $c_N$  for the standard normal maximum and ensures that the asymptotic Gumbel approximation holds.

### Calculation $d_N$ normalising constant

From Equation 13, we choose  $d_N = H^{-1}\left(1 - \frac{1}{N}\right)$ . Intuitively,  $d_n$  represents the level such that the probability of exceeding it is  $\frac{1}{N}$ , i.e., the level of the expected maximum in a sample of size  $N$ . Equivalently, we can express this using the survival function  $\bar{H}(x) := 1 - H(x)$  which gives us gives the probability of exceeding a value  $x$ :

$$\bar{H}(d_N) = 1 - H(d_N) = \frac{1}{N}$$

Taking the negative logarithm of both sides, we obtain

---


$$-\ln \bar{H}(d_N) = -\ln\left(\frac{1}{N}\right) = \ln N$$

This step shows explicitly how the survival function transforms the original inverse CDF condition into a logarithmic equation, which is convenient for solving  $d_N$  asymptotically.

For the tail of the normal distribution,  $\bar{H}(x) = \frac{\phi(x)}{x} = \frac{1}{\sqrt{2\pi}x} e^{-\frac{x^2}{2}}$ , so taking the logarithm and using its properties gives

$$-\ln \bar{H}(d_N) = \frac{1}{\sqrt{2\pi}d_N} e^{-\frac{d_N^2}{2}} = -\left[-\frac{d_N^2}{2} - \ln(\sqrt{2\pi}d_N)\right]$$

Finally, we can split the log to obtain

$$-\ln \bar{H}(d_N) = \frac{1}{2}d_N^2 + \ln(d_N) + \frac{1}{2}\ln 2\pi$$

Now we can solve for  $d_N$ , but since this equation is non-linear, we'll find the asymptotic solution for large  $N$  using a Taylor expansion. The leading-order term

$$\frac{1}{2}d_N^2 \approx \ln N \quad \Rightarrow \quad d_N \sim \sqrt{2 \ln N}$$

Including the next-order correction from  $\ln(d_N) + \frac{1}{2}\ln(2\pi)$ , we expand the equation and solve for  $d_N$  asymptotically:

$$d_N \approx \sqrt{2 \ln N} - \frac{\ln(\sqrt{2 \ln N}) + \frac{1}{2}\ln(2\pi)}{\sqrt{2 \ln N}}$$

Simplifying the logarithms yields the refined expansion:

$$d_N \approx \sqrt{2 \ln N} - \frac{\ln(\ln N) + \ln(4\pi)}{2\sqrt{2 \ln N}} + O((\ln N)^{-1/2})$$

where the  $O((\ln N)^{-1/2})$  term represents higher-order terms in the asymptotic expansion, i.e., terms that are smaller than the retained correction and vanish relative to the main terms as  $N \rightarrow \infty$ .

Since these higher-order contributions are negligible for large  $N$ , we often drop them, leaving the practical approximation:

$$d_N \approx \sqrt{2 \ln N} - \frac{\ln(\ln N) + \ln(4\pi)}{2\sqrt{2 \ln N}} \tag{17}$$

This shows explicitly that the leading-order term  $\sqrt{2\ln N}$  dominates, the next-order logarithmic correction refines the approximation, and all remaining terms fall away asymptotically.

### Calculation $c_N$ normalising constant

From Equation 13, we choose  $c_N = a(d_N)$ . Since  $a(x) \sim \frac{1}{x}$  as  $x \rightarrow \infty$  from Equation 11, we can use the leading order term for  $d_N \sim \sqrt{2\ln N}$  we therefore get

$$c_N = a(d_N) \sim \frac{1}{d_N} \sim \frac{1}{\sqrt{2\ln N}} = (2\ln N)^{-\frac{1}{2}} \quad (18)$$

