Stock Investment Risk Analysis

Executive Summary

By Sarika Gaind

The primary motivation is to perform risk analysis and investigate the profile of the customer by using financial stocks which will be separated into different categories that are Low- Level risk and High-level risk. Our mission is to use data driven strategies to optimize our client's investment objectives. For our analysis, we will be evaluating the volatility of stocks based on their "beta" values.

The target variable for regression is the daily returns of individual stocks. The regressions are further used to determine the risk of returns, by examining the regression coefficient "beta" which is a common measure of a stock's volatility (Kenton, 2024).

We fetched data for Tesla, Apple, Microsoft, and Walmart from Yahoo Finance. We also fetched the returns from the S&P 500. The S&P 500 index is commonly used as the benchmark for calculating beta which will help to measure the relationship between two variables and S&P 500 is a widely followed stock market index that includes 500 large-cap companies from various industry sectors (Beers, 2023).

We utilized libraries such as quantstats, numpy, matplotlib, seaborn, scikitlearn to process, clean, visualize, and model our data with Python. After downloading the datasets, we examined and cleaned it to ensure that the data is consistently formatted with no missing values or otherwise dirty data. This involved identifying and managing outliers. After data cleaning, boxplots were generated for each individual stock and highlighted with distinct colors to see the outliers are shown as dots on boxplot. We visualized the relationships between the four companies and their returns with scatter plots for each company. For example, the relationship between the Apple and Microsoft returns and the S&P 500 returns seems fairly linear. On the

other hand, the relationship between the Tesla and Walmart returns and the S&P 500 returns appear less linear, and a different kind of model may be a better fit.

A risk classification function was used for our goal of categorizing stocks as low-risk or high risk. A beta value above 1.0 indicates the stock more volatile than the market overall. Based on our data exploration and visualizations, we decided to try linear regression models and polynomial regression models. The slope of this line is the beta-value we are seeking to evaluate the stock's riskiness. We ran multiple model types because we would like to use the beta value from the model that best fits our data. We ran linear regression and polynomial regression for each of the four subject stocks, and generated a visualization, the beta-value, the Mean Square Error (MSE) and the R-square for the model.

The coefficient of determination ("R-Square") is used to measure how well a model fits the data. R-square measures the proportion of the variation in the target variable that is explained by the regression model and analyzes how well the model fits (Jaggia et al., 2022). Based on our R-Square values, the polynomial models fit the data slightly better than the linear models for all stocks. However, the difference was very small. Based on our analysis of the beta values from each regression, we concluded that Apple, Tesla, and Microsoft are high-risk options relative to the market, and Walmart is low risk. This information can be used to help us optimize our clients' portfolios based on their preferences and risk-tolerance.

References:

Beers, B. (2022, August 30). *Why do Investors Use the S&P 500 as a Benchmark.* Investopedia. https://www.investopedia.com/ask/answers/041315/what-are-pros-and-cons-using-sp-500-benchmark.asp

Jaggia, S. & Kelly, A. (2022). *Business Statistics: Communicating with Numbers* (4th ed.). McGraw Hill LLC.

Kenton, W. (2024, February 22). *Beta: Definition, Calculation, and Explanation for Investors*. Investopedia. https://www.investopedia.com/terms/b/beta.asp