Portfolio Performance Analysis

Harshil Prajapati, Chirag Jadav
University of Maryland Baltimore County
1000 Hilltop Circle, Baltimore, MD 21250

(Dated: May 10, 2023)

Abstract

A portfolio analysis helps users to determine the next step in their investment in order to make profit in the future or avoid the loss. In this paper, we demonstrate how a portfolio's performance is analyzed to determine the risk and forecast its return to see future profitability. Our tool incorporates user input for selecting stocks, choosing a market index as a benchmark, specifying start and end dates, allocating weights to the asset and determining the amount to be invested. The dataset is fetched using YahooFinance and FRED API (Federal Reserve Bank web services). By leveraging statistical techniques and models, we provide comprehensive analysis of the portfolio's performance.Our tool offers multiple features allowing users to have detailed analysis of their portfolio. First, we show an asset comparison which shows basic details about the asset the user has selected. As a second part, portfolio's performance is evaluated to show risk and return tradeoff. Third feature forecasts the portfolio's return on investment for the next 30 days

Keywords: time-series analysis, portfolio performance analysis, asset selection, benchmark comparison, statistical techniques, investment management, evaluating risk.

.

I. INTRODUCTION

Portfolio performance analysis is a crucial aspect of investment management. It is to evaluate a portfolio's return against the benchmark's return [1]. An investor likes to see what profit or loss the portfolio has experienced on the investment, which asset is more risky compared to its return, etc. This allows investors to take decisions such as diversification, assigning more weights to the best performing asset, and many other approaches to reduce investment risk.

This project aims to provide a comprehensive portfolio performance analysis program that enables users to evaluate their investment performance and forecast its portfolio's return. By leveraging statistical models and techniques, the program offers insights into key performance metrics, risk-return trade-off, and forecasting capabilities. The feature of allowing users to provide the investment amount and weights for each asset adds a better customizable factor in the project. The project has multiple features as follow:

1. Asset Comparison
2. Portfolio Performance Analysis
3. Portfolio Return Forecast

The paper briefly explains the aforementioned features. It includes methodology, implementation and results of each asset. Statistical methods and modeling techniques are incorporated in this tool to achieve the results for each feature. Python programming language is used to carry out the analysis and forecasting.Overall the project aims to offer a valuable tool for analyzing the portfolio's performance based on the user's selection of asset, date range, amount to be invested and weight allocation using evidence-based approaches and established theories from the academic literature. This tool can be beneficial to those involved in or wishes to be involved in passive portfolio management [2].

This paper is organized as follows: section (II) we summarizes some of the similar work done in the industry and approaches of the portfolio management. Section (III) we demonstrate how the dataset is gathered before carrying out the analysis. The (IV) section involves detailed explanation on the methodologies incorporated in each feature of the tool. Section (V) Summarizes the results of the model. And the final section (VI) highlights the conclusion and future directions.

II. LITERATURE REVIEW

The field of portfolio analysis has garnered significant attention from researchers and practitioners in the realm of investment management. To compare portfolio returns against benchmark returns and assess the impact of investment strategies, researchers have focused on creating robust methodologies . By introducing the risk-adjusted return and Capital Asset Pricing Model (CAPM), Sharpe (1966) and Jensen(1968) established a cornerstone in portfolio performance evaluation [3]. These studies laid the foundation for subsequent research on risk-adjusted measures, such as Sharpe ratio, Treynor ratio, and Information Ratio, which are widely used in assessing portfolio performance. In our tool, we have used Sharpe ratio to assess the portfolio performance.

Statistics play a vital role in portfolio analysis allowing investors to dig deep in understanding the underlying patterns in the historical data. Time series methodologies, including AutoRegressive Integrated Moving Average models, have been extensively used to capture the underlying dynamics of asset prices and forecast future returns [4]. There are other models which have given promising results in forecasting time series data especially economics and financial data, such as LSTM [5]. Many studies have also used machine learning as well as deep learning techniques to predict the time series data. Makridakis et al. [6] compares various statistical, machine learning and deep learning forecasting methods to show how the time series forecasting has evolved over time.

