Quantitative Portfolio Management: Theory and Practice

Chapter 2:
Performance Measurement

Raman Uppal
EDHEC Business School

2025-2026

The big picture: Plan for the entire book

Part A: **Preliminaries**

Chapter 1: Managing financial data

Chapter 2: Performance measurement (especially out of sample)

Part B: **Modern portfolio management**

Chapter 3: Mean-variance portfolios that *ignore* estimation error

Chapter 4: Mean-variance portfolios that *adjust for* estimation error

Part C: **Post-Modern Portfolio Management**

Chapter 5: CAPM-based portfolios: Black-Litterman model

Chapter 6: Factor-based portfolios: Parametric portfolio policies

Chapter 7: Volatility-timed factor portfolios

Chapter 8: Portfolios exploiting systematic risk factors *and* unsystematic risk

Table of contents

Road map

What do we want to do in Chapter 2

We first study how to measure the performance of a portfolio strategy; in particular, the importance of evaluating portfolio returns that are out-of-sample, instead of in-sample, which suffer from a *look-ahead* bias.

Then, we examine different metrics for evaluating portfolio performance. The key measure we will use is the Sharpe ratio.

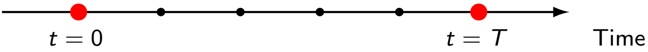
We conclude by studying how to test if the difference between the Sharpe ratios of two portfolios is statistically significant.

Road map

Static, myopic, and dynamic portfolios . . . I

- **Static** portfolio: Hold the same portfolio between $t = 0$ and $t = T$:
  - That is, form portfolio at $t = 0$ and hold it until T.

Static: Hold the same portfolio between $t = 0$ and $t = T$

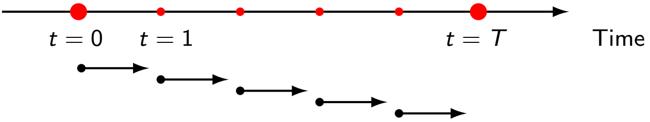


- An example of a static portfolio is:
  - A 25-year old person designs a portfolio for retirement at $t = 0$
  - This person holds the same portfolio for the next 40 years, until T.

Static, myopic, and dynamic portfolios . . . II

- **Myopic** portfolio: Revise portfolio at each date t, but ignore the possibility of rebalancing at future dates.
  - That is, when forming the portfolio at t, ignore the possibility that you will be rebalancing the portfolio again at $t + 1$.

Myopic: Rebalance portfolio at each date, but looking only one period ahead

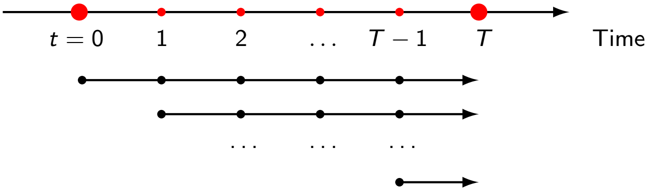


- An example of a myopic portfolio is:
  - A 25-year old person designs a portfolio for retirement at $t = 0$;
  - The person rebalances the portfolio each year for the next 40 years;
  - When rebalancing, the person optimizes for only the next year.

Static, myopic, and dynamic portfolios . . . III

- **Dynamic** portfolio: Choose portfolio at t, knowing we will be rebalancing again at $t+1$.

Dynamic: Rebalance portfolio at each date, anticipating future rebalancing



- An example of a dynamic portfolio is:
  - A 25-year old person designs a portfolio for retirement at $t=0$
  - The person rebalances the portfolio each year for the next 40 years;
  - When rebalancing, the person optimizes for next year anticipating that she will be rebalancing again the following year.

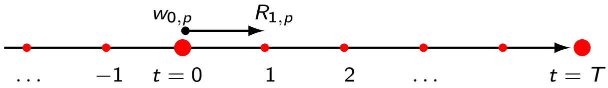
Static, myopic, and dynamic portfolios . . . IV

- We do not study static portfolios because they are sub-optimal relative to myopic (and dynamic) portfolios.
- We do not study dynamic portfolios because to implement them, we need to estimate a larger number of parameters than for myopic portfolios, and so these portfolios perform poorly out of sample.
  - For an excellent description of dynamic portfolios, see: Merton (1971, 1990) and Campbell and Viceira (2002).
- In our course, we will focus on myopic portfolios
  (even though, if one ignores estimation problems, myopic portfolios are suboptimal relative to dynamic portfolios).

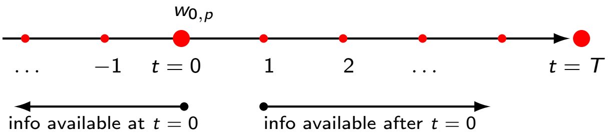
Road map

Portfolio returns

- When evaluating portfolio performance, we will focus on the return of the portfolio.
  - In principle, one could look at other properties of the portfolio, e.g., ESG properties.
- Denote the gross return at date t of portfolio p by $R_{t,p}$.
  - We will adopt the convention of $R_{\text{date},\text{asset-name}}$
- For a myopic portfolio formed at date $t=0$, $w_{0,p}$, we will want to evaluate its return at date $t=1$, $R_{1,p}$.

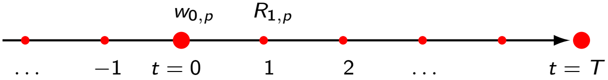


Portfolio weights should depend only on past information



- Consider choosing at date $t = 0$ a myopic portfolio $w_{0,p}$.
- When choosing $w_{0,p}$, it is important that only information available at date $t = 0$ be used to choose the portfolio.
  - $w_{0,p}$ can depend on information available at $t = 0, -1, -2, \ldots$
  - $w_{0,p}$ should not depend on information from the future; that is, no information from $t = 1, 2, \ldots$ should be used.
- It is not realistic to choose weights that depend on future information.

In-sample and out-of-sample portfolio performance



- If $w_{0,p}$ depends only on past information, then $R_{1,p}$ is called the out-of-sample return—because the weights depend on information until $t = 0$ and the return is for $t = 1$, which is out of sample.
- But, if $w_{0,p}$ is chosen using also future information that is available at $t = 1, 2, \ldots$, then $R_{1,p}$ is called the in-sample return.
- We will always want to choose portfolio weights that depend on currently available information and not future information.
- Therefore, we will always study out-of-sample performance of the myopic portfolio strategies we consider.

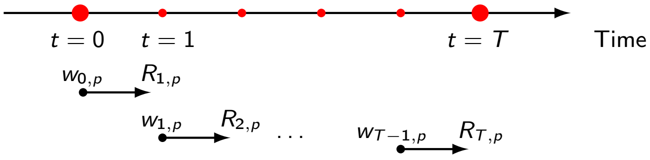
Out-of-sample returns of myopic portfolios

- A myopic portfolio chosen at t will realize its return at $t+1$, where $t = \{0, 1, 2, \ldots, T-1\}$.
- So, the sequence of portfolios

  $$w_p = \{w_{0,p}, w_{1,p}, \ldots, w_{T-1,p}\},$$

  will generate the sequence of returns

  $$R_p = \{R_{1,p}, R_{2,p}, \ldots, R_{T,p}\}.$$



Rolling vs. expanding window . . . I

- Suppose when choosing her portfolio, an investor estimates future expected returns and risk using historical data for a constant interval.
- For example, suppose the investor looks at the past 5 years of monthly data when making her decision at the start of 2020.
  - Then, when making her decision in 2020, the investor will consider data for the past 5 years 2019, 2018, 2017, 2016, and 2015.
  - Similarly, when making her decision in 2021, the investor will consider data for the past 5 years 2020, 2019, 2018, 2017, and 2016.
  - Finally, when making her decision in 2022, the investor will consider data for the past 5 years 2021, 2020, 2019, 2018, and 2017.
- This fixed-length lookback window is called a rolling window.

Rolling vs. expanding window . . . II

- An alternative approach is to estimate future expected returns and risk using historical data for an expanding interval.
  - For example, suppose when making her decision in 2020, the investor considers data until 2015; that is, data for the past 5 years: 2019, 2018, 2017, 2016, and 2015.
  - When making her decision in 2021, the investor will consider data until 2015; that is, data for the past 6 years: 2020, 2019, 2018, 2017, 2016, and 2015.
  - Finally, when making her decision in 2022, the investor will consider data until 2015; that is, data for the past 7 years: 2021, 2020, 2019, 2018, 2017, 2016, and 2015.
- This increasing lookback window is called an expanding window.

Rolling vs. expanding window . . . III

1. We should use a rolling window if we feel that
   - the distant past is not relevant to current decisions,
   - or, if the distant past is different from the more recent past.
2. We should use an expanding window if we feel that
   - the distant past is similar to the recent past.
3. Another alternative is to have an exponentially declining weight on historical data, with the weight decreasing as one goes back in time.

Python functions for rolling and expanding windows

- The pandas library in Python has
  - special functions for doing rolling- and expanding-window analysis;
  - the library also allows for exponentially declining weights on the past.
- Using the links below, you can read about the functions for
  - rolling windows, and
  - expanding windows.
- These functions can be used for some of the course assignments.

Main objective of this course

- The main objective of this course is to understand
  - how to choose the portfolio weights

$$w_p = \{w_{0,p}, w_{1,p}, \ldots, w_{T-1,p}\},$$

  - in order to generate portfolio returns

$$R_p = \{R_{1,p}, R_{2,p}, \ldots, R_{T,p}\}.$$

  - that have good performance.
- In the rest of this class, we will study various metrics of portfolio performance, which we will then use for the rest of the course.

Road map

Measures of portfolio performance

For the return on each portfolio strategy, we compute various performance metrics that can be divided into three groups.

1. Measures of mean (average) returns.
2. Measures of risk of returns.
3. Measures combining mean returns and risk of returns.

Road map

Measures of mean returns

- For the portfolio measures of returns, we use
  1. Total portfolio mean (average) return.
  2. Systematic component of the portfolio mean return based on some factor model for expected returns (see below).
     - The systematic component of portfolio mean returns is the reward for bearing systematic risk.
  3. Alpha component of the portfolio mean return, which is the component of expected returns unexplained by the factor model.
     - The unsystematic component of portfolio mean returns is the reward for bearing unsystematic risk.
  4. Outperformance frequency, which is the average fraction of times that the portfolio being evaluated has a higher cumulative return than the benchmark portfolio within twelve months from the beginning of each such period.

Return factors

- One may believe that the returns on an asset (stocks or portfolios) depend on the exposure of that asset to systematic risk factors.
- One kind of systematic risk factor that is considered in finance is the return on a long-short portfolio.
- Long-short portfolios that have been considered in finance include:
  - $R_{\text{mkt}} - R_f$ = return on market minus return on risk-free asset;
  - R_{smb} = small market capitalization minus big (size);
  - R_{hml} = high book-to-market ratio minus low (value);
  - R_{umd} = up positive momentum minus down momentum;
  - R_{rmw} = robust minus weak profitability;
  - R_{cma} = conservative (low) minus aggressive (high) investment firms.