### Proposition (iii): Convergence of Normalized Moments in the Gumbel MDA

Let  $X_1, X_2 \dots X_N$  be i.i.d random variables with common distribution function  $H$  with a right endpoint ( $x_F$ ). Let

$$M_N := \max\{X_1, X_2, \dots X_N\} \underset{N \rightarrow \infty}{\overset{a}{\sim}}$$

denote the sample maximum. Define the norming constants

$$d_N = H^{-1}\left(1 - \frac{1}{N}\right), \quad c_N = a(d_N),$$

where  $a(X)$  is the auxiliary function of  $H$ . If for some integer ( $k > 0$ )

$$\int_{-\infty}^{x_H} |x|^k H(dx) < \infty,$$

then

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[ \left( \frac{M_N - d_N}{c_N} \right)^k \right] = \int_{-\infty}^{\infty} x^k \Lambda(dx) = (-1)^k \Gamma^{(k)}(1), \quad (19)$$

where  $\Gamma^{(k)}(1) = \gamma \approx 0.5772$  is the  $k$ -th derivative of the Gamma function evaluated at  $x = 1$  which is Euler–Mascheroni constant.

### Application to the Standard Normal Maximum

We can use Equation 14 and Equation 19 to show the approximation to the normalized maximum of the standard normal

$$\frac{M_N - d_N}{c_N} \xrightarrow[n \rightarrow \infty]{d} G$$

Using Proposition (iii) on moments Equation 19 to get the Gumbel limit where  $k=1$

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[ \left( \frac{M_N - d_N}{c_N} \right)^1 \right] \approx \int_{-\infty}^{\infty} x^1 \Lambda(dx) \approx \Gamma^{(1)}(1) \approx \gamma$$

Solving for the  $\mathbb{E}[M_N]$  we get the following

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[ \left( \frac{M_N - d_N}{c_N} \right)^1 \right] \approx \gamma \Rightarrow \mathbb{E}[M_N] \approx d_N + c_N \gamma \approx (1 - \gamma)d_N + \gamma(d_N + c_N)$$

Remembering that  $d_N = H^{-1} \left( 1 - \frac{1}{N} \right) = Z^{-1} \left( 1 - \frac{1}{N} \right)$  from Equation 13 and using the result of the closure property Equation 16. Also from Equation 13  $c_N = a(d_N) = \frac{1}{Z^{-1} \left( 1 - \frac{1}{N} \right)}$

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

Equivalently,

$$\boxed{\mathbb{E}[M_N] \approx (1 - \gamma)Z^{-1} \left( 1 - \frac{1}{N} \right) + \gamma Z^{-1} \left( 1 - \frac{1}{N} e^{-1} \right)} \quad (20)$$

The left-hand term is equivalent to the result Equation 17 i.e.  $Z^{-1} \left( 1 - \frac{1}{N} \right) = d_N$ , this tells us the size of the maximum. The right-hand term is equivalent to Equation 17 + Equation 18 i.e.  $Z^{-1} \left( 1 - \frac{1}{N} e^{-1} \right) \sim d_N + c_N$ , this gives the appropriate scaling of the maximum so that when we normalize  $M_N$  it converges in distribution to a standard Gumbel. Calculating  $d_N$  and  $c_N$  is crucial because it grounds the general limit theorem in the specific case of the Normal distribution and makes the result practically useful. These constants let us explicitly approximate  $\mathbb{E}[M_N]$ , turning a purely theoretical Gumbel convergence (as guaranteed by the Fisher–Tippett–Gnedenko theorem) into a concrete formula that quantifies both the location of the maximum and the scale of its fluctuations

### Question 3: Minimum Backtest Length to Avoid Overfitting

Derive and discuss the minimum back test length  $T_{min}$  :

**Theorem 1.2.** The minimum back test length  $T_{min}$  needed to avoid selecting a strategy with an in-sample Sharpe Ratio as the average  $\mathbb{E}[max_N]$  among  $N$  independent strategies with an out-of-sample Sharpe Ratio of zero is:

$$T_{min} < \frac{2\ln(N)}{\mathbb{E}[\mathbb{E}[max_N]]^2}$$

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. The key point here: the analyst that does not report the number of simulations used to select a particular strategy configuration makes it very difficult to assess the overall risk of strategy overfitting

#### Null Hypothesis

The goal of Theorem 1.2 is to determine the minimum backtest length required to avoid selecting a strategy that only appears profitable due to chance. In practice, researchers test many candidate strategies and naturally choose the one with the highest in-sample Sharpe ratio. If none of the strategies has genuine skill, this maximum Sharpe ratio arises purely from noise.