III. DATASET

The dataset used in the analysis and forecast is fetched from the Yahoo Finance database and FRED database [7] which is publicly available. The historical data for the asset was downloaded from Yahoo Finance database using *yfinance* library in Python. The tool prompts users to select the asset and market index of their choice. It takes comma-separated ticker symbols for each asset to fetch the data from Yahoo Finance. Start and End dates for the analysis are taken as an input from the investor. Only daily Adjusted Closing price is considered for the analysis of the asset as well as market index selected by the users. (see exhibit 1)

	AAPL	JPM	META	NVDA	TSLA
Date					
2021-01-04	127.503639	117.236755	268.940002	130.875916	243.256668
2021-01-05	129.080063	117.874672	270.970001	133.782700	245.036667
2021-01-06	124.735039	123.409576	263.309998	125.895798	251.993332
2021-01-07	128.991379	127.462250	268.739990	133.176361	272.013336
2021-01-08	130.104752	127.602982	267.570007	132.505219	293.339996

EXHIBIT I : The above table shows pandas dataframe created after downloading the Adjusted Closing price for the assets selected by the users for a desired date range.

The FRED database was used to download the Treasury Bill Rate (also referred to as risk-free rate which is often used in this paper) to calculate the risk-adjusted performance of the portfolio. The date range of the T-rate is the same as the assets'. Generally the risk-free rate equal to the interest paid on the T-bill for 3-months is considered as a safe risk-free rate for the investment [8].

IV. METHODOLOGY

The aim of the analysis carried out in this section is to evaluate the contributing factors that decides portfolio performance and employ statistical learning models to predict the portfolio performance for the next 30-day horizon of the end date. The tool consists of three parts and methodologies employed in each section are explained in detail in this section.

A. Asset Comparison

The first includes an asset comparison. Once the asset ticker is provided by the user along with the date range, we download its historical data to show an Adjusted Closing price chart for each asset. Other comparisons in this section include each asset's return between the start and end date,correlation between each asset, and 30 days rolling volatility and drawdown.

Volatility in this case can be defined as the statistical measure to see how returns are dispersed for a given security or a market index [9]. We calculated rolling volatility with the below formula

$$\text{Rolling Volatility } (\sigma) = \sqrt{(1/M * \Sigma((Xi - \bar{X})^2))}$$

Where, σ represents the rolling volatility of the asset, M is the rolling window size (30 in our case), Xi represents the individual asset within the rolling window and $\bar{X}$ represents the mean return of the asset's value within the given window.

Adam et al. [10] explains a Maximum Drawdown (MDD) as the maximum observed loss from a peak to trough of a portfolio, before a new peak is attained. In simple equation is can expressed as:

$$MDD = Peak\ Value - Trough\ Value / Peak\ Value$$

In our analysis, we calculated a 30 days rolling drawdown for each asset. First we calculated the *running maximum* (RM) which is the highest value observed in historical daily stock price over a rolling window of 30 days. It is calculated using:

$$RM(t) = \max(P(t), P(t-1), ..., P(t-29))$$

Where P(t) represents an asset price at time t. *Daily drawdown* (DD) is evaluated with the help of running maximum as expressed below:

$$DD(t) = (P(t) / RM(t)) - 1$$

Once we have these values, we computed the *maximum daily drawdown* (MDD) as percentage declined in rolling 30 days. It is computed as shown below:

$$MDD(t) = \min(DD(t), DD(t-1), ..., DD(t-29)) * 100$$

Each asset's return for a selected date range is calculated to showcase how it performed for a selected date range. Here is a snippet provided as an example when Tesla, Apple, Meta, J.P. Morgan Chase & Co. and Nvidia stocks were selected for illustration purposes between between 2021-01-01 to 2022-01-01:

```
Asset Returns in Portfolio from 2021-01-
01 to 2022-01-01 is:
AAPL: 38.06%
JPM: 28.96%
META: 25.07%
NVDA: 124.48%
```

```
TSLA: 44.81%
```