Factor models of expected returns . . . I

- The factor models of expected returns that one can use to compute the systematic and unsystematic component of returns are:

1. One-factor market model of Sharpe (1964),

$$\mathbb{E}[R_p] - R_f = \beta_{p,\text{mkt}}\,\mathbb{E}[R_{\text{mkt}} - R_f].$$

2. Three-factor model of Fama and French (1993),

$$\mathbb{E}[R_p] - R_f = \beta_{p,\text{mkt}}\,\mathbb{E}[R_{\text{mkt}} - R_f] + \beta_{p,\text{smb}}\,\mathbb{E}[R_{\text{smb}}] + \beta_{p,\text{hml}}\,\mathbb{E}[R_{\text{hml}}]$$

3. Four-factor model (three-factors + momentum Carhart (1997),

$$\begin{aligned}\mathbb{E}[R_p] - R_f = {} & \beta_{p,\text{mkt}}\,\mathbb{E}[R_{\text{mkt}} - R_f] + \beta_{p,\text{smb}}\,\mathbb{E}[R_{\text{smb}}] + \beta_{p,\text{hml}}\,\mathbb{E}[R_{\text{hml}}] \\ & + \beta_{p,\text{umd}}\mathbb{E}[R_{\text{umd}}]\end{aligned}$$

Factor models of expected returns . . . II

4. Five-factor model of Fama and French (2015),

$$\begin{aligned}\mathbb{E}[R_p] - R_f = {} & \beta_{p,\mathsf{mkt}}\,\mathbb{E}[R_{\mathsf{mkt}} - R_f] + \beta_{p,\mathsf{smb}}\,\mathbb{E}[R_{\mathsf{smb}}] + \beta_{p,\mathsf{hml}}\,\mathbb{E}[R_{\mathsf{hml}}] \\ & + \beta_{p,\mathsf{rmw}}\,\mathbb{E}[R_{\mathsf{rmw}}] + \beta_{p,\mathsf{cma}}\,\mathbb{E}[R_{\mathsf{cma}}]\end{aligned}$$

5. Six-factor model (five-factors + momentum (Fama and French 2018)),

$$\begin{aligned}\mathbb{E}[R_p] - R_f = {} & \beta_{p,\mathsf{mkt}}\,\mathbb{E}[R_{\mathsf{mkt}} - R_f] + \beta_{p,\mathsf{smb}}\,\mathbb{E}[R_{\mathsf{smb}}] + \beta_{p,\mathsf{hml}}\,\mathbb{E}[R_{\mathsf{hml}}] \\ & + \beta_{p,\mathsf{rmw}}\,\mathbb{E}[R_{\mathsf{rmw}}] + \beta_{p,\mathsf{cma}}\,\mathbb{E}[R_{\mathsf{cma}}] \\ & + \beta_{p,\mathsf{umd}}\,\mathbb{E}[R_{\mathsf{umd}}]\end{aligned}$$

6. See also the series of papers on q-factor models that offer an alternative perspective to Fama and French models: Hou, Xue, and Zhang (2015, 2017a, 2017b) and Hou, Mo, Xue, and Zhang (2019).

Factor models of expected returns . . . III

7. K-"factor" model, where factors are principal components (PCs) of a principal component analysis (PCA) of returns (Connor, Goldberg, and Korajczyk 2010; Kozak, Nagel, and Santosh 2018, 2020).
   - These are statistical factors (obtained from an eigenvalue-eigenvector decomposition of the covariance matrix of returns).
   - These factors are often called "latent" (unobservable), in contrast to the observable Fama-French factors.
   - We will study models with latent factors later in the course.

Factor models of expected returns . . . IV

- A recent paper by Dello-Preite, Uppal, Zaffaroni, and Zviadadze (2025) shows that
  - perhaps it is important to adjust mean returns also for unsystematic risk, and
  - once we adjust mean returns for unsystematic risk, the only systematic risk factor needed is the market (or the market and size).
- We will study this paper later in the course.

Road map

Measures of the riskiness of returns . . . I

- We measure the riskiness of portfolio returns using
  1. Volatility (or standard deviation or square root of variance) of portfolio returns.
  2. Semi-volatility (square root of semi-variance) of portfolio returns is calculated in the same manner as the volatility, but only those observations that fall below the mean are included in the calculation.

$$\text{Semi-volatility} = \sqrt{\frac{1}{T}\sum_{\substack{t=1 \\ x_t < \hat{\mu}}}^{T} \left[x_t - \hat{\mu}\right]^2}, \quad \text{where } \hat{\mu} \text{ is the estimated mean.}$$

  3. Skewness of portfolio returns (explained in the previous chapter).
  4. Kurtosis of the portfolio returns (explained in the previous chapter).

Measures of the riskiness of returns . . . II

5. Average maximum drawdown (MDD), defined as the time-series average of the maximum percentage loss of the portfolio value $V(\tau)$ over any period from τ_1 to τ_2 during the last twelve months:

$$\text{MDD} = \frac{1}{T-13}\sum_{t=12}^{T-1} \max_{t-11 \leq \tau_1 < \tau_2 \leq t} \left\{0, \frac{V(\tau_1)}{V(\tau_2)} - 1\right\} \times 100.$$

- This page by Vamshi Jandhyala shows how to use Python to compute drawdown.

Measures of the riskiness of returns . . . III

6. Value at Risk (VaR) measures how much a portfolio may lose (with a given probability, usually 1% or 5%), given normal market conditions, over a set time period such as a day or a month.
   - VaR is typically used by firms and regulators in the financial industry to gauge the amount of assets needed to cover possible losses.
   - For example, if a portfolio has a one-week 99% VaR of $10 million, that means that there is a 1% probability that the portfolio will fall in value by more than $10 million over a one-week period.
   - Thus, a loss of $10 million or more is expected 1 week out of 100 (because of the 1% probability), so about once every two years.
   - See Wikipedia for further details and, in particular, the weaknesses of the VaR measure.

Measures of the riskiness of returns . . . IV

7. Conditional value at risk (CVaR) at $q\%$ level is the expected return on the portfolio in the worst $q\%$ of cases.
   - CVaR is also called: average value at risk (AVaR), expected shortfall, expected tail loss, and superquantile.
   - CVaR is an alternative to value at risk that is more sensitive to the shape of the tail of the loss distribution.
   - See Wikipedia for further details.

Python code

- This web page by Vamshi Jandhyala shows how to use Python to compute:
  - Skewness
  - Kurtosis
  - Semi-volatility
  - Value at Risk, and
  - Conditional Value at Risk (expected shortfall).

Road map

Measures combining mean and risk of returns . . . I

Measures of the risk-return tradeoff include the following.

1. Sharpe ratio (SR), which is the mean return of the portfolio in excess of the risk-free rate divided by the volatility of portfolio returns:

$$\mathrm{SR}_p = \frac{\mathbb{E}[R_p - R_f]}{\sqrt{\mathbb{V}[R_p - R_f]}} = \frac{\text{excess mean return}}{\text{total portfolio risk}}.$$

- Note that (for IID returns) the Sharpe ratio scales with $\sqrt{T}$:

$$\mathrm{SR}_p = \frac{\mathbb{E}[R_p - R_f]}{\sqrt{\mathbb{V}[R_p - R_f]}} = \frac{\text{numerator scales with } T}{\text{denominator scales with } \sqrt{T}} = \text{SR scales with } \sqrt{T}.$$

- Therefore, to annualize the Sharpe ratio of monthly returns, one needs to multiply it by $\sqrt{12} \approx 3.46$.

Measures combining mean and risk of returns . . . II

- Correct scaling to annualize the Sharpe ratio
  - Note that multiplying by $\sqrt{12} \approx 3.46$ is correct only if returns are IID with no serial correlation;
  - Lo (2002, eq. (20) & Table 2) shows that the correct multiplier depends on the serial correlation of the portfolio returns.
  - The correct multiplier can be much smaller than $\sqrt{12}$ (if returns have positive serial correlation);
  - The correct multiplier can be much larger (if returns have negative serial correlation).

Measures combining mean and risk of returns . . . III

- The Sharpe ratio has several limitations:
  - Sharpe ratio is appropriate for efficient portfolios but not for individual assets because it ignores correlation between these assets.
  - Sharpe ratio, because it depends on the mean and variance of returns, is not appropriate if returns have high skewness and kurtosis.
  - Sharpe ratio looks at first two moments, but ignores the distribution of returns. In contrast, the Omega measure uses the entire distribution; see Wikipedia.
  - Despite these limitations, the Sharpe ratio is the main performance measure we will use throughout the course.
- This web page by Vamshi Jandhyala shows how to use Python to compute the Sharpe ratio.
- Links to other websites with code for computing the Sharpe ratio are provided below.

Measures combining mean and risk of returns . . . IV

2. Sortino ratio (SoR) is the mean return in excess of a target rate, divided by the downside variance (DV) of portfolio returns relative to the target rate, R_{target}, which could be the risk-free rate:

$$\text{SoR}_p = \frac{\mathbb{E}[R_p] - R_{\text{target}}}{\sqrt{\text{DV}}}, \quad \text{where}$$

$$\text{DV} = \frac{1}{T}\sum_{i=1}^{T}\Big(\min\big\{R_p - R_{target}, 0\big\}\Big)^2,$$

- Compared to Sharpe ratio, Sortino ratio penalizes only downward deviations of the portfolio return from the target rate of return.

Measures combining mean and risk of returns . . . V

- This article by by John Doherty presents a step-wise explanation of computing the Sortino ratio.
- This article by Rollinger and Hoffman presents another example of computing the Sortino ratio.
- This web page from Quantitative Finance explains how to code the Sortino ratio using Python.

Measures combining mean and risk of returns . . . VI

3. Treynor ratio (TR) is the mean return of a portfolio (or individual asset) in excess of the risk-free rate divided by the market beta of returns:

$$\mathrm{TR}_p = \frac{\mathbb{E}[R_p - R_f]}{\beta_{p,\mathrm{mkt}}} = \frac{\text{excess mean return}}{\text{systematic risk}},$$

where $\beta_{p,\mathrm{mkt}} = \frac{\mathrm{Cov}[R_p, R_{\mathbf{mkt}}]}{\mathrm{Var}[R_{\mathbf{mkt}}]} = \frac{\mathbb{C}[R_p, R_{\mathbf{mkt}}]}{\mathbb{V}[R_{\mathbf{mkt}}]}$.

- Because the Treynor ratio ignores unsystematic risk, two portfolios (or assets) with the same Treynor ratio may have very different levels of unsystematic risk.
- Both the Sharpe ratio and Treynor ratio are reasonable for ranking investments, but neither tells us by how much one investment is better than another.

Measures combining mean and risk of returns . . . VII

4. Risk-adjusted performance (RAP or M^2 or M2), developed by Franco Modigliani and his granddaughter, Leah Modigliani, is the excess return of a portfolio adjusted for its risk relative to that of a benchmark, B.

$$\begin{aligned}\text{RAP}_p &= \mathbb{E}[R_p - R_f] \times \frac{\mathbb{V}[R_B]}{\mathbb{V}[R_p - R_f]} \\ &= [\text{excess mean return of } p] \times \frac{\text{return variance of benchmark}}{\text{return variance of } p}.\end{aligned}$$

- So, if the return variance of the benchmark portfolio is half of that of portfolio p, then the risk-adjusted performance of p will be half of the unadjusted performance, $\mathbb{E}[R_p - R_f]$.
- RAP is useful because, when plotting cumulative returns of different strategies, important to adjust returns so they all have same variance.

Measures combining mean and risk of returns . . . VIII

5. Information Ratio (IR) is the ratio of the active return of a portfolio (or asset) in excess of the benchmark's return to the volatility of the active return (i.e., active risk or benchmark tracking risk)

$$\mathrm{IR}_p = \frac{\mathbb{E}[R_p - R_B]}{\sqrt{\mathbb{V}[R_p - R_B]}} = \frac{\text{mean return in excess of benchmark}}{\text{volatility of excess return}}.$$

- In contrast, the Sharpe ratio uses the risk-free return as the benchmark, whereas IR uses a risky index as a benchmark (such as the market index).
- The Sharpe ratio is useful for evaluating absolute portfolio returns, whereas IR is useful for evaluating relative returns of a portfolio.

