# The Evolution of Portfolio Theory: Integrating Machine Learning with Markowitz Optimization

*Junhao* Xu*

School of Economics, The University of Edinburgh, Edinburgh EH89YL, UK

**Abstract.** Modern Portfolio Theory (MPT), developed by Harry Markowitz, transformed investment practices by wisely balancing risk and return. Nonetheless, its efficacy wanes in fluctuating financial markets due to its dependence on historical data and fixed assumptions. This paper investigates incorporating Machine Learning (ML) techniques into the traditional Markowitz optimization framework to enhance portfolio construction and risk management processes. It highlights the use of supervised learning for forecasting asset returns, unsupervised learning for asset clustering, and reinforcement learning for adjusting portfolios dynamically. An empirical analysis utilizing recent U.S. market data reveals that ML models improve risk assessment, asset selection, and adaptive portfolio allocation. Techniques such as linear regression, clustering algorithms, and principal component analysis (PCA) facilitate superior forecasting and portfolio design in various market environments. The research also shows that ML can enhance Sharpe ratios in specific market conditions compared to conventional MPT. ML increases portfolio flexibility and robustness by aligning predictive modeling with optimization objectives. This evolving methodology lays the foundation for a more responsive, data-informed investment strategy in today's finance, providing a viable alternative to the limitations of traditional models.

## 1 Introduction

The changing landscape of financial markets demands ongoing improvements in investment strategies. Since Harry Markowitz introduced Modern Portfolio Theory (MPT) in 1952, it has played a significant role in finance by providing a quantitative approach to balance expected returns against associated risks [1]. Through mean-variance optimization, MPT empowers investors to build portfolios that maximize returns for a specified level of risk or minimize risk for a target return. The core equations define expected portfolio return as a weighted average of individual asset returns and portfolio variance based on asset variances and covariances. Although MPT is elegant and impactful, it is limited by its assumptions: dependence on historical data, return normality, fixed correlation structures, and a static single-period model. These constraints weaken its effectiveness in addressing real-world market behaviors, especially during volatility, structural breaks, or regime shifts. The

* Corresponding author: s2297956@ed.ac.uk

foundational role of MPT in shaping modern investment practices and the continued academic dialogue around its evolution is well-documented in the literature [2].

This research seeks to tackle these challenges by integrating Machine Learning (ML) techniques within the traditional Markowitz optimization framework. It investigates how data-driven, adaptive models can improve portfolio construction, risk management, and overall investment performance. Unlike conventional statistical models, ML algorithms can identify non-linear relationships, update in real time, and adjust to changing market conditions. This study employs a hybrid methodology, using ML techniques such as supervised learning (e.g., linear regression) to predict asset returns, unsupervised learning (e.g., clustering and principal component analysis) to uncover hidden patterns and market states, and reinforcement learning principles to support dynamic asset reallocation. This research presents a decision-focused learning framework to enhance real-world investment results by connecting predictive modeling with optimization objectives.

The methodology focuses on an empirical analysis utilizing U.S. market data spanning February 2022 to February 2025. This dataset includes a varied range of assets from different sectors and regions, such as technology (Apple, Amazon), energy (ExxonMobil), finance (HSBC), emerging markets (Tencent, BYD), and alternative assets (Bitcoin, gold futures). Monthly closing prices sourced from Yahoo Finance underwent analysis using Markowitz optimization to pinpoint portfolios that maximize the Sharpe ratio, considering both scenarios with and without short-selling constraints. Initial findings indicated that permitting short-selling resulted in extreme portfolio weights and significantly high Sharpe ratios, whereas imposing no-short-selling limits diminished expected returns and diversification.

