Portfolio Theory A1

MV Backtesting and Out-of-Sample Performance

Christopher Eason (ESNCHR001)

2025-09-11

Table of contents

1 Part I: Introduction to Strategy Backtesting **2**
1.1 Question 1: Asymptotic distribution of the estimated annualized Sharpe Ratio 2
1.1.1 Definitions and Notation 2
1.1.2 Central Limit Theorem 2
1.1.3 Asymptotic Joint Distribution 3
1.1.4 Delta Method 4
1.1.5 Annualisation 5
1.2 Question 2: Question 2: Expected Maximum of a Sample of IID Normal Variables 6
1.2.1 Definition of the von Mises Function 6
1.2.2 Showing that the Normal Distribution is a von Mises Function . . . 7
1.2.3 Proposition 3.3.25: von Mises Functions and the Max-Domain of Attraction (MDA) 8
1.2.4 Linking the von Mises–MDA Proposition to the Standard Normal Distribution 9
1.2.5 9
1.3 Question 3: Minimum Backtest Length to Avoid Overfitting 10

References **11**

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 $y \to \infty$ to:

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

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

$$SR = \frac{\mu}{\sigma}\sqrt{q} \tag{1}$$

- Since μ and σ 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, ..., 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

$$\hat{SR} = \frac{\hat{\mu}}{\hat{\sigma}}\sqrt{q} \tag{2}$$

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 χ^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

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

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 χ^2 distribution imply that

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

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

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

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.

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

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 $\hat{\sigma}^2$.

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

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 μ and σ as parameters, and uses Equation 1. This means that $g(\hat{\mu}, \hat{\sigma^2}) = \hat{SR}$ 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 μ and σ^2 through the nonlinear function $g(\mu,\ \sigma^2) = SR$. This allows us to derive the asymptotic distribution of the Sharpe ratio estimator $\hat{SR}$ using the gradient of g and the covariance matrix of μ and σ^2 (Lo, 2002).

First, we can re-write Equation 6 as

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \underset{T\to\infty}{\overset{a}{\sim}} \mathrm{N}(0, \mathbf{V}_{\boldsymbol{\theta}})$$

Employing the delta method:

$$\sqrt{T}\left(g(\hat{\boldsymbol{\theta}}) - g(\boldsymbol{\theta})\right) \underset{T\to\infty}{\overset{a}{\sim}} \mathrm{N}\left(0, \left(\frac{\partial g}{\partial \boldsymbol{\theta}}\right)' \mathbf{V}_{\boldsymbol{\theta}} \frac{\partial g}{\partial \boldsymbol{\theta}}\right) \tag{7}$$

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

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

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

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

$$\left(\frac{\partial g}{\partial \boldsymbol{\theta}}\right)' \mathbf{V}_{\theta} \frac{\partial g}{\partial \boldsymbol{\theta}} = q + \frac{q\mu^2}{2\sigma^2} = q + \frac{SR^2}{2} = q\left(1 + \frac{SR^2}{2q}\right)$$

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 $g(\hat{\boldsymbol{\theta}}) = \hat{SR}$ and $g(\boldsymbol{\theta}) = SR$

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

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

Annualisation

Since $T = yq$, we express the asymptotic variance per year by switching the limiting argument from $T \to \infty$ to $y \to \infty$ 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.

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

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

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 μ and σ^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 γ

To approximate the expected maximum of N i.i.d. Normal random variables, we proceed in three steps. First, we show that the Normal distribution is a von Mises function using Example 3.3.29 (Embrechts et al., 1997). By Proposition 3.3.25, every von Mises function lies in the maximum domain of attraction of the Gumbel distribution (MDA(Λ)) (Embrechts et al., 1997). Second, by the Fisher–Tippett–Gnedenko theorem, this implies that the properly normalised sample maximum converges in distribution to the Gumbel law. Finally, applying Resnick's Proposition (iii) on moment convergence, we obtain that the expectation of the normalised maximum converges to the Euler–Mascheroni constant γ (Resnick, 1987). Together, these results yield the approximation $\mathbb{E} \approx \alpha + \gamma\beta$ with α, β 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 randomvariable 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:

- $a : (z,\ x_f) \to (0,\ \infty))$ is a 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.
- $c : (z,\ x_F) \to (0,\ \infty)$ is a positive function such that $\lim_{x \to x_f} c(x) = c > 0$ 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

$$\bar{F}(x) = c(x)exp\left(-\int_z^x \frac{1}{a(t)}dt\right) \qquad z < x < x_F \tag{8}$$

Showing that the Normal Distribution is a von Mises Function

Let $X\ N(0,\ 1)$ 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

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

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

$$\lim_{x\to\infty} \frac{\bar{Z}(x)}{\frac{\phi(x)}{x}} = \lim_{x\to\infty} \frac{\frac{d}{dx}(\bar{Z}(x))}{\frac{d}{dx}\frac{\phi(x)}{x}} = \lim_{x\to\infty} \frac{\frac{d}{dx}(1 - Z(x))}{\frac{\phi'(x)x - \phi(x)}{x^2}} = \lim_{x\to\infty} \frac{-Z(x)}{\frac{-x\phi(x)\cdot x - \phi(x)}{x^2}} =$$

$$= \lim_{x\to\infty} \frac{-\phi(x)}{-\phi(x)\left(1 + \frac{1}{x^2}\right)} = \lim_{x\to\infty} \frac{-\phi(x)}{-\phi(x)} = 1$$

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

$$\bar{Z}(x) \sim \frac{\phi(x)}{x}$$

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

$$a(x) = \frac{\bar{Z(x)}}{\phi(x)} = \frac{\frac{\bar{\phi(x)}}{x}}{\phi(x)} \sim \frac{1}{x} \tag{9}$$

- Therefore solving for c(x)

$$c(x) = \overline{Z}(x) \exp\Big(\int_z^x \frac{1}{a(t)} dt\Big) = \overline{Z}(x) \exp\Big(\int_z^x \frac{1}{\frac{1}{x}} dt\Big) = \frac{\varphi(x)}{x} \cdot \exp\Big(\frac{x^2}{2}\Big).$$

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

Note that $c(x) > 0$ as $x \to \infty$ for unbounded distributions like the normal, the slowly varying factor c(x) may converge to zero as long as it varies slower than the exponential decay.

Therefore, with $a(x) \sim \frac{1}{x}$ and $c(x) \sim \frac{1}{x\sqrt{2\pi}}$, using Equation 8 we have

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

which is exactly the standard tail approximation (Mills' ratio) for the standard normal distribution. 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 $X_1,\ X_2,\ \ldots X_N$ are IID random variables with distribution F, then the properly normalized maximum

$$M_N := max\{X_1,\ X_2,\ \ldots X_N\} \underset{N\to\infty}{\overset{a}{\sim}} G \tag{10}$$

  Where G is the Gumbel distribution i.e. $F \in$ MDA(Λ)

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

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

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

$$\frac{M_N - d_N}{c_N} \underset{N\to\infty}{\overset{d}{\to}} G$$

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 9 that $Z(x)$ is a von Mises function with auxiliary function $a(x) \sim \frac{1}{x}$ as $x \to \infty$, the standard normal belongs to the max-domain of attraction (MDA) of the Gumbel distribution i.e. $F \in$ MDA(Λ) Equation 10.

We now calculate the norming constants: $d_n = (2ln(n))^{frac12} - \frac{ln(ln(n))+ln(4\pi)}{2(2ln(n))^{\frac{1}{2}}} + O\left((ln(n))^{\frac{1}{2}}\right)$ $c_n = (2ln(n))^{-\frac{1}{2}}$

Question 3: Minimum Backtest Length to Avoid Overfitting

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