

Developing an Investment Portfolio Utilizing Machine Learning

Warren Denetsosie & Munqiz Minhas

Summary

Step 1: Determining portfolio allocation using Rule of 110, 5% in crypto and remaining percentage in bonds

Step 2: Used machine learning to model top 25 stocks in the S&P for performance in the short-term (Prophet, regression, decision trees regressor). Utilized other metrics: Sharpe ratios, Beta

Step 3: Monte Carlo projections for stock portfolio performance over the long-term

Data Techniques

- Data sources: Yahoo Finance API, U.S. Bureau of Labor Statistics, and CoinCodex. We gathered 5-10 years of data, depending on the analysis performed.
- Reasoning for data selection: Needed contemporaneous independent features (CPI and unemployment data) for machine learning models
- Collection, exploration and cleaning process

Step 1: Portfolio Allocation

- Rule of 100 vs. Rule of 110: The stock portion of a retirement portfolio used to be determined by subtracting the investor's age from 100. Now that people are living longer and working later in life, the rule of thumb is to subtract age from 110. Example: $110 - 35 \text{ year old} = 75\%$ in stocks.
- Crypto portfolio: The general advice is to include 1-5% of crypto assets in a retirement portfolio.
- The remaining allocation is in bonds and REITs.

Portfolio Allocation (cont'd)

For purposes of stock selection, we picked 3 portfolios from the top 25 S&P500 companies (determined by market capitalization).

Aggressive portfolio: Tesla, Nvidia, Apple, Amazon, and Meta

Moderate portfolio: Google, Microsoft, JP Morgan, Apple, and Amazon

Conservative portfolio: Coca-Cola, Pepsi, Proctor & Gamble, Apple, and Amazon

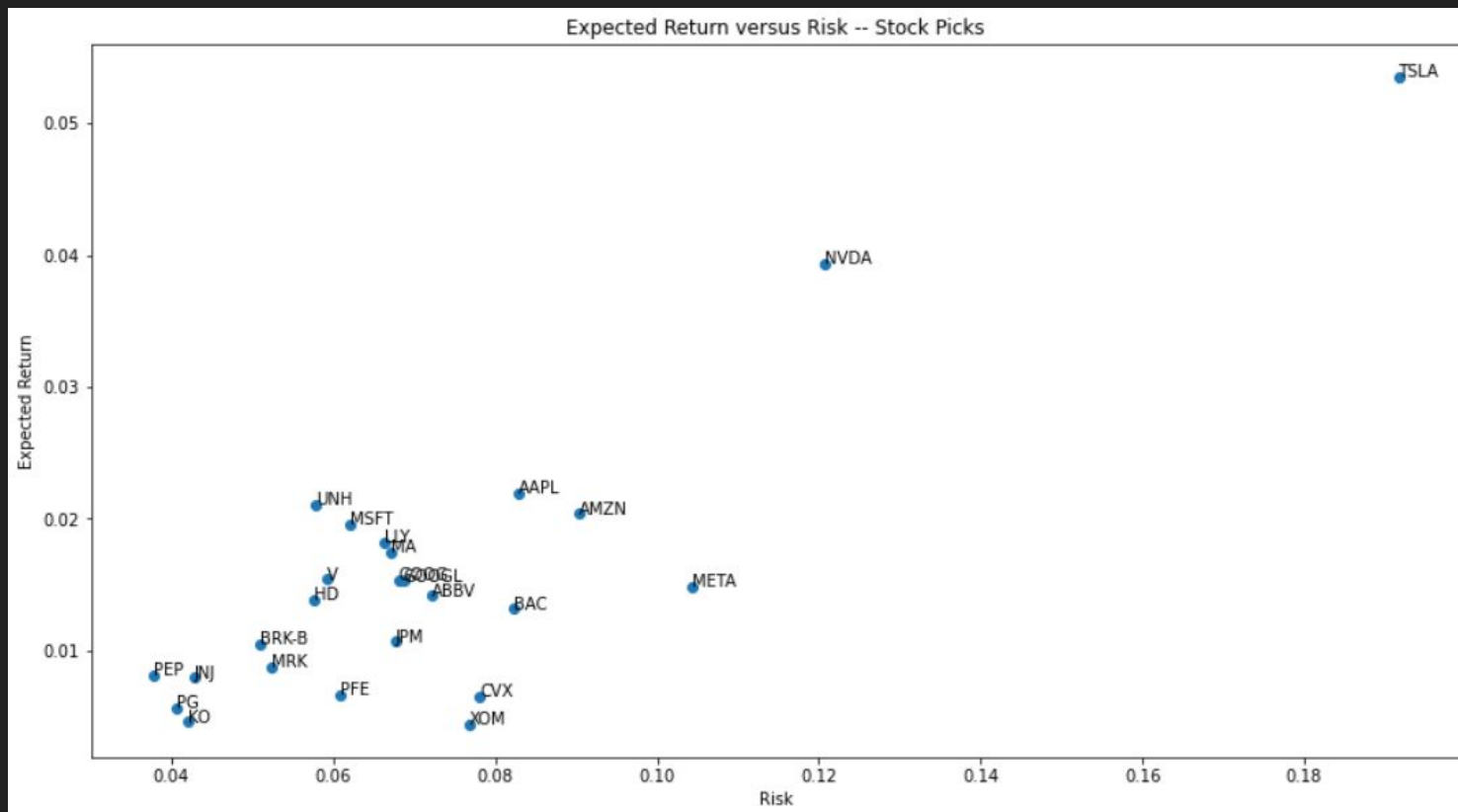
These were selected based on Machine Learning and traditional finance metrics.