To capture this scenario, we impose the null hypothesis:

$$H_0 : SR = 0$$

#### Relating the Maximum Sharpe Ratio to the Maximum of Normal Random Variables

From Question 1 result Equation 8 we know that the annualised Sharpe ratio estimator satisfies:

$$\hat{SR} \underset{y \rightarrow \infty}{\overset{a}{\sim}} N\left(SR, \frac{1 + \frac{SR^2}{2q}}{y}\right)$$

Under  $H_0 : SR = 0$ , this simplifies to

$$\hat{SR} \underset{y \rightarrow \infty}{\overset{a}{\sim}} N\left(0, \frac{1 + \frac{0}{2q}}{y}\right) \underset{y \rightarrow \infty}{\overset{a}{\sim}} N\left(0, \frac{1}{y}\right)$$

Thus any nonzero in-sample Sharpe ratio is entirely due to sampling variability. We can then rescale  $\hat{SR}$  by  $\sqrt{y}$ :

$$Z := \sqrt{y}\hat{SR} \sim N(0, 1) \tag{21}$$

Thus, each estimated Sharpe ratio under the null can be represented as a standard normal variable scaled by  $\frac{1}{\sqrt{y}}$ .

Now suppose we evaluate  $N$  independent strategies under the null, each estimated Sharpe ratio has the form

$$\hat{S}R_i = \frac{Z_i}{\sqrt{y}} \quad Z_i \sim N(0, 1)$$

so the in-sample maximum Sharpe ratio across strategies the  $N$  strategies can be written as

$$\max_{1 \leq i \leq N} \hat{S}R = \frac{1}{\sqrt{y}} \max_{1 \leq i \leq N} Z_i$$

This expression makes explicit how the problem of evaluating the maximum estimated Sharpe ratio reduces to the classical problem of the maximum of  $N$  standard normal random variables, already solved in Question 2. Therefore, the expected maximum Sharpe ratio under the null can be written to highlight the two layers of expectation:

$$\mathbb{E} \left[ \underbrace{\max_i \hat{S}R_i}_{\substack{\text{outer expectation over} \\ \text{Sharpe samples}}} \right] = \frac{1}{\sqrt{y}} \mathbb{E} \left[ \underbrace{\max_i Z_i}_{\substack{\text{inner expectation from extreme-value} \\ \text{distribution of standard normals } Z_i}} \right] = \frac{1}{\sqrt{y}} \mathbb{E}[\mathbb{E}[M_N]] \quad (22)$$

where  $M_N := \max_{1 \leq i \leq N} Z_i$  and  $Z_i \sim N(0, 1)$ . Note the following:

- The inner expectation  $\mathbb{E}[M_N]$  comes from the extreme-value approximation for the maximum of  $N$  IID standard normals, using the normalising constants  $d_N$  at Equation 17 and  $c_N$  at Equation 18 together with the approximation for  $\mathbb{E}[M_N]$  at Equation 20.
- The outer expectation accounts for the sampling variability of  $\hat{S}R$  itself, which under the null scales like  $\frac{1}{\sqrt{y}}$

This bridge between the two proofs is crucial, because it lets us transfer the extreme-value asymptotics from Question 2 into the analysis of the expected in-sample Sharpe ratio across  $N$  strategies.

### Deriving the Minimum Backtest Length $T_{min}$

The next step is to use the relationship between the expected maximum Sharpe ratio and the sample length to determine the minimum backtest length needed to avoid selecting a purely lucky strategy. We can rearrange Equation 22 to solve for the number of years of data  $y$  which gives



$$y = \frac{\mathbb{E}[\mathbb{E}[M_N]]}{(\mathbb{E}[\max_i \hat{S}R_i])^2}$$

To obtain a conservative bound for the minimum backtest length  $T_{min}$  we make 3 key substitutions:

1. Substitute  $y$  with  $T_{min}$  because we are solving for the minimum number of years of data required such that the expected maximum Sharpe ratio is not artificially large under the null hypothesis due to random luck.
2. Substitute the numerator  $\mathbb{E}[\mathbb{E}[M_N]]$  with the upper bound  $\sqrt{2\ln N}$  because using our final result from Question 2 Equation 14, we see that

