

Problem:

* There is a secret portfolio, a financial black-blox, which publicly reports its returns weekly.
* We do not know either which instruments the portfolio contains or what weight each instrument has;
* We do know the vast investments vehicles available in the markets: the **HFRX Index**, the **MSCI World Index**, the **MSCI World All Country Index**, the **Barclays Bloomberg Global Aggregate Bond Index** and a range of **Futures contracts**

Our goal is to guess what the inside of the black box looks like.

1. RX1 = Bund (10 Yrs Ger)
2. CO1 = Brent (Oil)
3. DU1 = Schatz (2yrs Gvt Ger)
4. ES1 = S&P 500 ( US Equity)
5. GC1 = Gold
6. LLL1 = MSCI Emerging Markets (EM Equity)
7. NQ1 = Nasdaq 100 (Tech Equity)
8. TP1 = Topix (Jap. Equity)
9. TU2 = 2Yrs US Treasury (US Gvt)
10. TY1 = 10Yrs US Treasury (US Gvt)
11. VG1 = Eurostoxx 50 (EU Equity)

* Data exploration [ What we can do and why]

*Statistical Analysis:* Mean, standard deviation, minimum, maximum and correlation coefficients for the weekly returns of each index. Provide insights into the individual characteristics and relationships, valuable in understanding the risk and return profiles of each index and identifying potential diversification opportunities.

*Time series Analysis:* Analyze trends, seasonality, and volatility clustering. Helps to understand the historical behavior of each index. In order to analyze trends: plot them, moving avarages to smooth put short-term fluctuations and highlighti long-term trends ( Simple moving average and exponential moving average, by comparing different moving avarages [e.g. 50-day,200 day], we can observe the direction and strength of the trend ), trend lines ( linear regression or exponential smoothing ), Volatility indicators like the average True Range (ATR) or Bollinger Brands, [**seasonal adjustments**](https://it.mathworks.com/help/econ/seasonal-adjustment-1.html)**,** conduct statistical trends such as Augmented Dickery-Fuller test to determine stationarity, relative strength index (RSI), or MACD and historical analysis.

*Visualization:* Charts, instograms, and scatter plots to compare the indexes and understand correlations.

*Factor Analysis:* Factor analysis techniques like PCA can help identify common factors or latent variables that explain the variations in the returns of different indexes. This information can guide you in selecting indexes that capture different factors and diversify the target portfolio across these factors, leading to a more robust and balanced allocation.

*Clustering:* Clustering analysis helps group indexes based on their return patterns. By identifying clusters of indexes that tend to move together, you can gain insights into the potential composition of the black box portfolio. If certain indexes consistently cluster together, it suggests that they may have similar underlying characteristics or exposures. This knowledge can guide you in selecting indexes that align with the clusters representing the desired investment themes or asset classes for the target portfolio.

*Performance comparison:* of the target and indexes.

* *Construction of the moster index*

Pay attention to the hypothesis of the models you’re going to use.

* Classical models for portfolio replica:

[Mean-Variance Optimization (MVO):](https://www.effisols.com/basics/) MVO is a popular model that aims to find the optimal portfolio allocation by balancing the expected returns and risk (variance). It considers the historical returns and covariance matrix of the assets to determine the portfolio weights that maximize the expected return for a given level of risk.

How to implement MVO: Define your target portfolio: Start by specifying the composition and weights of your target portfolio. Although you mentioned that you only know the weekly returns of the target portfolio, you'll need to provide additional information about the desired asset allocation and target weights for the assets you want to replicate.

1. Collect historical data: Gather historical data for the futures contracts and any other relevant data required for the MVO analysis. This includes historical prices or returns for the futures contracts, as well as any necessary data for estimating risk and return parameters (e.g., risk-free rate, market data, historical correlations).
2. Calculate the inputs for MVO: a. Calculate the expected returns: Use the known weekly returns of your target portfolio to estimate the expected returns for each futures contract. You can use statistical methods like historical mean returns or other forecasting techniques to derive the expected returns.

b. Estimate covariance matrix: Compute the covariance matrix of the futures contracts' returns. The covariance matrix represents the relationships and co-movements between the returns of different futures contracts. It is a crucial input for MVO and can be estimated using historical return data.

1. Formulate the MVO problem: Set up the MVO problem to find the optimal weights for the futures contracts that replicate the target portfolio. The objective is to maximize the expected return of the replicated portfolio while minimizing its risk (variance or standard deviation). The constraints can include budget constraints, target portfolio weights, and other limitations specific to your investment objectives.
2. Solve the MVO problem: Use optimization techniques to solve the MVO problem and find the optimal weights for the futures contracts. MATLAB provides optimization functions like **fmincon** or **quadprog** that can be used for this purpose. These functions allow you to define the objective function, constraints, and optimization options.
3. Validate and rebalance: Once you have the optimal weights, validate the results by comparing the replicated portfolio's characteristics (expected return, risk, and allocation) against the target portfolio. It's important to monitor and periodically rebalance the replicated portfolio to maintain the desired allocation as market conditions and asset prices change.