Step 2: Analysis

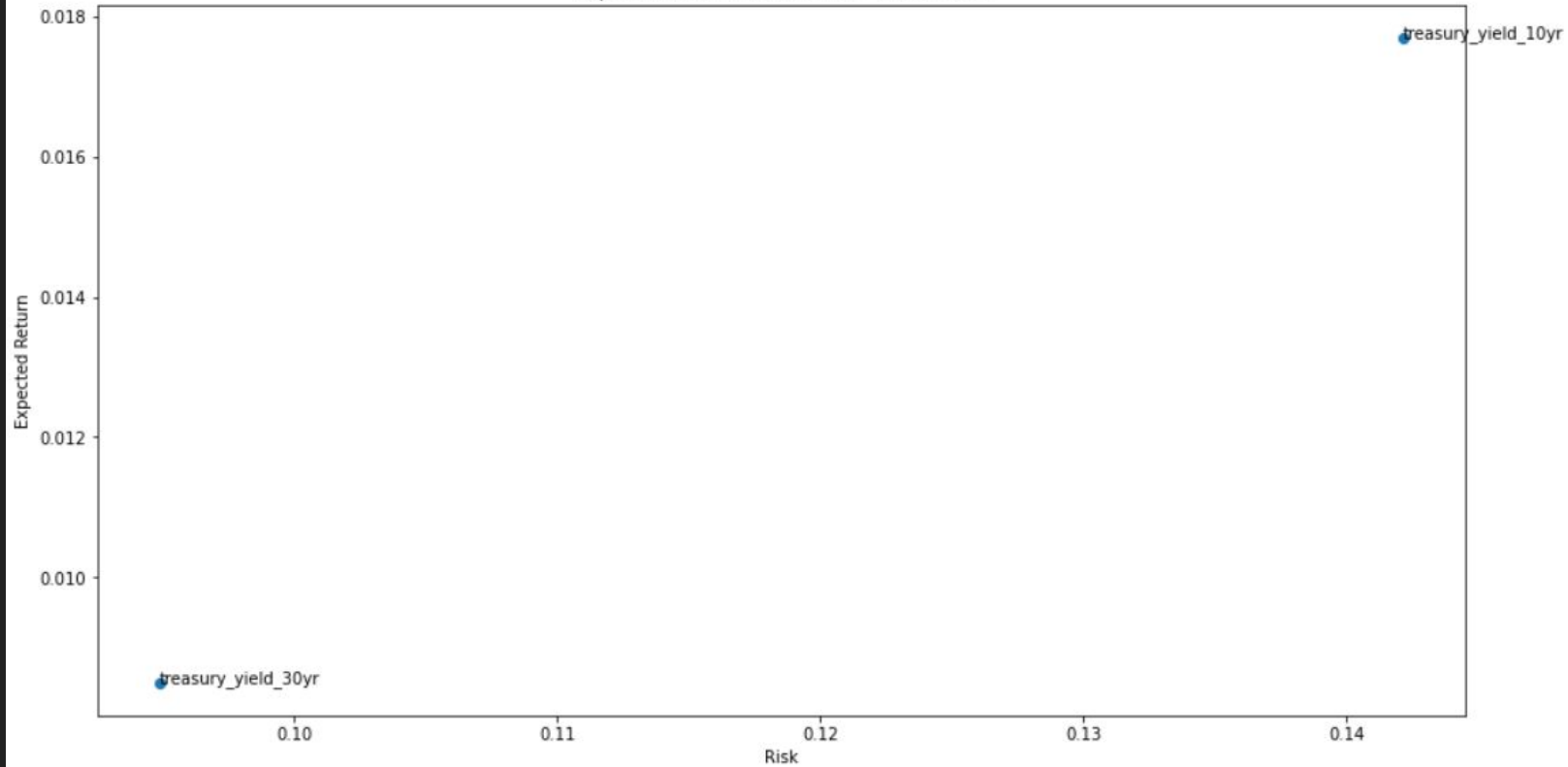
Utilizing the following metrics, we selected the stocks for portfolio selection and long-term portfolio analysis:

- Traditional metrics:
 - Expected return vs. risk
 - Sharpe ratios: Each of the stocks chosen had ratios above 0.6 (and up to 1.2 for NVIDIA and Tesla).
 - Beta: The beta of each of the stocks was near 1.0.

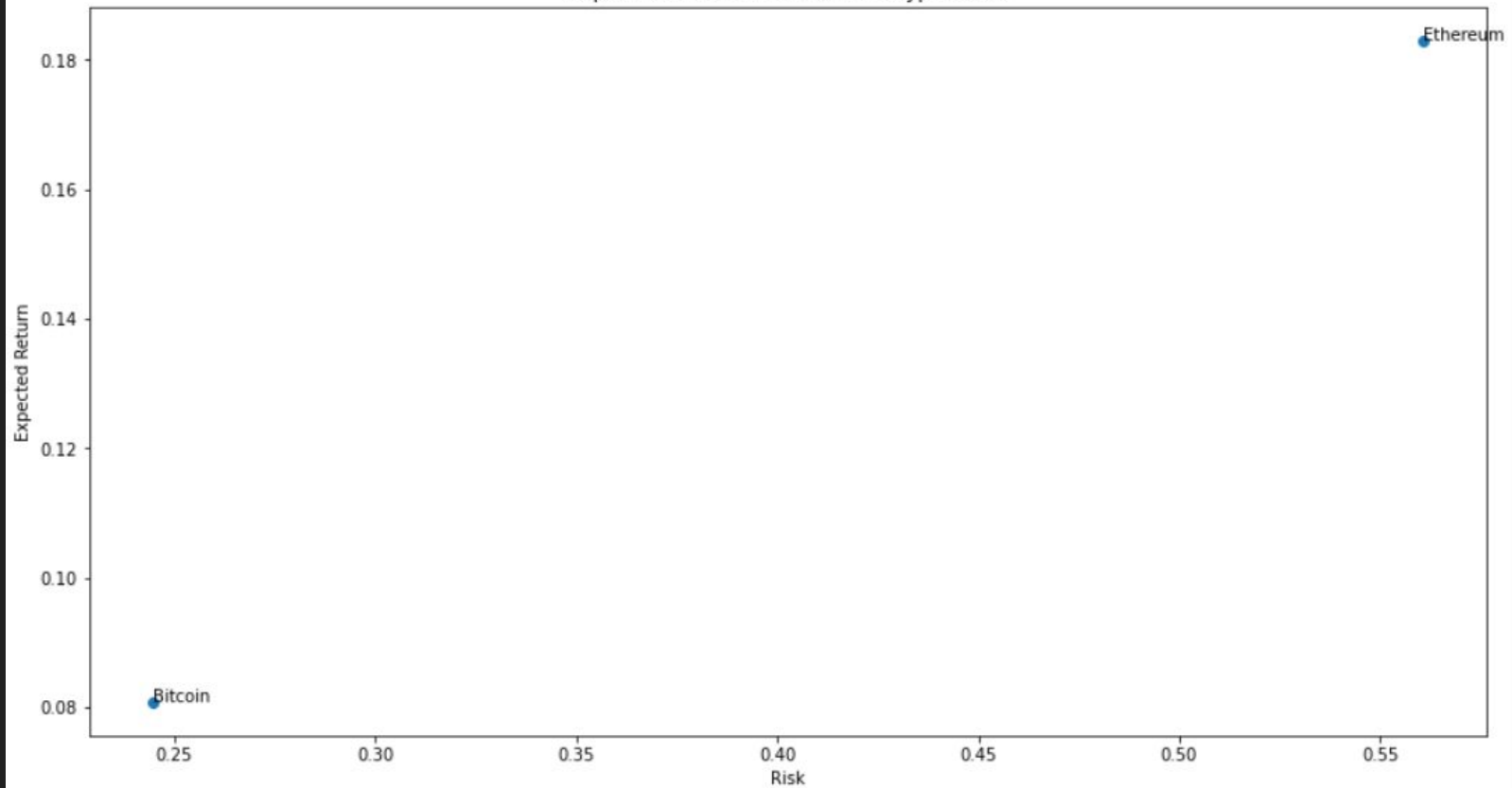
How we selected assets for the portfolio:



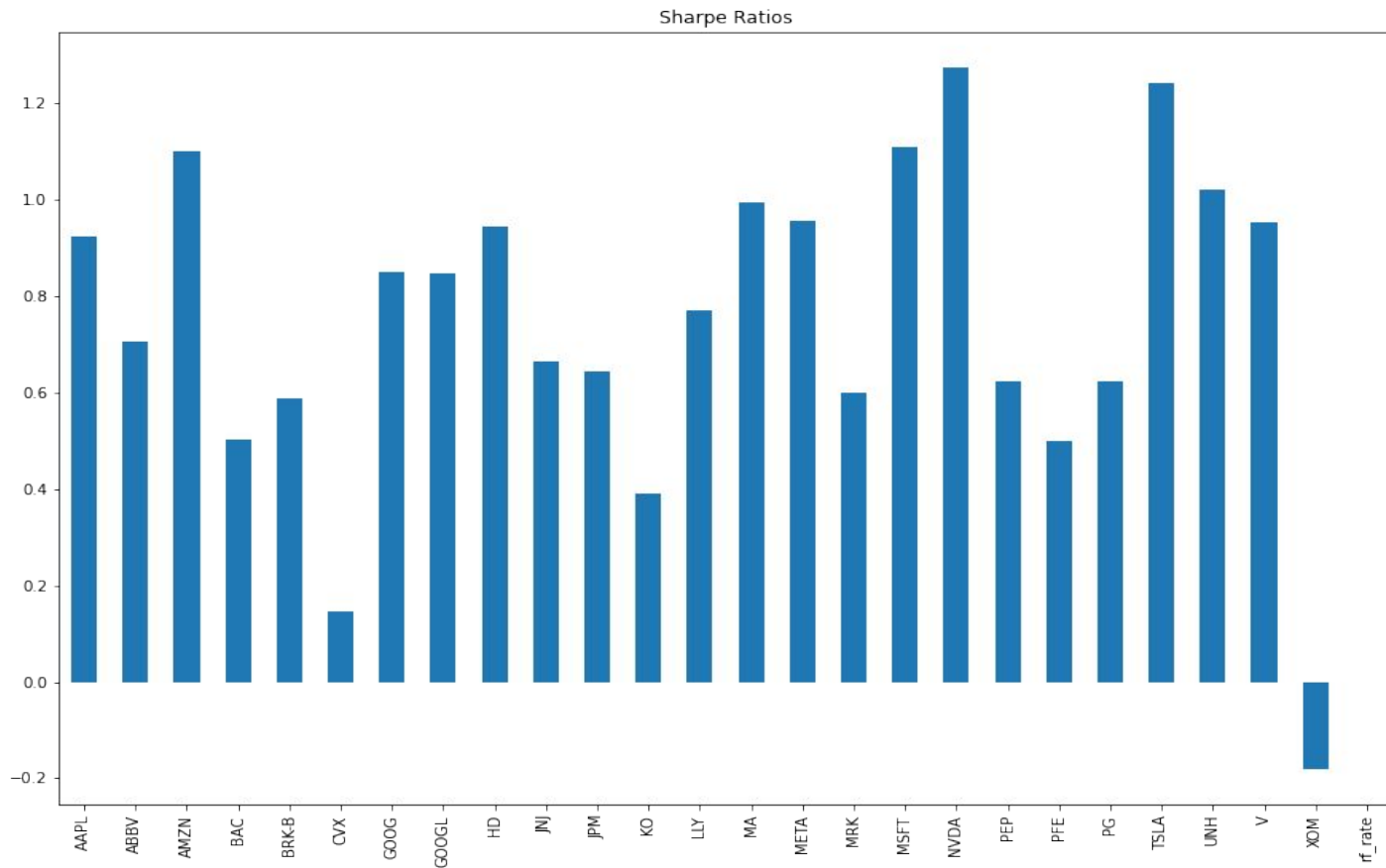
Expected Return versus Risk -- Bond Picks