B. Portfolio Performance Analysis

Second part provides analysis on the portfolio's performance. For this analysis, the user is prompted to allocate weights to each asset. Users have a flexibility to choose from two options: either assign weights manually or distribute them equally across the assets in the portfolio. If R_i is a return (percentage changes) of asset i, w_i is weight assigned to asset i for n number of assets in the portfolio, then portfolio return (R_p) is calculated as for amount M selected to invest:

$$R\_p = (w\_1 * R\_1 + w\_2 * R\_2 + \dots + w\_n * R\_n) * M$$

To ensure the weights add up to 1, the weights are normalized by dividing each weight by the total weight. Once the values are prompted by user, the portfolio will show analysis as:

```
Asset Allocation
aapl: 40.00% (4000.00 USD)
jpm: 10.00% (1000.00 USD)
meta: 20.00% (2000.00 USD)
nvda: 17.00% (1700.00 USD)
tsla: 13.00% (1300.00 USD)

Your investment of $10000.00 on 2021-01-01 would be worth $14586.91 by 2022-01-01 in your Portfolio.

You would have made a profit of $4586.91 (45.87%)
```

This feature compares the portfolio return with the selected benchmark return (See EXHIBIT III) It also compares Sharpe ratio for each asset to see the risk factor associated with each asset. Sharpe ratio can be defined as the measure of an investment's risk-adjusted performance [11]. The formula for calculating Sharpe ratio is:

$$\text{Sharpe Ratio} = (\mu - rf) / \sigma$$

Where μ (mu) represents the average excess return if the asset over a rolling window, rf (risk-free-rate) represents the risk-free rate of return which is subtracted from the asset's excess return and σ represents the standard deviation of asset's excess return. Excess return (ER) is a difference taken between the asset's return(Ri) and risk-free (Rf) rate (ER = Ri - Rf). This allows users to gauge potential risk associated with each asset in the portfolio and get an insight on the asset's risk-return tradeoff (see EXHIBIT II)

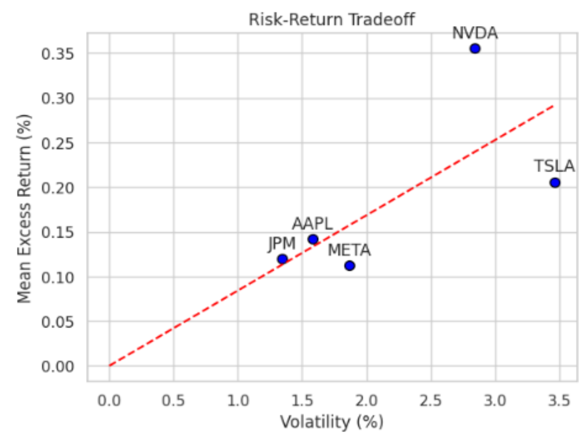

EXHIBIT II: The graph shows risk-return tradeoff between the asset's excess return and volatility of the portfolio.