Measures combining mean and risk of returns . . . IX

6. Jensen's alpha (α_J) is the ex-post abnormal return of a portfolio(or asset) over the theoretical (model-implied) return

$$\alpha_J = \underbrace{(R_p - R_f)}_{\text{realized excess return}} - \underbrace{\left(R_{\text{model}} - R_f\right)}_{\text{model-implied excess return}} .$$

- If the benchmark model is the CAPM: $\left(R_{\text{model}} - R_f\right) = \beta_{p,\text{mkt}}\left(R_{\text{mkt}} - R_f\right)$, then

$$\alpha_J^{\text{CAPM}} = \underbrace{(R_p - R_f)}_{\text{realized excess return}} \underbrace{- \beta_{p,\text{mkt}}\left(R_{\text{mkt}} - R_f\right)}_{\text{CAPM return}} \quad \ldots \text{CAPM alpha}$$

- If the benchmark model is the Fama-French-Carhart four-factor model, then we get the four-factor alpha:

$$\alpha_J^{\text{FFC}} = \underbrace{(R_p - R_f)}_{\substack{\text{realized}\\ \text{excess return}}} - \underbrace{\left[\beta_{p,\text{mkt}}\left(R_{\text{mkt}} - R_f\right) + \beta_{p,\text{smb}} R_{\text{smb}} + \beta_{p,\text{hml}} R_{\text{hml}} + \beta_{p,\text{umd}} R_{\text{umd}}\right]}_{\text{FFC return}}$$

Measures combining mean and risk of returns . . . X

7. One could also compute the certainty equivalent return assuming an investor has a particular utility function.
    - The benefit of this measure is that it adjusts returns not just for variance but also for higher moments of return, such as skewness and kurtosis.
    - A disadvantage of this measure is that one needs to specify the parameters of the investor's utility function; in particular, the investor's degree of risk aversion.
    - Because in many cases, the conclusions one draws from evaluating Sharpe ratios are similar to those from evaluating the certainty equivalent return, we will focus on the Sharpe ratio.

Python code

- This page from CodeArmo has Python code to compute the
  - Sharpe ratio,
  - Sortino ratio,
  - Max drawdown, and
  - Calmar ratio.
- This page from Turing Finance has Python code for a large number of measures of risk-adjusted return. In particular:
  - Measures of return adjusted for risk:
    Treynor ratio, Sharpe ratio, Information ratio, Modigliani ratio.
  - Measures of return adjusted for Value-at-Risk
  - Measures of return adjusted for risk based on partial moments:
    Omega, Sortino, Kappa, gain-loss, and upside-potential ratios.
  - Measures of return adjusted for drawdown risk:
    Calmar ratio, Sterling ratio, and Burke ratio.

Road map

Turnover for an individual asset

- When a portfolio is rebalanced, some assets need to be sold while others need to be purchased.
- When the portfolio is rebalanced at time $t+1$, it gives rise to a trade in each asset n of the magnitude

$$\left|w_{t+1,n} - w_{t^+,n}\right|, \quad \text{where}$$

  - $w_{t+1,n}$ is the (new) weight in asset n at date $t+1$, and
  - $w_{t^+,n}$ denotes the (old) date-t weight evaluated at date $t+1$ prices.
- This sale and purchase of assets is called turnover, and it incurs a transaction cost.
- Usually, more sophisticated portfolio strategies have larger turnover.

Turnover for an entire portfolio

- The turnover of a portfolio between dates t and $t+1$ is the sum of absolute changes in weights over the N stocks in the portfolio:

$$\text{Turnover} = \sum_{n=1}^{N} \left(\left| w_{n,t+1} - w_{n,t^+} \right| \right), \quad \text{where}$$

  - $w_{t+1,n}$ is the (new) weight in asset n at date $t+1$, and
  - $w_{t^+,n}$ is the (old) date-t weight evaluated at date $t+1$ prices.

Average turnover for a portfolio over time

- The annualized average monthly turnover of a portfolio is
  - the time-series mean (over the T monthly rebalancing dates) of
  - the sum over the N stocks in the portfolio of the absolute changes in weights
  - multiplied by twelve:

$$\text{Turnover} = \frac{1}{T}\sum_{t=0}^{T-1}\sum_{n=1}^{N}\left(\left|w_{n,t+1} - w_{n,t^+}\right|\right) \times 12,$$

where w_{n,t^+} are the weights of the portfolio using the prices at $t+1$.

Portfolio turnover ratio

- The investment industry often reports the portfolio turnover ratio.
- This is computed as:

$$\text{Portfolio turnover ratio} = \frac{\min(\text{securities sold}, \text{securities purchased})}{\text{average net assets}} \times 100$$

- Example:
  - A portfolio purchased \$10m of securities and sold \$9m of securities, over a one-year time period, so $\min(10, 9) = \$9\text{m}$.
  - Over the one-year period, the portfolio's average net assets were \$100m.
  - The portfolio turnover ratio is: $(\$9\text{m}/\$100\text{m}) \times 100 = 9\%$.
- Note: The measure of turnover (on the previous slide) includes both sales and purchases, while the turnover ratio (on this slide) includes only one side, sales or purchases, depending on which one is smaller.

Turnover and transaction cost . . . I

- To measure the economic cost of turnover, one can compute the transaction costs generated by this turnover.
- One simple approach to measuring transaction cost is to assume that there are only proportional costs for trading (i.e., no fixed costs).
- Then, the transaction cost of a portfolio strategy at $t + 1$ is the
  - turnover for date $t + 1$ multiplied by
  - κ, the average proportional cost of trading a stock in that portfolio:

$$\kappa \times \sum_{n=1}^{N} \left| w_{t+1,n} - w_{t^+,n} \right|.$$

Turnover and transaction cost . . . II

- Therefore, the net-of-costs evolution of wealth for portfolio p is

$$W_{t+1,p} = W_{t,p} \underbrace{(1 + R_{t+1,p})}_{\text{gross return}} \times \underbrace{\left(1 - \kappa \times \sum_{n=1}^{N} |w_{t+1,n} - w_{t^+,n}|\right)}_{\substack{\text{adjustment for} \\ \text{proportional transaction cost}}},$$

so that the return net of transactions costs is

$$\frac{W_{t+1,p}}{W_{t,p}} - 1 = \underbrace{(1 + R_{t+1,p})}_{\text{gross return}} \times \underbrace{\left(1 - \kappa \times \sum_{n=1}^{N} |w_{t+1,n} - w_{t^+,n}|\right)}_{\substack{\text{adjustment for} \\ \text{proportional transaction cost}}} -1.$$

Turnover and transaction cost . . . III

- One can then set the proportional transaction cost κ equal to an estimate of the average cost of trading the assets in that portfolio.
  - 50 basis points per transaction, as assumed in Balduzzi and Lynch (1999), based on the studies of the cost per transaction for individual stocks on the NYSE by Stoll and Whaley (1983), Bhardwaj and Brooks (1992), and Lesmond, Ogden, and Trzcinka (1999).
  - 10 to 20 basis points if trading large stocks in more recent times.
  - 0 to 5 basis points if you are a hedge fund.
- For a discussion of how the cost of investing has evolved over time, see French (2008).

Limitations of measuring transaction costs this way

- In the slides above, we have said that the three steps for adjusting the return of a portfolio for transaction costs are
  1. First identify the optimal portfolio
  2. Then, measure the turnover of this portfolio
  3. Finally, multiply turnover by the average transaction cost, κ.
- Each of these three steps is not the best way of doing things:
  1. First, transaction costs should be accounted for when identifying the optimal portfolio, not after the portfolio weights are optimized;
  2. Then, the turnover used should be of the portfolio that accounts for transaction costs at the optimization stage (which will be smaller) instead of one that ignores transaction costs (which will be larger).
  3. Finally, one should use the transaction cost for each stock at each date, $\kappa_{t,n}$, instead of a time-series and cross-sectional average, κ.

How to incorporate transaction costs more accurately

- Later on in the course, we will study how to incorporate transaction costs more accurately by
  1. Taking transaction costs into account when optimizing portfolio weights, instead of after choosing the portfolio weights;
  2. Using the transaction cost for each stock at each date, $\kappa_{t,n}$, instead of using some time-series and cross-sectional average;
  3. Finally, netting out trades (across factors) before computing transaction costs.
- For details of how this is done, see, for instance
  - DeMiguel, Martín-Utrera, Nogales, and Uppal (2020),
  - DeMiguel, Martín-Utrera, and Uppal (2024), and
  - Other papers cited in these two papers.

Turnover and capital-gains taxes

- Turnover, in addition to incurring transaction costs, also leads to capital-gains taxes, which we will not study in the course.
- For details on how to account for capital-gains taxes (and transaction costs) when constructing an optimal portfolio, see
  - Dybvig and Koo (1996),
  - DeMiguel and Uppal (2005),
  - and other papers cited in these two papers.

Portfolio turnover for different investment styles

- Active vs. passive management
  - Active managers have higher portfolio turnover, while passive managers, who follow an index, have lower turnover.
- Growth vs. value investing
  - Growth investors invest in rapidly expanding companies so have higher turnover, whereas value investors invest in undervalued companies with long-term potential so have lower turnover.
- Quantitative vs. qualitative analysis
  - Quantitative investors, who rely on data-driven models (as in this course), have higher turnover, while qualitative investors, who focus on fundamentals and management quality, have lower turnover.

Some strategies for controlling portfolio turnover

- When rebalancing the portfolio, account for transaction costs.
  - Taking transaction costs into account when choosing a portfolio can help reduce costs while maintaining an optimal risk-return tradeoff.
- Adopt a long-term investment horizon
  - Focusing on long-term growth can reduce turnover.
- Implement a buy-and-hold strategy
  - Holding assets for longer periods can reduce turnover and taxes.
- Utilize tax-efficient investment vehicles
  - Tax-efficient funds can reduce the impact of turnover on taxes.

Road map

Review of hypothesis testing: Steps involved

- Throughout the course, we will want to evaluate portfolio models.
- This evaluation (inference) will require a test of the hypothesis that our model is better than some benchmark model.
- A typical hypothesis test has the following steps:
  1. Stating the hypothesis.
  2. Finding an appropriate test statistic and its probability distribution.
  3. Specifying the significance level.
  4. Stating the decision rule.
  5. Collecting the data
  6. Calculating the test statistic.
  7. Making the statistical decision.
  8. Making the investment decision.

Review: The null hypothesis . . . I

- If testing a parameter, we want to know if it equals a certain value.
  - For example, we can ask if the mean $\mu = \mu_0$.
- The hypothesis $\mu = \mu_0$ is called the null hypothesis, H_0.
- We can choose the sample mean, $\bar{X}$, as a test statistic.
- If $\bar{X}$ is distributed as a normal with mean μ_0 and variance $\frac{\sigma^2}{T}$, i.e. if the null hypothesis is true, then we expect that $\bar{X}$ should be sufficiently close to μ_0.
- So, the test consists of not rejecting H_0 if $\bar{X}$ is "close" to μ_0.