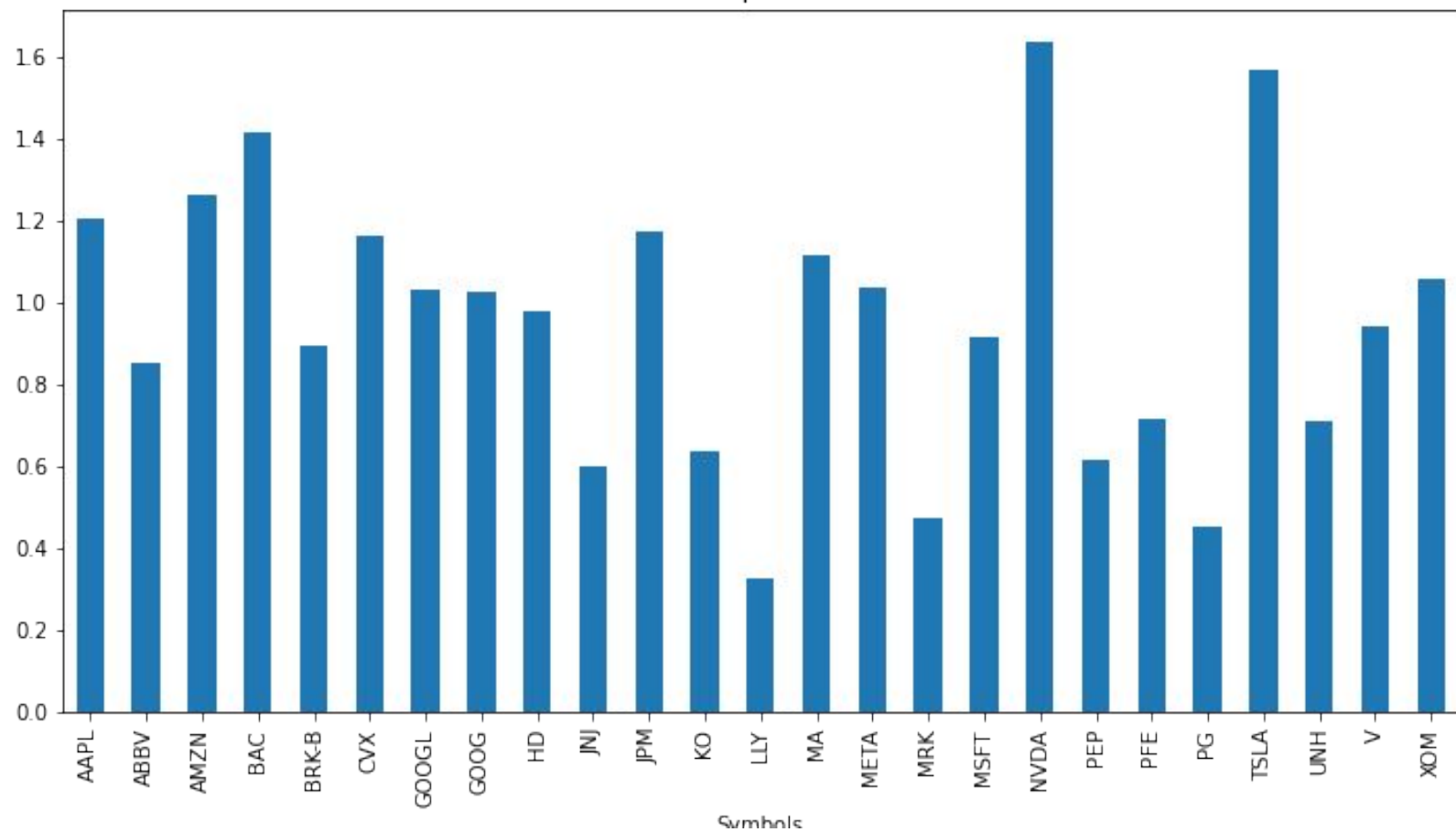
Expected Return versus Risk -- Crypto Picks



Sharpe Ratios



Beta of Top 25 S&P500



Step 2: Analysis (machine learning)

Stock selection was also done utilizing the following machine learning models:

- Logistic Regression
- Decision Trees Regressor
- Prophet

We tested the first two models to select next day stocks based on both historical stock data *only* and based on additional CPI and unemployment data. The results were nearly the same.