<https://www.kaggle.com/code/vijipai/lesson-5-mean-variance-optimization-of-portfolios>

<https://www.kaggle.com/code/louyueze/garch-mvo>

Black-Litterman Model: The Black-Litterman model is an extension of MVO that incorporates investor views and incorporates them into the optimization process. It allows you to incorporate your own beliefs or expectations about the future performance of assets to adjust the portfolio allocation.

Capital Asset Pricing Model (CAPM): CAPM is a model that relates the expected return of an asset to its systematic risk (beta). It assumes that the expected return is a linear function of the risk-free rate, the asset's beta, and the market risk premium. By estimating the beta of your monster index and the risk-free rate, you can use CAPM to determine the expected return of the portfolio.

How to implement CAMP: Define your target portfolio: Start by specifying the composition and weights of your target portfolio. Although you mentioned that you only know the weekly returns of the target portfolio, you'll need to provide additional information about the desired asset allocation and target weights for the assets you want to replicate.

1. Collect historical data: Gather historical data for the futures contracts and any other relevant data required for the CAPM analysis. This includes historical prices or returns for the futures contracts, as well as data on the risk-free rate and market returns.
2. Estimate the beta values: Beta represents the sensitivity of an asset's returns to market movements. Estimate the beta values for each futures contract in your set using historical return data. You can calculate the beta using regression analysis, where the dependent variable is the return of the futures contract and the independent variable is the return of a suitable market index or benchmark.
3. Estimate the market risk premium: The market risk premium is the excess return expected from the market above the risk-free rate. You can estimate the market risk premium using historical data by subtracting the risk-free rate from the average return of a suitable market index or benchmark.
4. Calculate the expected return: Use the CAPM formula to calculate the expected return for each futures contract. The CAPM formula is as follows: Expected Return = Risk-Free Rate + Beta \* Market Risk Premium

Plug in the estimated beta values and the estimated market risk premium to calculate the expected return for each futures contract.

1. Determine the replication weights: With the estimated expected returns for the futures contracts, use the CAPM model to determine the replication weights. The replication weights represent the optimal allocation of the futures contracts that replicates the target portfolio. The weights should be proportional to the expected returns and inversely proportional to the risk (beta) of each futures contract.
2. Validate and rebalance: Once you have the replication weights, validate the results by comparing the replicated portfolio's characteristics (expected return, risk, and allocation) against the target portfolio. It's important to monitor and periodically rebalance the replicated portfolio to maintain the desired allocation as market conditions and asset prices change.

Arbitrage Pricing Theory (APT): APT is an alternative to CAPM that considers multiple risk factors instead of just the market risk. It assumes that the expected return of an asset is determined by its exposure to various systematic risk factors. By identifying and estimating the relevant risk factors for your monster index, you can use APT to replicate the portfolio.

How to use it:

1. Define your target portfolio: Start by specifying the composition and weights of your target portfolio. Although you mentioned that you only know the weekly returns of the target portfolio, you'll need to provide additional information about the desired asset allocation and target weights for the assets you want to replicate.
2. Collect historical data: Gather historical data for the futures contracts and any other relevant data required for the APT analysis. This includes historical prices or returns for the futures contracts, as well as data on potential risk factors that may drive the returns of your target portfolio.
3. Identify potential risk factors: APT assumes that asset returns are driven by multiple systematic risk factors. Identify the potential risk factors that may influence the returns of your target portfolio. These factors can be macroeconomic variables, industry-specific variables, or any other factors that you believe are relevant to the assets you want to replicate.
4. Estimate factor loadings: Determine the factor loadings or sensitivities of each futures contract to the identified risk factors. Use statistical techniques such as regression analysis to estimate the factor loadings. The dependent variable will be the returns of the futures contract, and the independent variables will be the returns or values of the identified risk factors.
5. Determine the risk premium for each factor: Estimate the risk premium or expected return associated with each identified risk factor. This can be done by analyzing historical data or using market expectations. The risk premium represents the excess return that investors require for bearing the risk associated with each factor.
6. Calculate the expected return: Multiply the factor loadings of each futures contract by the corresponding risk premium for each factor. Sum up these expected returns to calculate the overall expected return for each futures contract.
7. Determine the replication weights: With the estimated expected returns for the futures contracts, determine the replication weights. The replication weights represent the optimal allocation of the futures contracts that replicates the target portfolio. The weights should be proportional to the expected returns of each futures contract.
8. Validate and rebalance: Once you have the replication weights, validate the results by comparing the replicated portfolio's characteristics (expected return, risk, and allocation) against the target portfolio. It's important to monitor and periodically rebalance the replicated portfolio to maintain the desired allocation as market conditions and asset prices change.

