**PORTFOLIO OPTIMIZATION USING PRINCIPAL COMPONENT ANALYSIS (PCA)**

**OVERVIEW**

The goal of this portfolio optimization is to maximize the Sharpe ratio by selecting optimal portfolio weights. The Sharpe ratio, a widely used metric in finance, measures the risk-adjusted return of an investment by comparing the excess return to its risk. This process involves analyzing historical returns and market signals from a selection of assets using Principal Component Analysis (PCA) for dimensionality reduction, then applying Ordinary Least Squares (OLS) regression to derive the optimal weights for portfolio allocation.

**OBJECTIVE**

* Maximize the Sharpe Ratio of the portfolio for in-sample data.
* Evaluate Sharpe Ratio performance on out-sample data to verify the model's robustness.
* Compare with a baseline to gauge improvements.

**DATA DESCRIPTION**

The optimization leverages historical asset data, divided into two segments:

* **In-Sample Data:** Used to train and fine-tune the model, providing the basis for weight calculations.
* **Out-Sample Data:** Serves as a test set to validate the portfolio's performance and assess if in-sample optimizations translate into real-world improvements.

**INPUT VARIABLES**

* **Returns Data:** Historical returns for each asset over a defined period.
* **Signals Data:** Market signals (e.g., macroeconomic indicators) potentially impacting asset returns.

**METHODOLOGY**

1. **Dimensionality Reduction With Principal Component Analysis (PCA)**

The in-sample signals are high-dimensional, possibly resulting in overfitting when used directly. By using PCA, we reduce dimensionality while retaining critical variance information.

* **PCA Transformation:** PCA identifies orthogonal components (principal components) that represent significant underlying patterns in the data, capturing the most variance with the least components.
* **Component Selection:** PCA components are chosen based on the number of assets, ensuring that the signals remain interpretable and manageable while providing a compressed, informative input for OLS regression.

**Mathematical Formulation of PCA**

Given a signals matrix X of shape m\*n (where m is the number of samples and n is the number of features), PCA seeks a transformation that reduces X to k dimensions (where k < n):

* **Mean Centering:** Subtract the mean from each feature in X.
* **Covariance Matrix:** Calculate the covariance matrix of X, denoted as £ = 1/m (XT X).
* **Eigen Decomposition:** Compute the eigen values and eigen vectors of £.
* **Transformation:** Select the top k eigen vectors, forming a matrix P that projects X into a lower-dimensional space, Xreduced = X . P.

1. **Weight Calculation with Ordinary Least Squares (OLS) Regression**

Using the reduced signals matrix, OLS regression is employed to estimate the optimal weights that align the portfolio’s expected returns with the historical asset returns.

**Mathematical Formulation of OLS**

The OLS regression model estimates a coefficient vector β that minimizes the sum of squared residuals between the predicted and actual returns:

* **Model:** y = Xreduced β + €, where y represents the returns and Xreduced the PCA transformed signals.
* **Solution:** The closed-form solution for β is derived as:

Β = (Xreduced T Xreduced)-1 Xreduced Ty.

1. **Weight Normalization**

To ensure practical portfolio allocation, the calculated weights are normalized so that they sum up to 1. This step ensures adherence to portfolio constraints and improves interpretability.

**Normalization Formula**

If w represents the initial weights vector derived from OLS, the normalized weights wnorm are computed as:

wnorm = w/(∑iwi)

1. **Sharpe Ratio Calculation**

The Sharpe ratio measures the average return in excess of a risk-free rate (assumed to be zero in this case) per unit of risk (standard deviation of returns). The Sharpe ratio, SR, is calculated as:

SR = E[R]/σR

Where:

* E[R]: Expected return (mean of portfolio returns).
* σR: Standard Deviation of portfolio returns.

The Sharpe ratio is annualized for a more standardized comparison by multiplying by √16 (assuming quarterly data), allowing us to gauge risk-adjusted returns effectively.

1. **Evaluation and Comparison**

The calculated Sharpe ratios for both in-sample and out-sample data allow for a comparison of portfolio performance against a baseline Sharpe ratio of 0.64. By examining the Sharpe ratio improvement, we determine the robustness of the PCA-OLS model.

**RESULTS**

**Sharpe Ratio Output**

* **In-Sample Sharpe Ratio:** Achieved a Sharpe ratio of **3.08**, indicating a substantial return-to-risk ratio based on the trained model.
* **Out-Sample Sharpe Ratio:** The model maintained a Sharpe ratio of **3.08** on out-sample data, suggesting that the portfolio allocation is stable across different data samples.

**Performance Comparison**

* The optimized Sharpe ratio of 3.08 significantly exceeds the baseline of 0.64, demonstrating that the PCA-OLS approach successfully enhances risk-adjusted returns.

**Visualization and Analysis**

Several visualizations are generated to support the analysis:

* **Return Distributions:** Histograms of in-sample and out-sample returns illustrate the data distribution, helping to verify assumptions of normality and check for outliers.
* **Cumulative Returns:** Line plots of cumulative returns highlight the growth trajectory of the portfolio over time, giving a comparative view of in-sample and out-sample performance.
* **Sharpe Ratio Scatter:** Scatter plots juxtapose in-sample and out-sample Sharpe ratios, with baseline comparisons, providing a visual confirmation of model performance across datasets.

**CONCLUSION**

This PCA-based portfolio optimization approach, followed by OLS regression, demonstrates a robust method for enhancing the Sharpe ratio in a 20-asset portfolio. PCA effectively reduced data dimensionality, improving computational efficiency and mitigating overfitting risks. The resulting weights maximized returns relative to risk, surpassing the baseline Sharpe ratio on both in-sample and out-sample datasets.