Logistic Regression results

Model 1: Lagged closing price data only

Accuracy: 0.4810810810810811

Model 2: Lagged closing prices, CPI, and unemployment data

Accuracy: 0.4967532467532468

Result: Independent economic data did not impact the model in any meaningful way.

Decision trees model

Single model utilizing CPI and unemployment data

Accuracy: 0.4967532467532468

Conclusion: Decision trees and logistic regression may be helpful when trading in the short-term. Prophet yields more meaningful results, as is shown below.

Prophet

Utilizing the same data, a prophet forecast yielded more meaningful results.

For any machine learning model, long-term forecasting does not have much value. A Monte Carlo forecast was more helpful calculating expected returns over a 5-year period of the above-defined portfolios with a 95% confidence interval.

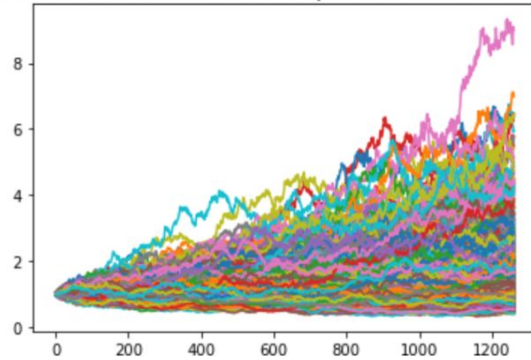
The accuracy rates for the various prophet forecasts are in the next chart.

Evaluating Prophet Forecasts using RMSE from Days 1 to 30

Stock Picks	Day 1 of Forecasts	Day 30 of Forecasts
TSLA	28.8790	36.27088
NVDA	29.76094	26.594314
META	21.5885	24.61883
AAPL	7.73256	10.62831
AMZN	11.84368	15.698068
GOOGL	7.090990	7.7429
MSFT	11.13465	12.2868
JPM	14.31650	15.08164
KO	3.32638	3.60321
PG	6.61769	7.7673
PEP	7.30246	8.10824

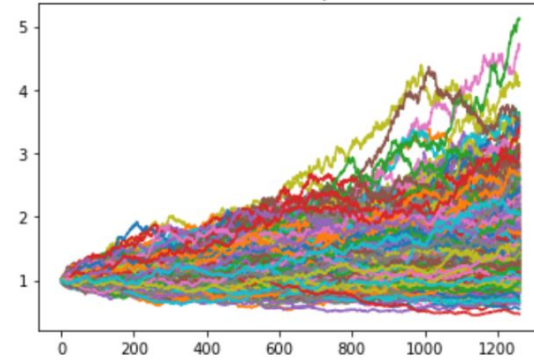
Aggressive Portfolio

500 Simulations of Cumulative Portfolio Return Trajectories Over the Next 1260 Trading Days.



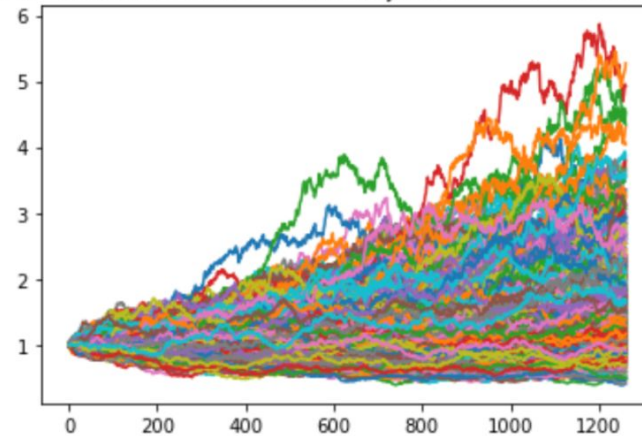
Conservative Portfolio

500 Simulations of Cumulative Portfolio Return Trajectories Over the Next 1260 Trading Days.



Moderate Portfolio

500 Simulations of Cumulative Portfolio Return Trajectories Over the Next 1260 Trading Days.



Q&A

Sources

- <https://coincodex.com/>
- <https://rapidapi.com/sparior/api/yahoo-finance15>
- <https://www.bls.gov/data/>
- <https://www.liberatedstocktrader.com/sp-500-companies/>