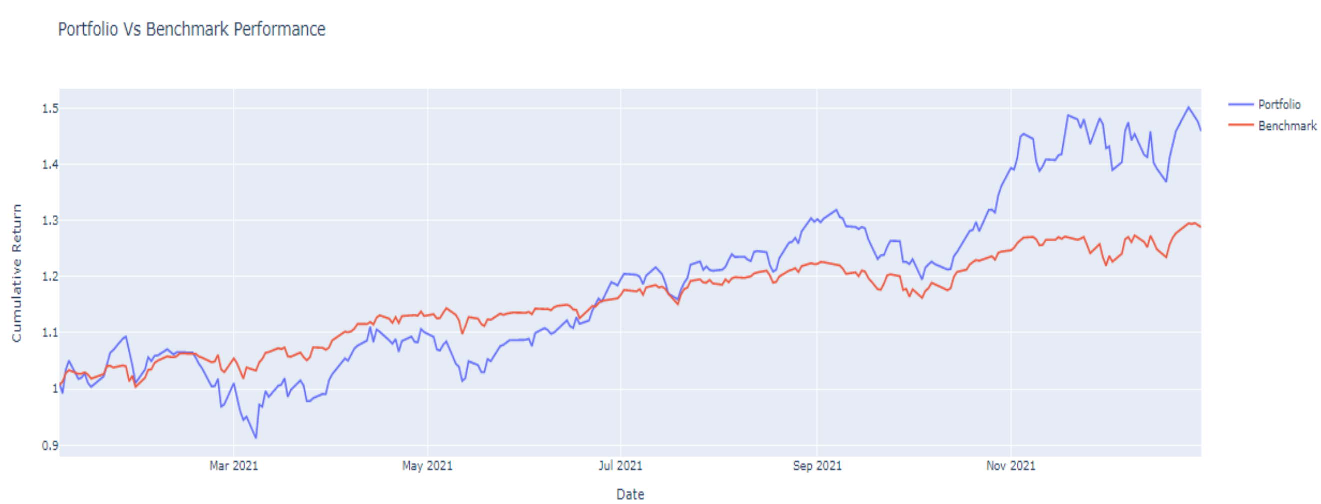

EXHIBIT III: Comparison of Portfolio Cumulative Return with the Benchmark for a selected date range.

C. Portfolio Return Forecast

Third feature of the portfolio analysis tool forecasts the portfolio return for the next 30 days of the end date. The Autoregressive Integrated Moving Average (ARIMA) model is employed to make the forecast. The dataset was preprocessed by taking the first difference to remove trend and seasonality. The aim is to find the optimal lag for the ARIMA model. To find the optimal lag, Akaike Information Criterion (AIC) values are computed to get the lag length having minimum AIC value. The results for lag values are stored in a list once determined from Partial Autocorrelation Function (PACF) method. Given a time series, denoted as PACF(k), is calculated using Yule-Walker equation:

$$PACF(k) = [\rho(k) - \Sigma\,(PACF(i) * \rho(k-i))],\ \text{for } i = 1 \text{ to } k-1$$

Where PACF(i) is partial autocorrelation at lag i and is $\rho(k)$ an autocorrelation evaluated as:

$$\rho(k) = cov(X(t), X(t-k)) / \sigma(X(t))^2$$

Where cov is covariance between X(t) and X(t-k). The optimal lag determined using AIC and PACF is used as p and q value in ARIMA model. If optimal lag is not found, then by default 8 is taken as p and q value in the ARIMA model. For training and testing the model, un-differenced data is taken as ARIMA model has d parameter, also known as degree of difference which does the same job. The dataset is split at 80-20% for training and testing purposes.

ARIMA model is a statistical time series forecasting model. It is an extension of the (Autoregressive Moving Average) ARMA model which is a merger of two components. (1) AR (Auto Regressive) which requires past values to predict the future values. (2) MA (Moving Average) which attempts to predict future values based on past forecasting errors. ARIMA model has additional I (Integration) component which relates the differencing information to the model [12]. Its order is (p,d,q) where p is the AR component, d is the I component and q is the MA component. AR model is formulated as:

$$Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t$$

Where Y_t is the value at time step t, ε_t is a constant and φ_i is the corresponding coefficient for each respective prior time step Y_{t-i}. The coefficients are to be estimated [12].

MA model can be defined as:

$$Y_t = \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

Where Y_t is the value at time step t, ε_t is a constant and θ_j is the corresponding coefficient for each respective prior error ε_{t-j}. And the I component is mathematically expressed as:

$$\Delta Y_t = Y_t - Y_{t-1}$$

Adding all these components makes ARIMA (p,d,q) where p is the order of the AR model component, d is the number of differences conducted on time series and q is the order of the MA component [12].