Review: The null hypothesis . . . II

- To determine closeness, we can use the following probability statement (based on the standard Normal distribution):

$$P\left[\mu - 1.96\frac{\sigma}{\sqrt{T}} \leq \bar{X} \leq \mu + 1.96\frac{\sigma}{\sqrt{T}}\right] = 0.95.$$

- Under H_0, we have that $\mu = \mu_0$, so we can replace μ by μ_0:

$$P\left[\mu_0 - 1.96\frac{\sigma}{\sqrt{T}} \leq \bar{X} \leq \mu_0 + 1.96\frac{\sigma}{\sqrt{T}}\right] = 0.95.$$

Review: The alternative hypothesis: Rejection region

- In this case, "sufficiently close" is about two times the standard deviation, $\frac{\sigma}{\sqrt{T}}$.
- The rejection region of H_0 is formed by the values of X in
  - $\bar{X} < \mu_0 - 1.96\frac{\sigma}{\sqrt{T}}$, or
  - $\bar{X} > \mu_0 + 1.96\frac{\sigma}{\sqrt{T}}$.
- The rejection region corresponds to the alternative hypothesis H_1, where $\bar{X}$ is Normal with a mean greater in absolute value than μ_0.

Review: Standardizing the parameter to be tested

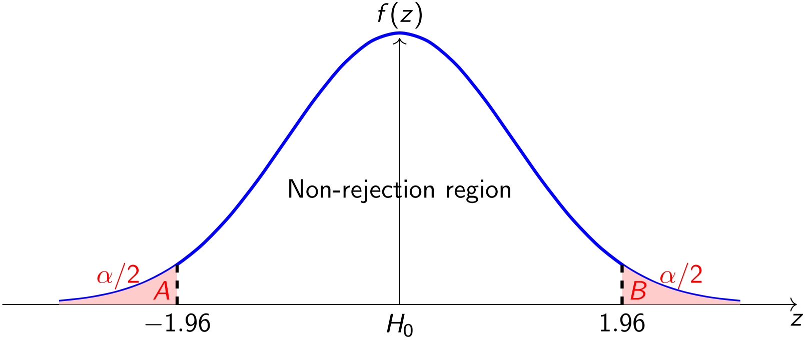
- To evaluate the rejection region, we standardize $\bar{X}$ so that its
  - mean is 0 ... done by subtracting the mean of $\bar{X}$, μ_0
  - variance is 1 ... done by dividing by standard deviation of $\bar{X}$, $\sigma/\sqrt{T}$.
- That is, instead of evaluating $\bar{X}$, we evaluate z, where:

$$z = \frac{\bar{X} - \mathbb{E}[\bar{X}]}{\sqrt{\mathbb{V}[\bar{X}]}} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{T}}.$$

- The rejection region is then formed by the values of z in
  - $A : \frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{T}}} = z < -1.96$, or
  - $B : \frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{T}}} = z > +1.96$.

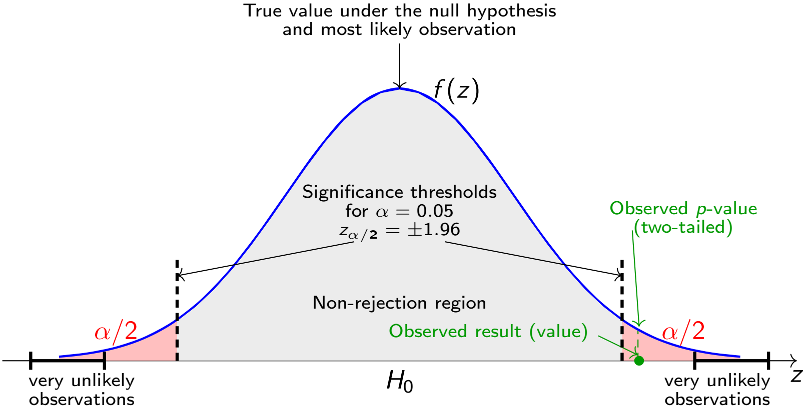
Review: Two-sided critical region

- The two regions A and B form the critical region of the test statistic.
- If the computed value for $\frac{\bar{X}-\mu_0}{\sigma/\sqrt{T}}$ is in this region, we reject H_0.



Review: P-value and rejection region

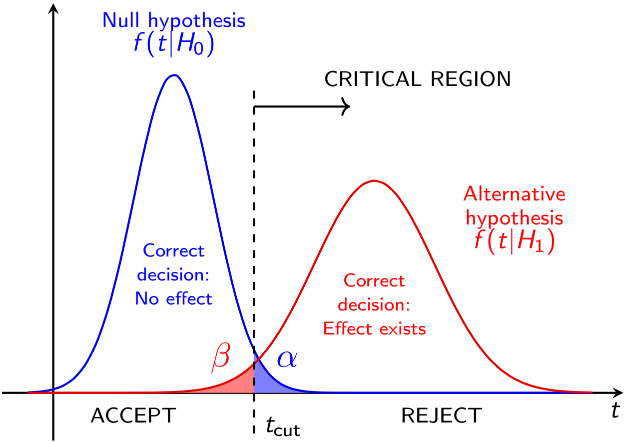
- The p-value is the probability of obtaining test results at least as extreme as the result actually observed, assuming the null hypothesis is correct.
- A very small p-value means that such an extreme observed outcome is very unlikely under the null hypothesis, so one can reject the null. (Wikipedia)



Review: Type I and Type II errors

- The value 1.96 results from building the interval with a confidence level of 95%.
  - Therefore, there is still a 5% probability that the true value is outside the interval.
- Similarly, we can make errors with hypothesis tests.
  - First, we can reject H_0 while it is true (since we left a 5% probability that it happens). We call this error a Type I error.
  - We can also not reject H_0 while it is false, that is, when one of the alternatives is true. This will be a Type II error.
- The probability of a Type II error depends on the particular alternative that is true in the set described by H_1.

Review: Type I and Type II errors



Type I error: reject H_0 while it is true.

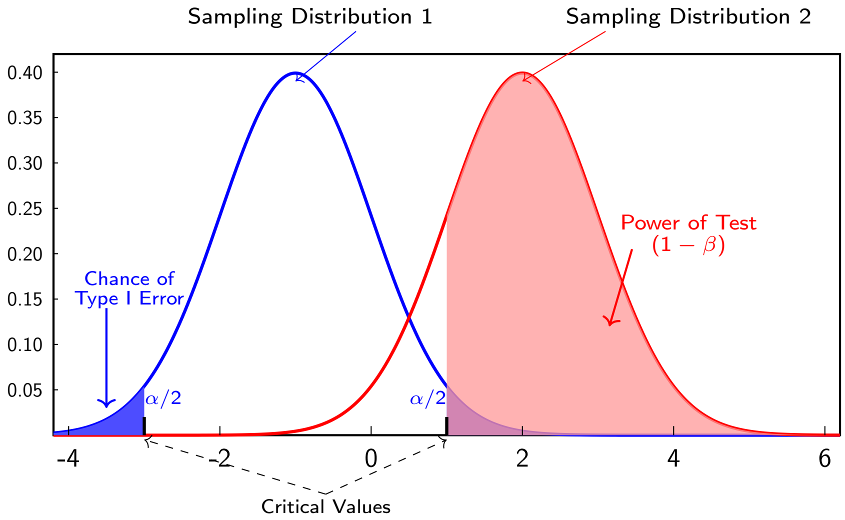
Type II error: not reject H_0 while it is false.

Review: Power of the test . . . I

- Ideally, we wish to minimize both error types, but this is impossible (if we reduce Type I error, we increase Type II error).
- The usual approach is to set Type I error at a certain level (usually at 5%), called the level of the test, and minimize Type II error.
- In fact, we maximize the power of the test, defined as.

$$\text{Power} = 1 - \text{Prob(Type II error)}$$

Review: Power of the test . . . II



Review: Power of the test ... III

- When drawing inferences using a test statistic, it is important to evaluate the power of the test statistic.
  - The power is the probability that the null hypothesis will be rejected given that an alternative hypothesis is true.
  - Low power indicates that the test is not useful to discriminate between the alternative and the null.
  - For a fixed value of N, considerable increases in power are possible with larger values of T.
  - The power gain is substantial when N is reduced for a fixed alternative.
  - But, when N is large and T is small, power will be low.

Example: Tests with standard Normal distribution . . . I

- Suppose we want to test if the mean of a population is equal to μ_0.
  - The null hypothesis is $H_0 : \mu = 15$.
  - The alternative hypothesis $H_1 : \mu \neq 15$.
- We formulate the hypothesis directly on the standardized statistic:

$$z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{T}}$$

- Suppose we find: $\bar{X} = 13, \sigma = 3.0456$, $T = 30$.
- For these values of $\bar{X}$ and σ and for the given null, the value for z is:

$$z = \frac{13 - 15}{3.0456/\sqrt{30}} = -3.597$$

Example: Tests with standard Normal distribution . . . II

- Should we accept or reject H_0 ?
- For α equal 0.05 :

$$P\left[-1.96 \leq \frac{\bar{X} - \mu_0}{\sigma/\sqrt{T}} \leq 1.96\right] = 0.95.$$

- We reject H_0 because the critical region is $z < -1.96$ and $z > 1.96$, and on the previous slide we calculated -3.597.
- The critical values depend on α and the degrees of freedom.

Example: Single-sided and two-sided test

- In the example above, the alternative hypothesis was of the form $H_1 : \mu \neq \mu_0$, which is
  - two-sided test because it involves the two tails of the distribution.
- If the alternative hypothesis was $H_1 : \mu < \mu_0$, we would have a
  - one-sided test because only the left tail would be relevant.
- In the previous example, for a one-sided test at 5%, the critical region will change from ± 1.96 to

$$z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} < -z_{5\%} = -1.645.$$

Example: Probability value (P-values)

- By convention, hypothesis tests have been carried out at levels of 1%,5% or 10%. However, a test can be carried out at any level.
- The probability value (p-value) is the lowest level at which a null hypothesis can be rejected.
- To find the p-value we evaluate the probability of the value found for z (-3.597 in the example) in the standard Normal distribution. $P[z < -3.597] = 0.000160954$.
- If we set α at 0.000160954, we could reject at this level the null hypothesis that $\mu_x = 15$.
- The advantage of reporting the p-value is that we avoid the arbitrariness of conventional levels, and we let the reader decide whether or not to reject a hypothesis.

Start of focus

Road map

Test of the difference in Sharpe ratios for Normal returns

- For evaluating the performance of a portfolio n relative to a benchmark m, we will compare the Sharpe ratios of n and m; i.e.,
  - instead of testing whether $\bar{X} = \mu_0$,
  - we will now test whether $\text{SR}_n - \text{SR}_m = 0$.
- So, compared to the earlier example,
  - $\bar{X}$ has been replaced by the difference in Sharpe ratios, $\text{SR}_n - \text{SR}_m$;
  - μ_0 has been replaced by 0.
- To test for the difference in Sharpe ratios, we will use two methods.
  1. "Delta method"
  2. "Bootstrap method"

Road map

The delta method

- The test for the difference in Sharpe ratios, using the delta method, was
  - developed by Jobson and Korkie (1981),
  - with a correction provided by Memmel (2003).
- Then, this test was extended to non-IID and non-Normal distributions by Lo (2002), Opdyke (2007), and Bailey and Lopez de Prado (2012).

Underlying logic of the Delta method

From: Wikipedia

- In statistics, the delta method is a method of deriving the asymptotic distribution of a random variable.
- It is applicable when the random variable being considered can be defined as a differentiable function of a random variable that is asymptotically Gaussian.
- The delta method allows us to approximate the distribution of a function of an estimator, assuming the estimator is asymptotically normal.

