Please answer as many questions marked by "Q" as you can. For questions that involve coding, please use the programming language of your choice (preferably Python), and submit your code together with the rest of the problem set.

1 Test of Methodology and Implementation

1.1 Problem setup

Assume we have a portfolio of N assets indexed by $i = 1, \ldots, N$ with return r_i, volatility σ_i, and correlation ρ_{ij} between the returns of assets i and j. If the portfolio's weight on asset i is ω_i, the portfolio return will be

$$r_p = \sum_i w_i r_i \tag{1}$$

Q1. What is the volatility of the portfolio σ_p expressed in the notation above?

Define Sharpe ratio of the portfolio as $SR_p = \dfrac{r_p}{\sigma_p}$, ignoring the risk-free rate.

In addition, the portfolio is assigned an exogenous benchmark (BM). We define the tracking error (TE) of the portfolio with respect to BM to be the standard deviation of the return difference, i.e.,

$$TE_p^2 = \mathrm{Var}(r_p - r_{\mathrm{BM}}) = \sigma_p^2 + \sigma_{\mathrm{BM}}^2 - 2\rho_{p,\mathrm{BM}}\sigma_p\sigma_{\mathrm{BM}} \tag{2}$$

1.2 Marginal contribution

If we define the marginal contribution (MC) of each asset to the portfolio metrics as

- MC to return: $MCR_i = \dfrac{\partial r_p}{\partial \omega_i}$
- MC to volatility: $MCV_i = \dfrac{\partial \sigma_p}{\partial \omega_i}$
- MC to Sharpe ratio: $MCS_i = \dfrac{\partial SR_p}{\partial \omega_i}$
- MC to tracking error: $MCT_i = \dfrac{\partial TE_p}{\partial \omega_i}$

then

Q2. What is the formula for MCR_i?

Q3. What is the formula for MCV_i?

Q4. What is the formula for MCS_i?

Q5. What is the formula for MCT_i?

1.3 Minimum Volatility Portfolio (MVP)

Given volatility σ_i and correlation ρ_{ij}, we would like to find the weight vector $\{\omega_1, \ldots, \omega_N\}$ such that the volatility of the portfolio σ_p (as derived in **Q1**) is minimized while satisfying $0 \leq \omega_i \leq \omega_{\max}, \forall i$ and $\sum_i \omega_i = 1$.

Q6. Provide a working implementation for calculating MVP using the programming language of your choice (preferably Python).

1.4 Minimum Tracking Error Portfolio

Given volatility σ_i and correlation ρ_{ij}, we would like to find the weight vector $\{\omega_1, \ldots, \omega_N\}$ such that TE_p (as derived in Equation 2) is minimized while satisfying $0 \leq \omega_i \leq \omega_{\max}, \forall i$ and $\sum_i \omega_i = 1$.

Q7. Provide a working implementation using the programming language of your choice.

1.5 Connection between min-TE portfolio and linear regression

In practice we don't know the true volatility σ_i or correlation ρ_{ij}, and they have to be estimated from data. Suppose we denote the historical daily return of asset i at time t as $x_{i,t}$ and the return of BM at time t as y_t. After obtaining the sample variance-covariance from the data, we can plug in the sample value to get the minimum TE Portfolio using the procedure derived in Q7.

Another way to approach this is simply running a constrained linear regression of y on x, where the linear coefficient β is constrained to be within $[0, \omega_{\max}]$.

Q8. What is the connection between the minimum TE portfolio procedure you derived above and this linear regression procedure? Describe the similarities and differences. Under what scenarios will one be better than the other? Use both mathematics and intuitive reasoning to narrate the answer.

2 Test of Analyzing Financial Data

The data set in *postion_trade.xlsx* contains information of a Long-Only Equity manager. Please refer to Data Field Glossary for detailed explanation of each column.

Q9. Please use the 'Fund Level Data' sheet to answer the questions below

(a) Calculate the following portfolio stats

- Total return in %
- Total P&L
- Annualized Sharpe ratio
- Simple annualized return in bps

(b) Generate charts to show

- Compounding cumulative return
- Net asset value and cash transfer over time

Do you spot any bad data or abnormal behavior? If so, can you fix the problem or explain the discrepancy?

Q10. Please use the 'Fund Level Data' and 'Holdings Data' sheets to answer the questions below

(a) Calculate the following portfolio stats

- Range of position counts
- Range of position weights
- Range of fund total exposure

(b) Generate charts to show

- Position counts over time
- Position weights distribution
- Fund total exposure over time

Do you spot any bad data or abnormal behavior? If so, can you fix the problem or explain the discrepancy?

Q11. Please use the 'Holdings Data' and 'Trades Data' sheets to answer the questions below

(a) Do the quantities reconcile between the two sheets? If not, can you fix the problem or explain the discrepancy?

(b) Do the trade prices reconcile between the two sheets? If not, can you fix the problem or explain the discrepancy?

Q12. Please use the 'Holdings Data' and 'Trades Data' sheets to answer the questions below

(a) Calculate the following portfolio stats

- Total return in %
- Total P&L
- Annualized Sharpe ratio
- Simple annualized return in bps

(b) Generate charts to show

- Compounding cumulative return from trade P&L and compare with the one calculated from **Q9**
- Number of trades over time

Do you spot any bad data or abnormal behavior? If so, can you fix the problem or explain the discrepancy?

A Data Field Glossary

- Fund Level Data
  - Net Asset Value(NAV): Fund NAV at market close
  - Cash Transfer: Capital Inflow/Outflow to/from the Fund
- Holdings Data
  - Date: date
  - Asset ID: unique identifier for the security
  - Asset Type ID: security type. 1 stands for common stock
  - Quantity: number of shares at market close
  - Local Price: local close price for security
  - Local Currency Id: local currency for security. 1 stands for USD
  - Base Currency Id; base currency for fund. 1 stands for USD
  - Base Market Value: base market value of the position
- Trades Data
  - Date: date
  - Asset Id: unique identifier for the security
  - Trade Type Id: trade type. 1 stands for buy, 2 stands for sell, 3 stands for cash dividend, 5 stands for stock split
  - Trade Quantity: For buy/sell/stock split, it represents the change in number of shares due to the transaction. For cash dividend, it represents the number of shares that generates dividend
  - Local Price: the price at which transaction is executed
  - Base Cash Movement: change in cash holdings due to transactions (in base currency)
  - Transaction Id: unique identifier for the trade