Once the values are forecasted, the analysis shows the total profit or loss that is incurred in the forecasted period. The graph is plotted for the user to track the actual and forecasted return of the portfolio (see EXHIBIT V).

```
You would incur a loss of $15.08 (0.10%)
over the forecasted period.
```

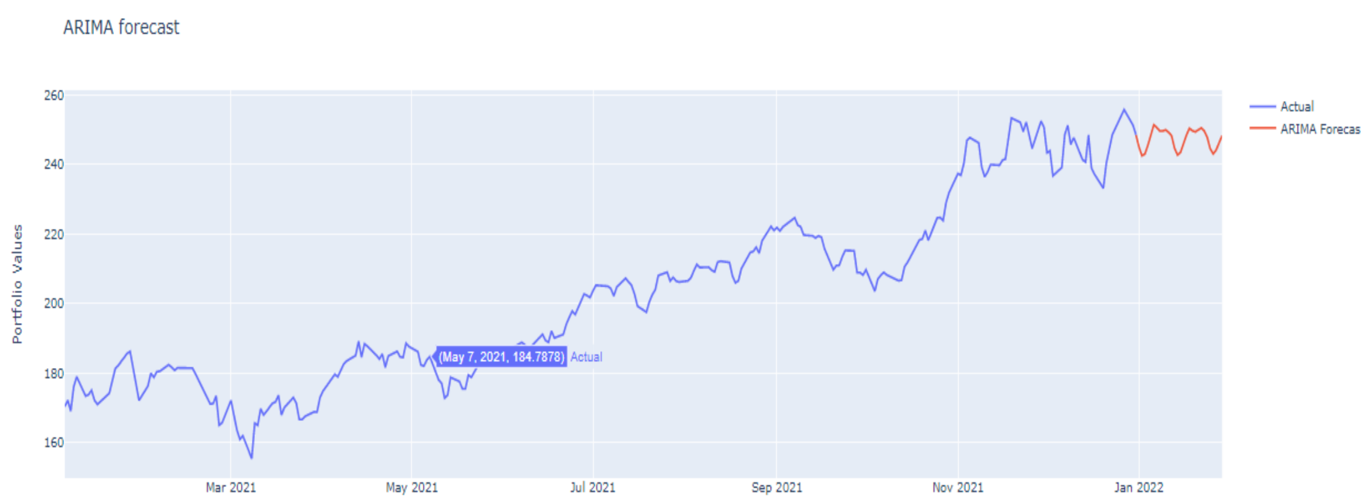

EXHIBIT V: Forecasting Portfolio Return for next 30 days of the selected end date. The red line represents the forecasted value whereas the blue line represents the actual value.

V. MODEL EVALUATION

The performance of the model was measured using error statistical models. These statistical models determines the accuracy by checking how much the predicted value deviated from the original value. The four model performance metrics used for the model evaluations are as follows:

Assuming $\hat{y}_i$ as the predicted values and y_i as the actual values for n number of datapoints, Mean Absolute Error (MAE) measures the magnitude of the errors in a set of predictions, without considering the direction.

$$MAE = (1/n) * \Sigma(|y_i - \hat{y}_i|)$$

Mean Absolute Percentage Error (MAPE) is the average percentage difference between actual and predicted values.

$$MAPE = (1/n) * \Sigma(|(y_i - \hat{y}_i)/y_i|) * 100$$

Mean Squared Error (MSE) is calculated by differencing predicted values from actual values and squaring them. This ensures removal of any negative signs.

$$MSE = (1/n) * \Sigma((y_i - \hat{y}_i)^2)$$

And Root Mean Squared Error takes the square root of the calculated MSE. With MSE, the model's badness can be underestimated if the error values are less than 0, hence RMSE is used. See the below results as an example of how model was evaluated using the four error matrices mentioned above.

mape_arima	mae_arima	mse_arima	rmse_arim
0.09893	24.242244	678.183247	26.04195

EXHIBIT IV :The model's performance is evaluated using error matrices to see how much the forecasted values deviates from the actual value.