Central Limit Theorem (CLT) setup

- Let $\theta_0 \in \mathbb{R}^p$ and an estimator $\hat{\theta}$ satisfy a CLT:

  $$\sqrt{T}\,(\hat{\theta} - \theta_0) \;\xrightarrow{d}\; \mathcal{N}(0,\;\Omega)\,,$$

  with Ω positive semidefinite.

- Typically, $\hat{\theta}$ is a sample mean and Ω is a (long-run) variance.

- But, $\hat{\theta}$ could also be a set of moments, not just a single moment.

Univariate Delta Method (1st order)

- If $f : \mathbb{R}^p \to \mathbb{R}$ is a differentiable function at θ_0 with gradient $\nabla f(\theta_0)$, then

$$\sqrt{T}\,\{f(\hat{\theta}) - f(\theta_0)\} \xrightarrow{d} \mathcal{N}(0,\ V)\,, \quad \text{where:} \quad V \equiv \nabla f(\theta_0)^\top\, \Omega\, \nabla f(\theta_0).$$

- Hence, an approximate standard error is

$$\widehat{\mathrm{se}}\{f(\hat{\theta})\} = \sqrt{\widehat{V}/T},$$

where $\widehat{V}$ plugs in consistent estimates.

Obtaining p-values via the delta method

- Test $H_0 : f(\theta_0) = c \quad$ vs. $\quad H_1 : f(\theta_0) \neq c$.
- Statistic:

$$z \equiv \frac{f(\hat{\theta}) - c}{\widehat{\mathrm{se}}\{f(\hat{\theta})\}} \approx \mathcal{N}(0, 1).$$

- Two-sided p-value: $p = 2 \cdot \Phi(-|z|)$.
- One-sided variants are analogous.

Intuition underlying the delta method

- The intuition of the delta method is that any such f function, in a “small enough" range of the function, can be approximated via a first-order Taylor series (which is basically a linear function).
- If the random variable is roughly normal, then a linear transformation of it is also normal.
- A small range can be achieved when approximating the function around the mean, when the variance is “small enough".
- When f is applied to a random variable such as the mean, the delta method tends to work better as the sample size increases, because it helps reduce the variance of the mean.
- Thus, the Taylor approximation is applied over a smaller range of the function f at the point of interest.

Standardizing the difference in Sharpe ratios

- Recall that when testing whether $\bar{X} = \mu_0$, we constructed a standardized variable

$$z = \frac{\bar{X} - \mathbb{E}[\bar{X}]}{\sqrt{\mathbb{V}[\bar{X}]}} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{T}}.$$

- Now, to test whether $\mathrm{SR}_n - \mathrm{SR}_m = 0$, we will again need to construct a standardized variable

$$\begin{aligned} z &= \frac{(\mathrm{SR}_n - \mathrm{SR}_m) - 0}{\sqrt{\mathbb{V}[\mathrm{SR}_n - \mathrm{SR}_m]}} \\ &= \frac{(\mathrm{SR}_n - \mathrm{SR}_m)}{\sqrt{\mathbb{V}[\mathrm{SR}_n - \mathrm{SR}_m]}}. \end{aligned}$$

- So, the only thing that we need for the above test is the variance of the difference in the Sharpe ratios, $\mathbb{V}[\mathrm{SR}_n - \mathrm{SR}_m]$.

Difference in Sharpe ratios (Jobson-Korkie-Memmel)

- We start by writing down the final result so that we all know where we are going, after which we will see how this result is derived.

- The test statistic is

$$z = \frac{\widehat{\mathrm{SR}}_n - \widehat{\mathrm{SR}}_m}{\sqrt{\mathbb{V}\left[\widehat{\mathrm{SR}}_n - \widehat{\mathrm{SR}}_m\right]}},$$

where the variance of the difference in Sharpe ratios is

$$\boxed{\mathbb{V}\left[\widehat{\mathrm{SR}}_n - \widehat{\mathrm{SR}}_m\right] = \frac{1}{T}\left[2 - 2\hat{\rho}_{nm} + \frac{1}{2}\left(\widehat{\mathrm{SR}}_n^2 + \widehat{\mathrm{SR}}_m^2 - 2\widehat{\mathrm{SR}}_n\widehat{\mathrm{SR}}_m\hat{\rho}_{nm}^2\right)\right]}$$

in which $\hat{\rho}_{nm}$ denotes the estimated correlation between the excess returns of portfolios n and m.

Understanding the test for difference in Sharpe ratios

- To understand the test for the difference in Sharpe ratios of two portfolios, we will proceed in **three small steps** by studying, for IID normal returns, the:
  1. Distribution of μ_n and σ_n^2;
  2. Distribution of $SR_n = f(\mu_n, \sigma_n) = \mu_n/\sigma_n$;
  3. Distribution of $SR_n - SR_m = f(\mu_n, \mu_m, \sigma_n, \sigma_m)$;
- Then, after this discussion, we will study what to do if returns are not IID normal.

Step 1: Distribution of μ_n and σ_n^2 . . . I

- Denote by
  - $u = \{\mu_n, \sigma_n^2\}$ the mean and variance of the excess returns;
  - $\hat{u}$ the empirical counterpart of u, based on T observations;
- Then, the large-sample result for $u = \{\mu_n, \sigma_n^2\}$ is:

$$\sqrt{T}\left(\begin{bmatrix} \hat{\mu}_n \\ \hat{\sigma}_n^2 \end{bmatrix} - \begin{bmatrix} \mu_n \\ \sigma_n^2 \end{bmatrix}\right) \rightarrow \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_n^2 & 0 \\ 0 & 2\sigma_n^4 \end{bmatrix}\right).$$

- This can be written more compactly as:

$$\sqrt{T}(\hat{u} - u) \rightarrow \mathcal{N}(0_2, \Omega), \quad \text{with} \quad \Omega = \begin{bmatrix} \sigma_n^2 & 0 \\ 0 & 2\sigma_n^4 \end{bmatrix}.$$

Step 1: Distribution of μ_n and σ_n^2 . . . II

- In plain language, the above says that as T increases,
  - $\sqrt{T}(\hat{u} - u)$ converges to a Normal distribution
  - with a 2-dimensional mean vector of zeros, and
  - a (2×2)-dimensional covariance matrix given by Ω.
- Thus, the asymptotic estimation errors (variances) of $\hat{\mu}_n$ and $\hat{\sigma}_n^2$ are

$$\mathbb{V}[\hat{\mu}_n] \stackrel{a}{=} \frac{\sigma_n^2}{T}, \quad \text{and} \quad \mathbb{V}[\hat{\sigma}_n^2] \stackrel{a}{=} \frac{2\sigma_n^4}{T}.$$

- Thus, as T increases, estimation error decreases.
- For the derivation of the result that $\mathbb{V}[\hat{\sigma}_n^2] = 2\sigma_n^4/T$, look for "Distribution of the sample variance" on this Wikipedia page (toward the bottom of the page).

Step 2: Distribution of SR_n . . . I

- On the previous slide, we have derived the asymptotic distribution of $u = \{\mu_n, \sigma_n^2\}$.

- The Sharpe Ratio is a function of u:

$$\mathrm{SR}_n(u) = \mathrm{SR}_n(\mu_n, \sigma_n^2) = \frac{\mu_n}{\sqrt{\sigma_n^2}} = \frac{\mu_n}{\sigma_n}.$$

- Therefore, we need to find the asymptotic distribution of $\mathrm{SR}_n(u)$.

Step 2: Distribution of SR_n ... II

- Denote by $f(\cdot)$ a differentiable function of u.
- Then, as we have seen above, a result from statistics tells us

$$\sqrt{T}\big(f(\hat{u}) - f(u)\big) \rightarrow \mathcal{N}\left(0, f_u^\top \Omega\, f_u\right), \tag{1}$$

  where f_u is the first derivative (gradient) of $f(u)$ with respect to u.
- In plain language, the above says that as T increases,
  - $\sqrt{T}\big(f(\hat{u}) - f(u)\big)$ converges to a Normal distribution
  - with mean 0, and
  - variance $f_u^\top \Omega\, f_u$ ... we compute this on the next page.

Step 2: Distribution of SR_n ... III

- Recall that

$$u = \{\mu_n, \sigma_n^2\}. \tag{2}$$

- Set the function $f(\cdot)$ to be

$$f(u) = f(\mu_n, \sigma_n^2) = \mathrm{SR}_n(u) = \frac{\mu_n}{\sqrt{\sigma_n^2}}. \tag{3}$$

- Calculating the first partial derivatives of $f(u)$ (i.e., differentiating $f(u)$ in (3) with respect to the elements of u in (2)), we get

$$\frac{\partial f}{\partial u} = \begin{bmatrix} \frac{\partial f}{\partial \mu_n} \\ \frac{\partial f}{\partial \sigma_n^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_n} \\ -\frac{\mu_n}{2\sigma_n^3} \end{bmatrix}. \tag{4}$$

Step 2: Distribution of SR_n ... IV

- We wish to compute $f_u^\top \Omega\, f_u$, where

$$\frac{\partial f}{\partial u} = f_u = \begin{bmatrix} \frac{1}{\sigma_n} \\ -\frac{\mu_n}{2\sigma_n^3} \end{bmatrix} \quad \text{and} \quad \Omega = \begin{bmatrix} \sigma_n^2 & 0 \\ 0 & 2\sigma_n^4 \end{bmatrix}.$$

- So, we need to multiply: (i) the transpose of f_u; (ii) Ω; (iii) f_u:

$$\begin{aligned} f_u^\top \Omega\, f_u &= \begin{bmatrix} \frac{1}{\sigma_n} & -\frac{\mu_n}{2\sigma_n^3} \end{bmatrix} \cdot \begin{bmatrix} \sigma_n^2 & 0 \\ 0 & 2\sigma_n^4 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sigma_n} \\ -\frac{\mu_n}{2\sigma_n^3} \end{bmatrix} \\ &= 1 + \frac{1}{2}\frac{\mu_n^2}{\sigma_n^2} \\ &= 1 + \frac{1}{2}\mathrm{SR}_n^2. \end{aligned}$$

Step 2: Distribution of SR_n . . . V

- Then, using the result in (1) that

$$\sqrt{T}\left(f(\hat{u}) - f(u)\right) \rightarrow \mathcal{N}\left(0, f_u^\top \Omega\, f_u\right),$$

we get the expression we need:

$$\sqrt{T}\left(\widehat{\mathrm{SR}}_n - \mathrm{SR}_n\right) \rightarrow \mathcal{N}\left(0, 1 + \frac{1}{2}\mathrm{SR}_n^2\right).$$

- So the variance of the estimated Sharpe ratio is

$$\mathbb{V}[\widehat{\mathrm{SR}}_n] = \frac{1}{T}\left[1 + \frac{1}{2}\mathrm{SR}_n^2\right].$$

Step 2: Distribution of $\mathrm{SR}_n \ldots$ VI

- Thus, the standard error (SE) for the Sharpe-ratio estimator $\widehat{\mathrm{SR}}_n$ is

$$\mathrm{SE}\left[\widehat{\mathrm{SR}}_n\right] = \sqrt{\frac{1}{T}\left[1 + \frac{1}{2}\mathrm{SR}_n^2\right]}.$$

- Thus, the standardized z-statistic is

$$z = \frac{\mathrm{SR}_n - \mathrm{SR}_0}{\mathrm{SE}\left[\widehat{\mathrm{SR}}_n\right]} = \frac{\mathrm{SR}_n - \mathrm{SR}_0}{\sqrt{\frac{1}{T}\left[1 + \frac{1}{2}\mathrm{SR}_n^2\right]}} \approx \mathcal{N}(0,1).$$

- The 95% Confidence Interval (CI) for the Sharpe ratio is

$$\mathrm{CI} = \widehat{\mathrm{SR}}_n \pm 1.96 \times \sqrt{\frac{1}{T}\left[1 + \frac{1}{2}\mathrm{SR}_n^2\right]}.$$

- Robust variants replace σ^2/T by a HAC long-run variance for the mean and adjust the variance-of-variance accordingly.