Acknowledging the volatility of results due to minor input changes—an established limitation in MPT—the study opted to employ ML models to enhance predictions and stability. A supervised learning model was created in R utilizing linear regression, with Amazon's monthly stock price as the outcome variable. Independent variables encompassed various other asset prices. Notably, Apple and Bitcoin demonstrated statistically significant positive impacts on Amazon's price, while several different variables (such as BYD, Gold, and Tencent) were excluded due to their lack of significance. This streamlining bolstered model reliability, reflected by a high adjusted $R^2$ (0.9371) and F-statistics. Leveraging this predictive framework, the research devised a new ML-predicted portfolio by allocating weights in proportion to each asset's predictive strength.

The study utilized unsupervised learning methods, specifically clustering and principal component analysis (PCA). Clustering methods organized monthly return data into distinct market states, highlighting bullish trends, stability, or high volatility periods. PCA was used to decrease dimensionality and uncover significant variance patterns across the asset universe. These approaches offered valuable insights into portfolio performance across different market conditions. For instance, the machine learning-driven portfolio outperformed those based on modern portfolio theory (MPT) during volatile or bearish markets by dynamically adjusting its exposures—especially by cutting back or completely removing investments in Amazon when market conditions worsened.

We calculated each market cluster's Sharpe ratio, expected return, and standard deviation to assess model effectiveness. The findings revealed that MPT is superior in stable conditions, whereas ML-driven strategies show increased robustness during turbulent times. A comparative table and visual representation demonstrated that hybrid methods—utilizing Markowitz optimization in stable environments alongside ML in more volatile conditions—produce the most reliable risk-adjusted returns. This emphasizes the importance of merging traditional finance concepts with contemporary AI technologies to develop flexible portfolio strategies that accommodate the complexities of real-world markets.

This paper is organized as follows: Section 2 examines the relevant literature on MPT, highlighting its critiques and advancements via machine learning; Section 3 details the data sources, the methodology for portfolio construction, and the ML implementation. Section 4 provides empirical results, comparing traditional portfolios with those integrated with ML across various market conditions. Section 5 explores the implications of the findings, including practical challenges like short-selling restrictions and issues related to market liquidity. Lastly, Section 6 wraps up with thoughts on future research, emphasizing that greater integration of deep learning and real-time analytics could advance adaptive investment management.

This research proves that incorporating machine learning into MPT can significantly improve portfolio optimization. It adds to the expanding body of literature on hybrid financial models by showing that predictive analytics, when aligned with optimization goals, allows for investment strategies that are more robust, flexible, and attuned to market conditions. As financial markets evolve with technological changes and rising uncertainty, these integrated methods are not merely beneficial; they are essential.

# 2 The Modern Portfolio Theory (MPT)

## 2.1 Theoretical framework

Harry Markowitz's mean-variance optimization model established the basis for Modern Portfolio Theory (MPT) by prioritizing the balance between expected returns and their associated risks [1].

His model aimed to minimize portfolio variance while maximizing expected returns. A portfolio's expected return (E(Rp)) is the weighted sum of the expected returns of the individual assets.

$$E(R_p) = \sum_{i=1}^{n} w_i * E(R_i) \tag{1}$$

Where $W_i$ is the weight of each asset, $E(R_p)$ is the portfolio's expected return, and $E(R_i)$ represents the expected return of each asset.

Meanwhile, Portfolio variance ($\sigma_p^2$) measures the portfolio's total risk.

$$\sigma_p{}^2 = \sum_{i=1}^{n} w_i{}^2 \sigma_i{}^2 + 2\sum_{i=1}^{n} \sum_{j=i+1}^{n} w_i * w_j * \sigma_{ij} \tag{2}$$

Where $\sigma_p^2$ is the portfolio variance, $\sigma_i^2$ is the variance of the asset, and $\sigma_{ij}$ is the covariance between asset $i$ and asset $j$.

Since the goal of the Markowitz optimization is to minimize portfolio variance subject to constraints, the method is interpreted as the following:

$$min\, \sigma_p{}^2 = \sum_{i=1}^{n} \sum_{j=i+1}^{n} w_i * w_j * \sigma_{ij} \tag{3}$$

Moreover, the model analysis is subject to several constraints, including the expected return constraint ($E(R_p)$), target portfolio return, weight constraint (sum of $W_i = 1$), and relevant shorting constraint (for example, $W_i \geqslant 0$).