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

$$\mathbb{E}[M_N] \approx (1 - \gamma)d_N + \gamma(d_N + c_N) \approx d_N + \gamma c_N$$

Both the  $d_N$  and  $d_N + c_N$  have the same dominant, leading terms of  $\sqrt{2\ln N}$  but with slightly different correction terms, which vanish asymptotically compared to the dominant term. Putting this together

$$Z^{-1}\left(1 - \frac{1}{N}\right) \sim Z^{-1}\left(1 - \frac{1}{N}e^{-1}\right) \sim \sqrt{2\ln N}$$

When substituting  $\sqrt{2\ln N}$  into the numerator of the  $T_{min}$  formula, we want an effective asymptotic ceiling that the expected maximum cannot grow faster than.

$$\mathbb{E}[M_N] \lesssim \sqrt{2\ln N} \Rightarrow \mathbb{E}[\mathbb{E}[M_N]] \lesssim \sqrt{2\ln N}$$

The  $\lesssim$  turns into a strict  $<$  in practice because using the ceiling slightly overestimates the true expectation, meaning the actual required backtest length is strictly less than the bound computed. This turns the equality into an inequality of:

$$T_{min} < \frac{2 \ln N}{(\mathbb{E}[\max_i \hat{S}R_i])^2}.$$

3. Substitute the denominator  $\mathbb{E}[\max_i \hat{S}R_i]$  with  $\mathbb{E}[\mathbb{E}[M_N]]$  because under the null  $H_0$  :  $SR = 0$ , the observed in-sample Sharpe ratio that drives selection bias is precisely given by the scaled expected maximum across the  $N$  strategies shown by Equation 22 i.e.

$$T_{min} < \frac{2 \ln N}{\left(\frac{1}{\sqrt{y}} \mathbb{E}[\mathbb{E}[M_N]]\right)^2} = \frac{2 \ln N}{(\mathbb{E}[\mathbb{E}[M_N]])^2} y$$

Here,  $y$  acts purely as a scaling factor. Since we are primarily interested in a conservative ceiling for the expected maximum, we can drop  $y$  to simplify the inequality. This

---

leads directly to the final, conservative bound:

$$T_{min} < \frac{2 \ln N}{(\mathbb{E}[\mathbb{E}[M_N]])^2}$$

This final form clearly highlights that  $T_{min}$  is bounded above by the square of the asymptotic ceiling of the expected maximum, ensuring a safe estimate for the minimum backtest length.

### Discussion: Implications of the Minimum Backtest Length

**Overfitting Risk Increases with Strategy Exploration** The bound on  $T_{min}$  directly highlights that as the number of independent strategies  $N$  grows, the required backtest length increases logarithmically. Practically, this means that testing more configurations or tuning more parameters without extending the data history proportionally significantly raises the risk of selecting a strategy that appears profitable purely by chance. Analysts must be transparent about how many variations they tested to allow realistic assessment of overfitting risk.

**Setting Realistic Expectations for Backtest Performance** Theorem 1.2 quantifies the minimum amount of data needed to meaningfully evaluate a strategy's Sharpe ratio. Even if an in-sample Sharpe ratio looks impressive, this result reminds practitioners that without sufficient historical data, apparent "skill" could simply reflect noise. This provides a concrete metric to calibrate confidence in backtest results and avoid over-interpreting short-term performance.

**Guiding Practical Backtest Design and Reporting.** The inequality  $T_{min} \lesssim \frac{2 \ln N}{(\mathbb{E}[\mathbb{E}[M_N]])^2}$  offers a practical tool for designing backtests: it informs how much data is required relative to the number of strategies being evaluated. From a reporting standpoint, it encourages analysts to disclose not only the best in-sample performance but also the number of trials and the backtest length, helping decision-makers gauge the robustness of the strategy before committing capital.

---

## References

- Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events: For insurance and finance* (Vol. 33). Springer-Verlag. <https://doi.org/10.1007/978-3-642-33483-2>
- Lo, A. W. (2002). The statistics of sharpe ratios. *Financial Analysts Journal*, 58(4), 36–52. <https://doi.org/10.2469/faj.v58.n4.2453>
- Resnick, S. I. (1987). *Extreme values, regular variation and point processes*. Springer. <https://doi.org/10.1007/978-0-387-75953-1>