

DATA 608 - 2022 SPRING TERM

Bruno de Melo

Final Project

Title: Mutual Funds portfolios comparison

Source: <https://www.kaggle.com/datasets/stefanoleone992/mutual-funds-and-etfs>

Content

The file contains 23,783 Mutual Funds with:

- General fund aspects (e.g. totalnetassets, fund family, inception date, etc.)
- Portfolio indicators (e.g. cash, stocks, bonds, sectors, etc.)
- Historical yearly and quarterly returns (e.g. yeartodate, 1-year, 3-years, etc.)
- Financial ratios (price/earning, Treynor and Sharpe ratios, alpha, and beta)

Project goal is to create a shiny app that can be used as a comparison tool of risk and return data of Mutual Funds portfolios - based on an investor's risk tolerance. Tool to display charts and tables that are visually easy to understand, not requiring great knowledge about complex concepts like value-at-risk, drawdowns, correlation, etc.

1: Data preparation

Dataset was scraped from yahoo.finance and contains 24k mutual funds with around 300 features. Funds have with multiple share classes and there was not a single flag that could distinguish a unique representative. Usually in datasets from commercial providers like Morninstar, flags such as "oldest share class", or "institutional share class" are used to determine uniqueness.

Here I used the '*distinct*' function on several variables to keep off duplicates: Fund_Long_Name, Fund_Category, Investment_Strategy, Asset_Allocation, Top_10_holdings. The end results was a reduction from 24k to 15k unique funds.

Moreover, throughout the project data quality issues arised such as misspecification of several variables like fund strategy, asset allocation breakdown, fund category. These were not corrected.

In addition, all periodic return was displayed in columns with no time stamp/date attached. I had to do string manipulation to have a date assigned to each return and use the '*pivot longer*' function to create a proper return stream that could be used in the tables and charts.

2: Data visualization

Visualization tool is created in ShinyApp and contains several charts and tables displaying various risk and return characteristics of portfolio(s) like return and risk over time, drawdowns, return distribution, portfolio breakdowns by ratings, sector, strategy, asset class. These portfolio are based on different risk tolerance levels: conservative, moderate, aggressive.

Risk tolerance levels are defined by the 3-year volatility: funds in the 75th percentile and above are deemed aggressive, funds in the 25th percentile are deemed conservative and the remainder is deemed conservative. This is a rather naive approach to portfolio construction but considering that portfolio construction techniques was not the main subject of the project, this method was chosen for simplicity sake.

Intention is to present a comparison tool that is easy to understand and to communicate in a straight forward way complex concepts like value-at-risk, drawdowns, correlation, etc.

This tool is important because it is able to show through simple visuals how risk tolerances affect risk and return over time, impact of drawdowns, portfolio composition, and trade-offs that are made when a different risk level is chosen.