It's worth noting that APT relies on certain assumptions, such as the absence of arbitrage opportunities and the existence of a linear relationship between asset returns and risk factors. Additionally, accurately identifying and estimating the relevant risk factors is crucial for the success of the replication. Therefore, it's important to carefully consider the limitations and risks associated with APT and futures trading, and evaluate the accuracy and reliability of the estimated inputs before making investment decisions.

Factor Models: Factor models analyze the returns of assets based on common risk factors that drive their returns. By identifying relevant factors that influence the returns of your monster index, you can construct a factor model and estimate the factor loadings to replicate the portfolio.

1. Single-Factor Model:
   * Example Factor: Market Index (e.g., S&P 500)
   * Regression Equation: Asset Return = α + β \* Factor Return + Error Term
   * Implementation: Use MATLAB's regression functions (e.g., **regress**, **fitlm**) to estimate the factor loading (β) and intercept (α) for each asset based on historical returns.
2. Multi-Factor Model:
   * Example Factors: Market Index, Interest Rates, Oil Prices
   * Regression Equation: Asset Return = α + β₁ \* Factor₁ Return + β₂ \* Factor₂ Return + β₃ \* Factor₃ Return + Error Term
   * Implementation: Similar to the single-factor model, use regression functions to estimate the factor loadings (β) and intercept (α) for each asset based on historical returns of multiple factors.
3. Fama-French Three-Factor Model:
   * Factors: Market Return, Size (Small Minus Big), Value (High Minus Low)
   * Regression Equation: Asset Return = α + β₁ \* Market Return + β₂ \* Size Factor + β₃ \* Value Factor + Error Term
   * Implementation: Estimate the factor loadings (β) and intercept (α) using regression functions. You may need to construct the Size and Value factors based on relevant data.
4. Carhart Four-Factor Model:
   * Factors: Market Return, Size, Value, Momentum
   * Regression Equation: Asset Return = α + β₁ \* Market Return + β₂ \* Size Factor + β₃ \* Value Factor + β₄ \* Momentum Factor + Error Term
   * Implementation: Estimate the factor loadings (β) and intercept (α) using regression functions. You will need to construct or obtain the necessary data for the Size, Value, and Momentum factors.

Machine Learning Models: Machine learning techniques, such as regression models, support vector machines (SVM), random forests, or neural networks, can also be used to replicate a portfolio. These models learn patterns and relationships from historical data to make predictions about future returns. By training the model with the historical returns of your monster index, you can use it to generate a replication of the portfolio.

Lasso, Ridge models

* Hybrid models

Yes, hybrid models do exist, and they are commonly used in various fields, including finance and portfolio management. Hybrid models combine elements from different modeling approaches or techniques to leverage the strengths of each component and create a more robust and accurate model.

In the context of portfolio replication, hybrid models can be constructed by combining multiple modeling techniques or methodologies. For example, you can create a hybrid model by integrating elements of mean-variance optimization (MVO), factor models, and machine learning algorithms. This allows you to take advantage of the benefits of each approach and overcome their individual limitations.

Hybrid models can be designed in different ways depending on the specific objectives and requirements. Here are a few examples:

MVO with Factor Model: You can incorporate factor models, such as the Fama-French Three-Factor Model or the Carhart Four-Factor Model, into the MVO framework. This helps capture the systematic risk factors that drive asset returns and refine the portfolio allocation based on these factors.

Machine Learning with Risk Measures: You can combine machine learning algorithms, such as random forests or neural networks, with risk measures like Value-at-Risk (VaR) or Conditional Value-at-Risk (CVaR). This allows you to incorporate both predictive power and risk management considerations in portfolio replication.

Bayesian Optimization with Constraints: You can use Bayesian optimization techniques to find the optimal portfolio allocation while incorporating specific constraints, such as sector or asset class exposure limits. This hybrid approach balances the optimization process with practical constraints.

These are just a few examples of how hybrid models can be constructed by integrating different methodologies. The specific design of a hybrid model depends on the objectives, available data, and the characteristics of the portfolio being replicated.

It's important to note that constructing and implementing hybrid models require careful consideration of the underlying assumptions, data requirements, and potential trade-offs. It's recommended to thoroughly test and validate the hybrid model's performance before applying it to real-world investment decisions.

Notes:

* + We can use papers to justify out choices: see [using futures for replication](https://www.cmegroup.com/trading/equity-index/report-a-cost-comparison-of-futures-and-etfs.html).
  + Replication of Buffet portfolio: <https://www.kaggle.com/code/shivavashishtha/replicating-warren-buffett-portfolio-with-python>