Step 3: Distribution of $(\mathrm{SR}_n - \mathrm{SR}_m) \ldots \mathrm{I}$

- We now derive the distribution of the difference in the Sharpe ratios of two portfolio returns.
- Consider two portfolios with excess returns over the risk-free rate given by $R_{t,n}$ and $R_{t,m}$.
- Assume that these excess returns are IID (independently and identically distributed) Normal; that is, there is
  - no change in the distribution over time; i.e., the probability distribution of $R_{t,n}$ is the same as $R_{t',n}$ for any two dates t and t';
  - no serial dependence; i.e., $R_{t,n}$ and $R_{t',n}$ are independent.

Step 3: Distribution of $(\mathrm{SR}_n - \mathrm{SR}_m)$ . . . II

- Denote by
  - $u = \{\mu_n, \mu_m, \sigma_n^2, \sigma_m^2\}$ the means and variances of the excess returns;
  - $\hat{u}$ the empirical counterpart of u, based on T observations;
  - σ_{nm} the covariance between $R_{t,n}$ and $R_{t,m}$, and
  - $\rho_{nm} = \sigma_{nm}/(\sigma_n \sigma_m)$ the correlation between the excess returns.
- Note that now $u = \{\mu_n, \mu_m, \sigma_n^2, \sigma_m^2\}$ has four elements in it.

Step 3: Distribution of $(\mathrm{SR}_n - \mathrm{SR}_m)$ ...III

- Then, the large-sample result for $u = \{\mu_n, \mu_m, \sigma_n^2, \sigma_m^2\}$ is:

$$\sqrt{T}(\hat{u} - u) \to N(0_4, \Omega), \quad \text{with}$$

$$\Omega = \begin{bmatrix} \sigma_n^2 & \sigma_{nm} & 0 & 0 \\ \sigma_{nm} & \sigma_m^2 & 0 & 0 \\ 0 & 0 & 2\sigma_n^4 & 2\sigma_{nm}^2 \\ 0 & 0 & 2\sigma_{nm}^2 & 2\sigma_m^4 \end{bmatrix}$$

- In plain language, the above says that as T increases,
  - $\sqrt{T}(\hat{u} - u)$ converges to a Normal distribution
  - with a 4-dimensional mean vector of zeros, and
  - a (4×4)-dimensional covariance matrix given by Ω.

Step 3: Distribution of $(\mathrm{SR}_n - \mathrm{SR}_m) \ldots$ IV

- On the previous slide, we have derived the asymptotic distribution of $u = \{\mu_n, \mu_m, \sigma_n^2, \sigma_m^2\}$.
- We want to find the asymptotic distribution of the function:

$$f(u) = f(\mu_n, \mu_m, \sigma_n^2, \sigma_m^2) = \mathrm{SR}_n - \mathrm{SR}_m = \frac{\mu_n}{\sigma_n} - \frac{\mu_m}{\sigma_m}.$$

Step 3: Distribution of $(\mathrm{SR}_n - \mathrm{SR}_m)$ . . . V

- Let $f(\cdot)$ be a differentiable function of u, then a result from statistics tells us that

$$\sqrt{T}\big(f(\hat{u}) - f(u)\big) \to \mathcal{N}\left(0, f_u^\top \Omega\, f_u\right), \tag{5}$$

  where f_u is the first derivative of $f(u)$ with respect to u.

- In plain language, the above says that as T increases,
  - $\sqrt{T}\big(f(\hat{u}) - f(u)\big)$ converges to a Normal distribution
  - with mean 0, and
  - variance $f_u^\top \Omega\, f_u$ . . . we compute this on the next page.

Step 3: Distribution of $(\text{SR}_n - \text{SR}_m) \ldots$ VI

- Note that now we define

$$u = \{\mu_n, \mu_m, \sigma_n^2, \sigma_m^2\}. \tag{6}$$

- Set the function $f(\cdot)$ to be

$$f(u) = \text{SR}_n - \text{SR}_m = \frac{\mu_n}{\sigma_n} - \frac{\mu_m}{\sigma_m}. \tag{7}$$

- Calculating the first partial derivative of $f(u)$ (i.e., differentiating $f(u)$ in (7) with respect to the elements of u in (6)), we get

$$\frac{\partial f}{\partial u} = \begin{bmatrix} \frac{\partial f}{\partial \mu_n} \\ \frac{\partial f}{\partial \mu_m} \\ \frac{\partial f}{\partial \sigma_n^2} \\ \frac{\partial f}{\partial \sigma_m^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_n} \\ -\frac{1}{\sigma_m} \\ -\frac{\mu_n}{2\sigma_n^3} \\ \frac{\mu_m}{2\sigma_m^3} \end{bmatrix}. \tag{8}$$

Step 3: Distribution of $(\mathrm{SR}_n - \mathrm{SR}_m)$ ... VII

- We wish to compute $f_u^\top \Omega\, f_u$, where

$$\frac{\partial f}{\partial u} = f_u = \begin{bmatrix} \frac{1}{\sigma_n} \\ -\frac{1}{\sigma_m} \\ -\frac{\mu_n}{2\sigma_n^3} \\ \frac{\mu_m}{2\sigma_m^3} \end{bmatrix} \quad \text{and} \quad \Omega = \begin{bmatrix} \sigma_n^2 & \sigma_{nm} & 0 & 0 \\ \sigma_{nm} & \sigma_m^2 & 0 & 0 \\ 0 & 0 & 2\sigma_n^4 & 2\sigma_{nm}^2 \\ 0 & 0 & 2\sigma_{nm}^2 & 2\sigma_m^4 \end{bmatrix}.$$

- So, we need to multiply: (i) the transpose of f_u; (ii) Ω; (iii) f_u:

$$f_u^\top \Omega\, f_u = \begin{bmatrix} \frac{1}{\sigma_n} & -\frac{1}{\sigma_m} & -\frac{\mu_n}{2\sigma_n^3} & \frac{\mu_m}{2\sigma_m^3} \end{bmatrix} \cdot \begin{bmatrix} \sigma_n^2 & \sigma_{nm} & 0 & 0 \\ \sigma_{nm} & \sigma_m^2 & 0 & 0 \\ 0 & 0 & 2\sigma_n^4 & 2\sigma_{nm}^2 \\ 0 & 0 & 2\sigma_{nm}^2 & 2\sigma_m^4 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sigma_n} \\ -\frac{1}{\sigma_m} \\ -\frac{\mu_n}{2\sigma_n^3} \\ \frac{\mu_m}{2\sigma_m^3} \end{bmatrix}$$

$$= 2 - 2\rho_{nm} + \frac{1}{2}\left(\mathrm{SR}_n^2 + \mathrm{SR}_m^2 - 2\mathrm{SR}_n\mathrm{SR}_m\rho_{nm}^2\right), \quad \text{where } \rho_{nm} = \sigma_{nm}/(\sigma_n\sigma_m).$$

Step 3: Distribution of $(\mathrm{SR}_n - \mathrm{SR}_m)$ ...VIII

- Then, using the result in (5) that

$$\sqrt{T}(f(\hat{u}) - f(u)) \to N\left(0, f_u^\top \Omega\, f_u\right), \tag{9}$$

we get the expression we need:

$$\sqrt{T}\left((\widehat{\mathrm{SR}}_n - \widehat{\mathrm{SR}}_m) - (\mathrm{SR}_n - \mathrm{SR}_m)\right) \to N\left(0, f_u^\top \Omega\, f_u\right),$$

- so the variance of the difference in Sharpe ratios is

$$\begin{aligned}
\mathbb{V}\left[\widehat{\mathrm{SR}}_n - \widehat{\mathrm{SR}}_m\right] &= \frac{1}{T}\left(\frac{\partial f}{\partial u}\right)^\top \Omega \left(\frac{\partial f}{\partial u}\right) \\
&= \frac{1}{T}\left[2 - 2\rho_{nm} + \frac{1}{2}\left(\mathrm{SR}_n^2 + \mathrm{SR}_m^2 - 2\mathrm{SR}_n\mathrm{SR}_m\rho_{nm}^2\right)\right] \\
&= \frac{1}{T}\left[\left(1 + \frac{1}{2}\mathrm{SR}_n^2\right) + \left(1 + \frac{1}{2}\mathrm{SR}_m^2\right)\right] \quad \ldots \text{if } \rho_{nm} = 0.
\end{aligned}$$

Step 3: Distribution of $(\mathrm{SR}_n - \mathrm{SR}_m) \ldots$ IX

- Just as we did for the Sharpe ratio, we can use the expression for $\mathbb{V}\left[\widehat{\mathrm{SR}}_n - \widehat{\mathrm{SR}}_m\right]$ on the previous page to:
  - Compute the Standard Error for the difference in Sharpe ratios;
  - Construct the Confidence Interval for the difference in Sharpe ratios.
  - Construct the p-value for the difference in Sharpe ratios.

Road map

Limitations of the delta method

- The test of Jobson and Korkie (1981) and Memmel (2003) is not valid if
  - **Non-normality:** Returns are heavy-tailed and skewed.
  - **Time dependence:** Autocorrelation and conditional heteroskedasticity (volatility clustering).
  - **Correlation across strategies:** Returns of the two strategies are not independent (e.g., because of common risk factors).
- The problems that arise with non-Normal returns are illustrated in this excellent article by QuantPy, which includes Python code.
  - I strongly recommend that you read this article.
- An alternative approach to the delta method is bootstrapping.

Bootstrapping

From Wikipedia

- Bootstrapping is a procedure for estimating the distribution of an estimator by resampling one's data (usually with replacement).
- Bootstrapping allows estimation of the sampling distribution of almost any statistic using random sampling methods.
- The bootstrap is often used as an alternative to statistical inference based on a parametric model when
  - the assumptions of the parametric model are in doubt, or
  - when parametric inference is impossible, or
  - the parametric model requires complex formulas to calculate standard errors.

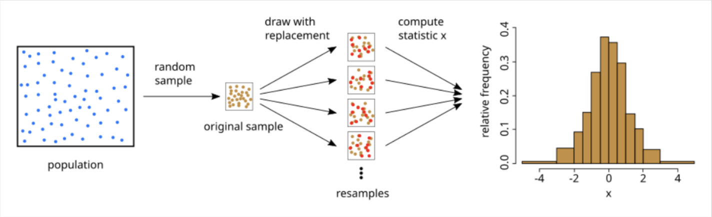
Bootstrap: Intuitive explanation . . . I

- Suppose we are interested in the average height of people in France.
- We sample only a small part—of size N—of the French population.
- From that single sample, we obtain only a single estimate of the mean.
- In order to reason about the population, we need some sense of the variability of the mean that we have computed.
- The simplest bootstrap method involves taking the original data set of heights, and sampling from it to form a new sample (called a 'resample' or 'bootstrap sample') that is also of size N.
- The bootstrap sample is taken from the original by using sampling with replacement (e.g. we might 'resample' 5 times from [1,2,3,4,5] and get [2,5,4,4,1]).