VI. CONCLUSIONS

In this article, we have shown how portfolio performance can be analyzed effectively by employing various statistical methods. The three features of the tool helps users get an insight about the changing in closing price of the asset to predicting the portfolio return for a 30-day horizon. The asset comparison section provided valuable insights based on individual assets historical data. Factors like volatility, drawdown allow users to assess the risk associated with the return of each asset.

Portfolio performance analysis feature focuses on visualizing the performance of the portfolio based on the weights allocated to each asset. Along with that, it highlights the total profit or loss that is incurred for the amount of money invested in the date range. The final feature forecasts the price for the next 30 days of the end date and calculates net profit or loss incurred on the forecasted period.

Allowing users to enter the data as per their selection gives them more manual control and provides results for their desired asset and weights allocated. By employing statistical learning models and considering various factors such as asset returns, volatility, drawdowns, and risk-adjusted performance, investors can gain valuable insights into portfolio management and make informed decisions to optimize their investment strategy.

For future scope, we wish to employ other learning models and compare their results with the ARIMA model to choose the best performing model, thriving for enhanced forecast accuracy. We also plan to devise analysis to suggest the weights to be allocated for optimized portfolio return. The tool is converted to a front-end web page for better user experience with the help of Streamlit. The web page was locally hosted for the illustration purpose. In future, for even enhanced user experience the tool can be created as a fully functional web page application and hosted online

ACKNOWLEDGEMENT

The authors would like to thank professor Ozgur Ozturk for guiding us throughout the project and professor Abdullah Karasan, PhD for sharing his expertise throughout this project.

REFERENCES

[1] Victorio Collin, "*Portfolio Performance*". Retrieved from, https://www.fe.training/free-resources/portfolio-management/portfolio-performance, (August, 2021)[2] Adam Hayes, "*Portfolio Management: Definition, Types, and Strategies* ". Retrieved from, https://www.investopedia.com/terms/p/portfoliomanagement.asp, (updated March, 2023)

[3] Sharpe,W.F., "*Mutual Fund Performance.*" The Journal of Business, https://www.jstor.org/stable/2351741, (Jan, 1966).

[4] Jensen, M.C., "*The Performance of Mutual Funds in the period of 1945-1964*",The Journal of FInance, https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.1968.t00815.x, (May, 1968)

[5] Sima S.N., Neda T., Akbar S.N., "*A Comparison of ARIMA and LSTM in Time Series Forecasting.*" IEEE. Retrieved form https://par.nsf.gov/servlets/purl/10186768, (2018)

[6] Spyros M, Evangelos S, Vassilios A, Artemios A.S, Gary M, Konstantinos N, "*Statistical, Machine Learning and Deep Learning Forecasting Methods: Comparison and Ways Forward*." The Journal of Operational Research Society.https://www.tandfonline.com/doi/full/10.1080/01605682.2022.2118629 (September, 2022)

[7] Federal Reserve Economic Data (FRED) https://fred.stlouisfed.org/docs/api/fred/

[8] CFI Team, "*Risk-Free Rate*" CFI, Updated March 2023. https://corporatefinanceinstitute.com/resources/valuation/risk-free-rate/

[9] Adam H, Charles P, Pete R., "*Volatility: Meaning in Finance and How it works with Stocks".* Retrieved from https://www.investopedia.com/terms/v/volatility.asp, (updated March, 2023)

[10] Adam H, Gordon S., "*Maximum Drawdown (MDD) Defined, with Formula for Calculation".* Retrieved from https://www.investopedia.com/terms/m/maximum-drawdown-mdd.asp, (updated March, 2023)

[11] Jason F., Margaret J., Katrina M., "*Sharpe Ratio formula and definition.*" Retrieved from (updated March, 2023) https://www.investopedia.com/terms/s/sharperatio.asp

[12] Lleyton A. "*A Thorough Introduction to ARIMA Models*" Retrieved on January 2021 form https://medium.com/analytics-vidhya/a-thorough-introduction-to-arima-models-987a24e9ff71