Bootstrap: Intuitive explanation . . . II

- Assuming N is sufficiently large, there is virtually zero probability that it will be identical to the original “real" sample.
- This resampling process is repeated a large number of times (typically 1,000 to 10,000), and for each bootstrap sample, we compute its mean (each of which is called a "bootstrap estimate").
- We can now create a histogram of bootstrap means.
- This histogram provides an estimate of the distribution of the sample mean, from which we can answer questions about how much the mean varies across samples.
- The method described here for the mean can be applied to almost any other statistic or estimator, not just the mean.

Bootstrap: Illustrated



A sample is drawn from a population. From this sample, resamples are generated by drawing with replacement (orange). Data points that were drawn more than once (which happens for approx. 26.4% of data points) are shown in red and slightly offsetted. From the resamples, the statistic x is calculated and, therefore, a histogram can be calculated to estimate the distribution of x.

From: Wikipedia

Bootstrapping: Additional resources

- To learn more about bootstrap sampling, see Introduction to bootstrap sampling, which includes Python code.
- Other articles explaining bootstrapping along with Python code:
  - A few examples of bootstrapping
  - Ditch p-values. Use Bootstrap confidence intervals instead.
- The definitive guide to bootstrapping is the book by Efron and Tibshirani (1994).

Block bootstrap

- The block bootstrap is used when the data (or the errors in a model) are correlated, as is often the case for financial data.
  - In this case, simple resampling will fail, because it is not able to replicate the correlation in the data.
  - The block bootstrap tries to replicate the correlation by resampling inside blocks of data. For additional details, see this article.
  - The block bootstrap has been used mainly with data correlated in time (i.e., time series) but can also be used with data correlated in space, or among groups (so-called cluster data).

Block bootstrap: Motivation

- **The problem with time-series data**
  - The classical bootstrap doesn't apply to time series because time series data has an inherent order to it that i.i.d. methods don't account for.
  - Each data point in a time series depends on previous data points.
    1. So we need a way to incorporate dependency and preserve the order.
    2. And we also want our samples to be random.
  - To achieve both objectives, we divide the data into a series of consecutive blocks, each containing a few observations, in order, and use a bootstrap on the blocks.

Block bootstrap: Details

- Suppose we have the series $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}$.
- We first need to determine how many observations to put in each group. Let's say we will use two data points in each block.
- Then our blocked data would look like

$$\overbrace{X_1, X_2}^{block1}, \quad \overbrace{X_3, X_4}^{block2}, \quad \overbrace{X_5, X_6}^{block3}, \quad \overbrace{X_7, X_8}^{block4}, \quad \overbrace{X_9, X_{10}}^{block5}.$$

- Then, we take a random sample of the blocks with replacement.
- A bootstrapped series of blocks might be

  block 3, block 1, block 5, block 2, block 5

- This has the effect of giving us a new series with the same (short term) dependence structure.

Moving block bootstrap

- An extension of the block bootstrap is the moving block bootstrap (MBB) where the series is split into overlapping blocks.
  The blocks would therefore look like this:

$$
\begin{aligned}
\text{block 1:} &\quad X_1, X_2 \\
\text{block 2:} &\quad X_2, X_3 \\
\text{block 3:} &\quad X_3, X_4 \\
\text{block 4:} &\quad X_4, X_5 \\
\vdots \quad &\quad \vdots \quad \vdots \\
\text{block 9:} &\quad X_9, X_{10}.
\end{aligned}
$$

- Once the blocks are defined, we can then take a bootstrap sample of the blocks as before.

The stationary bootstrap

- The stationary bootstrap is another variant of the block bootstrap with a different sort of randomness.
- The stationary bootstrap uses a random block length.
- An additional source of randomness comes from selecting the starting points for the blocks at random.

Circular block bootstrap

- Circular block bootstrapping treats data as circular, enabling blocks to wrap around from the end of the series back to the beginning.
- Thus, points on the edges have the same probability of being sampled as points in the middle.
  - For the paper that developed the circular block bootstrap, see Politis and Romano (1994).
  - For a nice explanation with pictures, see this article.
  - For Python code see this link.
- Ledoit and Wolf (2008) apply the circular block bootstrap to study the difference in Sharpe ratios, which is what we wish to study.

Bootstrap *p*-values: Ledoit and Wolf (2008)

- Ledoit and Wolf (2008) construct a bootstrap confidence interval for the difference in Sharpe ratios, which does not require particular assumptions regarding the return distributions.
  - Link to the paper.
  - Link to code from Michael Wolf's page (in Matlab and R).
  - Python code for the bootstrap method in Ledoit and Wolf (2008) is available at RobustSharpeRatioHAC by Michael Mark.

Bootstrap p-values for ΔSR

- **Hypotheses:**

$$H_0 : SR_n = SR_m \quad \text{vs.} \quad H_1 : SR_n \neq SR_m.$$

- **Define**

$$\Delta SR = SR_n - SR_m, \qquad \text{with estimated values:} \quad \widehat{\Delta SR} = \widehat{SR}_n - \widehat{SR}_m.$$

- **Test statistic:** A robust test statistic is a studentized statistic:

$$\mathcal{S} = \frac{\widehat{\Delta SR}}{\widehat{\mathrm{SE}}(\widehat{\Delta SR})},$$

where $\widehat{\mathrm{SE}}(\widehat{\Delta SR})$ is a heteroskedasticity/autocorrelation consistent (HAC) estimate or a bootstrap standard error.

Resampling design and p-value

- **Block bootstrap (time-series robust):** Resample pairs of the two returns in blocks of data to preserve serial dependence. Common choices:
  - fixed-length blocks,
  - stationary bootstrap with random block lengths, or
  - circular blocks (this is what I recommend).
- Compute bootstrap replicates $\mathcal{S}_b^*$ of the statistic under a resampling scheme that respects the null or, more practically, for SR differences, under the empirical distribution.
- Then, the two-sided p-value is

$$\hat{p} = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(|\mathcal{S}_b^*| \geq |\mathcal{S}_{\text{obs}}|\right).$$

Bootstrapping: Practical recipe

1. Compute $\widehat{SR}_A, \widehat{SR}_B$ and $\mathcal{S}_{\text{obs}} = \widehat{\Delta SR}/\widehat{\mathrm{SE}}$.
2. Choose block length ℓ (rule-of-thumb: $\ell \propto T^{1/3}$) or use stationary bootstrap with mean length ℓ.
3. For $b = 1, \ldots, B$:
   - 3.1 Draw a bootstrap time series of pairs $\{(R_{t,b}^{n*}, R_{t,b}^{m*})\}_{t=1}^{T}$ by concatenating random blocks from the original series.
   - 3.2 Compute $\widehat{SR}_n^{*(b)}$, $\widehat{SR}_m^{*(b)}$, $\widehat{\Delta SR}^{*(b)}$.
   - 3.3 Studentize: $\mathcal{S}_b^* = \widehat{\Delta SR}^{*(b)}/\widehat{\mathrm{SE}}^{*(b)}$, where $\widehat{\mathrm{SE}}^{*(b)}$ is computed on the bootstrap sample in the same way as in the original sample.
4. Estimated p-value: $\hat{p} = \frac{1}{B}\sum_b \mathbb{I}(|\mathcal{S}_b^*| \geq |\mathcal{S}_{\text{obs}}|)$.

End of focus

Road map

Sample Python code

- The next few slides contain sample Python code.
- Studying this code is optional.
- This code is provided without any guarantees.
- Feel free to play with the code.
- If you like fun and adventure,
  - please reproduce Tables 1 and 2 of Ledoit and Wolf (2008)
  - using your own code or code written by ChatGPT or someone else.

Python code for bootstrapping the mean and percentile CI

Code for bootstrapping the mean and a percentile CI

```python
import numpy as np
from numpy.random import default_rng   # rng = random-number-generator

rng = default_rng(42)
n = 200
data = rng.normal(loc=0.0, scale=1.0, size=n)

B = 5000
boot_means = np.empty(B)
for b in range(B):
    sample = rng.choice(data, size=n, replace=True)
    boot_means[b] = sample.mean()

se_mean = boot_means.std(ddof=1)
ci_mean = np.quantile(boot_means, [0.025, 0.975])

print(data.mean(), se_mean, ci_mean)
```

Python code for bootstrapping a quantile

Code for bootstrapping a quantile

```python
alpha = 0.05
boot_q = np.empty(B)
for b in range(B):
    sample = rng.choice(data, size=n, replace=True)
    boot_q[b] = np.quantile(sample, alpha)

se_q = boot_q.std(ddof=1)
ci_q = np.quantile(boot_q, [0.025, 0.975])

print(np.quantile(data, alpha), se_q, ci_q)
```

Python code for BCa interval . . . I

BCa interval

```
from math import isfinite
from scipy.stats import norm

def bca_interval(x, stat_fn, conf=0.95, B=5000, rng=None):
    if rng is None:
        rng = default_rng()
    x = np.asarray(x)
    n = x.size
    t0 = stat_fn(x)

    # 1) Bootstrap distribution
    boots = np.empty(B)
    for b in range(B):
        boots[b] = stat_fn(rng.choice(x, size=n, replace=True))

    # 2) Bias-correction z0
    prop = np.mean(boots < t0)  # strictly less than t0
    prop = min(max(prop, 1e-12), 1-1e-12)
    z0 = norm.ppf(prop)

    # 3) Jackknife for acceleration a       ... contd. on next page
```

Python code for BCa interval . . . II

```
    jack = np.empty(n)
    for i in range(n):
        jack[i] = stat_fn(np.delete(x, i))
    jack_mean = jack.mean()
    num = np.sum((jack_mean - jack)**3)
    den = np.sum((jack_mean - jack)**2)**1.5 + 1e-12
    a = (1/6) * (num / den)

    # 4) Adjusted alpha levels
    alpha_lo, alpha_hi = (1 - conf)/2, (1 + conf)/2
    z_lo, z_hi = norm.ppf(alpha_lo), norm.ppf(alpha_hi)
    adj_lo = norm.cdf(z0 + (z_lo - z0)/(1 - a*(z_lo - z0)))
    adj_hi = norm.cdf(z0 + (z_hi - z0)/(1 - a*(z_hi - z0)))

    # 5) Quantiles of bootstrap distribution
    lo, hi = np.quantile(boots, [adj_lo, adj_hi])
    return lo, hi

# Example: BCa for the mean
bca_interval_mean = bca_interval(data, np.mean, conf=0.95, B=5000, rng=
     rng)
print(bca_interval_mean)
```

Code for Sharpe ratio and IID and block bootstrap . . . I

Code for Sharpe ratio, and for IID and block bootstrap

```
def sharpe(x):
    x = np.asarray(x)
    s = x.std(ddof=1)
    return np.nan if s == 0 else x.mean()/s

def iid_pairs_bootstrap(a, b, rng=None):
    if rng is None:
        rng = default_rng()
    n = len(a)
    idx = rng.integers(0, n, size=n)
    return a[idx], b[idx]

def block_bootstrap_pairs(a, b, ell, rng=None):
    """
    Moving-block bootstrap for paired series a,b (same length).
    Concatenate random starting points of length ell, then truncate to n
     .
    """
    if rng is None:
        rng = default_rng()
```

Code for Sharpe ratio and IID and block bootstrap . . . II

```
    n = len(a)
    num_blocks = int(np.ceil(n/ell))
    starts = rng.integers(0, n-ell+1, size=num_blocks)
    idx = np.concatenate([np.arange(s, s+ell) for s in starts])[:n]
    return a[idx], b[idx]
```

Python code for HAC-style SE for ΔSR ... I

Code for HAC-style SE for ΔSR

```
def _newey_west_variance_of_mean(series, max_lag=None):
    """
    NW variance for the sample mean of a (possibly) dependent series.
    Returns var(mean(series)).
    """
    x = np.asarray(series)
    n = x.size
    x = x - x.mean()
    if max_lag is None:
        # Same default spirit as Mathematica code: ~ 4*(n/100)^(2/9)
        max_lag = int(np.floor(4*(n/100)**(2/9)))
    # gamma_0
    gamma0 = np.dot(x, x)/n
    # auto-covariances (population-scale)
    def cov_k(k):
        return np.dot(x[:-k], x[k:])/(n - k)
    w = lambda k: 1 - k/(max_lag + 1.0)
    s = gamma0 + 2.0*sum(w(k)*cov_k(k) for k in range(1, max_lag+1))
    return s / n   #                ... continued on the next page
```

Python code for HAC-style SE for ΔSR ... II

```
def delta_sr_se_hac(a, b, max_lag=None):
    """
    HAC (NW) SE for Delta SR using an influence-function-style proxy.
    """
    a = np.asarray(a); b = np.asarray(b)
    muA, muB = a.mean(), b.mean()
    sA, sB = a.std(ddof=1), b.std(ddof=1)
    if sA == 0 or sB == 0:
        return np.nan
    # Influence-proxy (same form as Mathematica comments)
    zA = (a - muA)/sA
    zB = (b - muB)/sB
    d = (a - muA)/sA - (b - muB)/sB \
        - (muA/(2*sA**3))*((a - muA)**2 - sA**2) \
        + (muB/(2*sB**3))*((b - muB)**2 - sB**2)

    var_mean = _newey_west_variance_of_mean(d, max_lag=max_lag)
    se = np.sqrt(var_mean)
    return se
#                                   ... continued on the next page
```

Python code for HAC-style SE for ΔSR ... III

```
def delta_sr_test_statistic(a, b, se_mode="HAC"):
    dSR = sharpe(a) - sharpe(b)
    if se_mode == "HAC":
        se = delta_sr_se_hac(a, b)
    else:
        se = delta_sr_se_hac(a, b)  # placeholder, same as Mathematica
     note
    T = np.nan if (se is None or not np.isfinite(se) or se == 0) else
     dSR/se
    return dSR, se, T
```

Python code for block bootstrap p-value . . . I

Code for block bootstrap p-value

```
def delta_sr_pvalue_block(a, b, ell, B=5000, se_mode="HAC", rng=None):
    if rng is None:
        rng = default_rng()
    dObs, seObs, TObs = delta_sr_test_statistic(a, b, se_mode)
    TBoot = []
    for _ in range(B):
        rA, rB = block_bootstrap_pairs(a, b, ell, rng=rng)
        seB = delta_sr_se_hac(rA, rB)
        dB = sharpe(rA) - sharpe(rB)
        if np.isfinite(seB) and seB != 0:
            TBoot.append(dB/seB)
    TBoot = np.asarray(TBoot)
    if TBoot.size == 0 or not np.isfinite(TObs):
        return np.nan
    return np.mean(np.abs(TBoot) >= np.abs(TObs))
#                                ... continued on the next page
```

Python code for block bootstrap p-value . . . II

```
# Example with simple AR(1)-like dependence
rng = default_rng(23)
n = 2000
phiA = 0.2
phiB = 0.1
epsA = rng.normal(0, 0.01, size=n)
epsB = rng.normal(0, 0.01, size=n)

retA = np.zeros(n)
retB = np.zeros(n)
for t in range(1, n):
    retA[t] = phiA*retA[t-1] + epsA[t]
    retB[t] = phiB*retB[t-1] + epsB[t]

ell = max(5, round(n**(1/3)))
pval_block = delta_sr_pvalue_block(retA, retB, ell, B=2000, se_mode="HAC
    ", rng=rng)
print("Block p-value:", pval_block)
```

Python code for stationary bootstrap . . . I

Code for stationary bootstrap

```
def stationary_bootstrap_pairs(a, b, mean_block_len, rng=None):
    """
    Stationary bootstrap for paired series (Politis & Romano).
    mean_block_len -> geometric restart prob p = 1/mean_block_len.
    """
    if rng is None:
        rng = default_rng()
    a = np.asarray(a); b = np.asarray(b)
    n = len(a)
    p = 1.0/float(mean_block_len)
    idx = np.empty(n, dtype=int)
    idx[0] = rng.integers(0, n)
    for t in range(1, n):
        if rng.random() < p:
            idx[t] = rng.integers(0, n)   # start new block
        else:
            idx[t] = (idx[t-1] + 1) % n   # continue (circular)
    return a[idx], b[idx]
#                                ... continued on the next page
```

Python code for stationary bootstrap . . . II

```
# Example usage (drop-in replacement inside the bootstrap loop)
def delta_sr_pvalue_stationary(a, b, mean_block_len, B=5000, se_mode="
    HAC", rng=None):
    if rng is None:
        rng = default_rng()
    dObs, seObs, TObs = delta_sr_test_statistic(a, b, se_mode)
    TBoot = []
    for _ in range(B):
        rA, rB = stationary_bootstrap_pairs(a, b, mean_block_len, rng=
    rng)
        seB = delta_sr_se_hac(rA, rB)
        dB = sharpe(rA) - sharpe(rB)
        if np.isfinite(seB) and seB != 0:
            TBoot.append(dB/seB)
    TBoot = np.asarray(TBoot)
    if TBoot.size == 0 or not np.isfinite(TObs):
        return np.nan
    return np.mean(np.abs(TBoot) >= np.abs(TObs))
```

Road map

What we plan to do in the next chapter

What's next?

In the next chapter, we will study classical mean-variance portfolios that *ignore* estimation error.

To do for next class

- Readings
  - Please read the section on "Working with Data" by Moreira (2021, Chapter 5) available online from this link.
- Assignment
  - You should start working on the first assignment.

Road map

Bibliography . . . I

Bailey, D. H., and M. Lopez de Prado. 2012. The Sharpe ratio efficient frontier. *Journal of Risk* 15 (2): 13. (Cited on page 83).

Balduzzi, P., and A. W. Lynch. 1999. Transaction costs and predictability: Some utility cost calculations. *Journal of Financial Economics* 52 (1): 47–78. (Cited on page 56).

Bhardwaj, R. K., and L. D. Brooks. 1992. The january anomaly: effects of low share price, transaction costs, and bid-ask bias. *Journal of Finance* 47 (2): 553–575. (Cited on page 56).

Campbell, J. Y., and L. M. Viceira. 2002. *Strategic asset allocation: Portfolio choice for long-term investors.* New York: Oxford University Press. (Cited on page 11).

Carhart, M. M. 1997. On persistence in mutual fund performance. *Journal of Finance* 52, no. 1 (March): 57–82. (Cited on page 27).

Connor, G., L. Goldberg, and R. A. Korajczyk. 2010. *Portfolio risk analysis.* Princeton, New Jersey: Princeton University Press. (Cited on page 29).

Dello-Preite, M., R. Uppal, P. Zaffaroni, and I. Zviadadze. 2025. Cross-sectional asset pricing with unsystematic risk. Available at SSRN 4135146. (Cited on page 30).

DeMiguel, V., A. Martín-Utrera, F. J. Nogales, and R. Uppal. 2020. A Transaction-Cost Perspective on the Multitude of Firm Characteristics. *Review of Financial Studies* 33, no. 5 (April): 2180–2222. (Cited on page 58).

Bibliography . . . II

DeMiguel, V., A. Martín-Utrera, and R. Uppal. 2024. A multifactor perspective on volatility-managed portfolios. *Journal of Finance* 79 (6): 3859–3891. (Cited on page 58).

DeMiguel, V., and R. Uppal. 2005. Portfolio investment with the exact tax basis via nonlinear programming. *Management Science* 51, no. 2 (February): 277–290. (Cited on page 59).

Dybvig, P., and H. K. Koo. 1996. Investment with taxes. Working Paper, Washington University. (Cited on page 59).

Efron, B., and R. J. Tibshirani. 1994. *An introduction to the bootstrap.* CRC press. (Cited on page 115).

Fama, E. F., and K. R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (1): 3–56. (Cited on page 27).

———. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116 (1): 1–22. (Cited on page 28).

———. 2018. Choosing factors. *Journal of Financial Economics* 128 (2): 234–252. (Cited on page 28).

French, K. R. 2008. Presidential address: The cost of active investing. *Journal of Finance* 63, no. 4 (August): 1537–1573. (Cited on page 56).

Bibliography . . . III

Hou, K., H. Mo, C. Xue, and L. Zhang. 2019. Which factors? *Review of Finance* 23 (1): 1–35. (Cited on page 28).

Hou, K., C. Xue, and L. Zhang. 2015. Digesting anomalies: An investment approach. *Review of Financial Studies* 28 (3): 650–705. (Cited on page 28).

———. 2017a. A comparison of new factor models. *Fisher College of Business Working Paper,* nos. 2015-03: 05. (Cited on page 28).

———. 2017b. Replicating anomalies. *The Review of Financial Studies.* (Cited on page 28).

Jobson, J. D., and B. M. Korkie. 1981. Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance:* 889–908. (Cited on pages 83, 110).

Kozak, S., S. Nagel, and S. Santosh. 2018. Interpreting factor models. *Journal of Finance* 73 (3): 1183–1223. (Cited on page 29).

———. 2020. Shrinking the cross-section. *Journal of Financial Economics* 135 (2): 271–292. (Cited on page 29).

Ledoit, O., and M. Wolf. 2008. Robust performance hypothesis resting with the Sharpe ratio. *Journal of Empirical Finance* 15:850–859. (Cited on pages 121, 122, 128).

Lesmond, D. A., J. P. Ogden, and C. A. Trzcinka. 1999. A new estimate of transaction costs. *Review of Financial Studies* 12 (5): 1113–1141. (Cited on page 56).

Bibliography ... IV

Lo, A. W. 2002. The statistics of Sharpe ratios. *Financial Analysts Journal* 58 (4): 36–52. (Cited on pages 39, 83).

Memmel, C. 2003. Performance hypothesis testing with the Sharpe ratio. Available at SSRN 412588. (Cited on pages 83, 110).

Merton, R. C. 1971. Optimum consumption and portfolio rules in a continuous time model. *Journal of Economic Theory* 3:373–413. (Cited on page 11).

———. 1990. *Continuous time finance.* Cambridge MA: Basil Blackwell. (Cited on page 11).

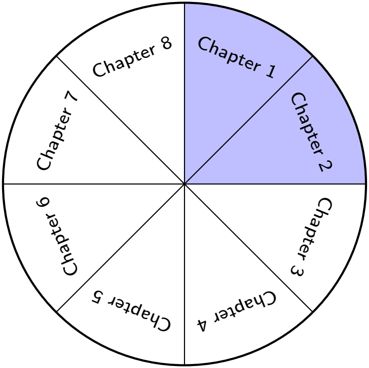
Moreira, A. 2021. *Quantitative investing.* Available online from this link. (Cited on page 144).

Opdyke, J. D. 2007. Comparing Sharpe ratios: So where are the p-values? *Journal of Asset Management* 8 (5): 308–336. (Cited on page 83).

Politis, D. N., and J. P. Romano. 1994. The stationary bootstrap. *Journal of the American Statistical Association* 89 (428): 1303–1313. (Cited on page 121).

Sharpe, W. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 1919 (3): 425–442. (Cited on page 27).

Stoll, H. R., and R. E. Whaley. 1983. Transaction costs and the small firm effect. *Journal of Financial Economics* 12 (1): 57–79. (Cited on page 56).



End of Chapter 2