## 2.2 Portfolio analysis

### *2.2.1 Data collection*

In this paper, we analyzed the monthly historical stock prices of companies from various industries in the US, including both traditional and emerging ones, from February 1, 2022, to February 1, 2025, in the US market. We select HSBC Holdings plc- one of the world's largest banking and financial services organizations, with its Asia-Pacific headquarters in Hong Kong; XOM (US): Exxon Mobil Corporation, a prominent American multinational oil and gas company known for its global energy operations and petroleum products; Amazon (US) - a leading U.S. e-commerce and cloud computing giant, widely recognized for revolutionizing online retail; Apple Inc.- a premier U.S. technology company known for its innovative consumer electronics, including the iPhone, Mac, and services ecosystem; Tencent Holdings Limited- a Chinese tech conglomerate headquartered in Shenzhen, best known for its social media, gaming, and fintech platforms like WeChat; BYD Company Limited- a major Chinese manufacturer of electric vehicles and batteries, with a growing international presence in green energy solutions. Moreover, we include gold futures, which are standardized contracts traded on exchanges to buy or sell gold at a future date and price, commonly used for hedging or speculation, and Bitcoin- the first and most prominent decentralized digital cryptocurrency, enabling peer-to-peer transactions on a blockchain network- into our portfolio to diversify the returns and variances.

We downloaded the statistics from Yahoo Finance regarding the monthly closing prices for the first day of each month during the period [3]. We intended to implement the Markowitz optimization to achieve the highest Sharpe ratio (which compares the return of an investment with its risk) in our portfolio, where the formula for the Sharpe ratio is given.

$$S = \frac{E(R_p) - R_f}{\sigma_p} \tag{4}$$

We used Excel to calculate the optimal portfolio, using the monthly overnight SOFR offer rate on 28th March 2025 as a reference for the risk-free rate, which is 4.33% [3,4]. Meanwhile, we plot the efficient frontier, which shows the different levels of risk for a given expected return.

### *2.2.2 Results*

Initially, we calculated the monthly return and obtained the average. After measuring the demeaned adjusted returns, which ensures that the returns are stationary and can be compared across different sample sizes, we determined the covariance matrix of the adjusted returns between each asset and its inverse. The formula for demeaned returns is provided below:

$$r_{adjusted} = \frac{r_t - \hat{r}}{\sqrt{T}} \tag{5}$$

Within Markowitz's original mean-variance optimization framework, we employ the following formula to derive the intermediate scalar:

$$A = 1^{-1}\sum 1 \qquad B = 1^{-1}\sum r \qquad C = r^{-1}\sum r \qquad D = AC - B^2 \tag{6}$$

The next step is calculating each asset's portfolio variance using the efficient frontier formula for a given R.

$$\sigma^2 = \frac{AR^2 - 2BR + C}{D} \tag{7}$$

Therefore, we could plot the scatter points of each asset on a chart and illustrate the efficient frontier along with the tangent line corresponding to the specific risk-free rate above them. Figure 1 visualizes the efficient frontier derived from Modern Portfolio Theory (MPT), plotting expected returns (Y-axis) against volatility or risk (X-axis) for both individual assets and constructed portfolios. The blue curve signifies the efficient frontier, and the tangent point where the orange line (capital market line) touches the curve represents the optimal portfolio, maximizing the Sharpe ratio. Bitcoin lies far to the right, indicating high volatility (18%) and high return (3.3%), making it a risky yet potentially rewarding asset. Meanwhile, Gold Futures, HSBC, Apple, and XOM are clustered in the lower left, representing low-risk, moderate-return assets. Additionally, Amazon, Tencent, and BYD fall into a mid-risk, mid-return range, with Tencent showing a moderate return but elevated volatility.

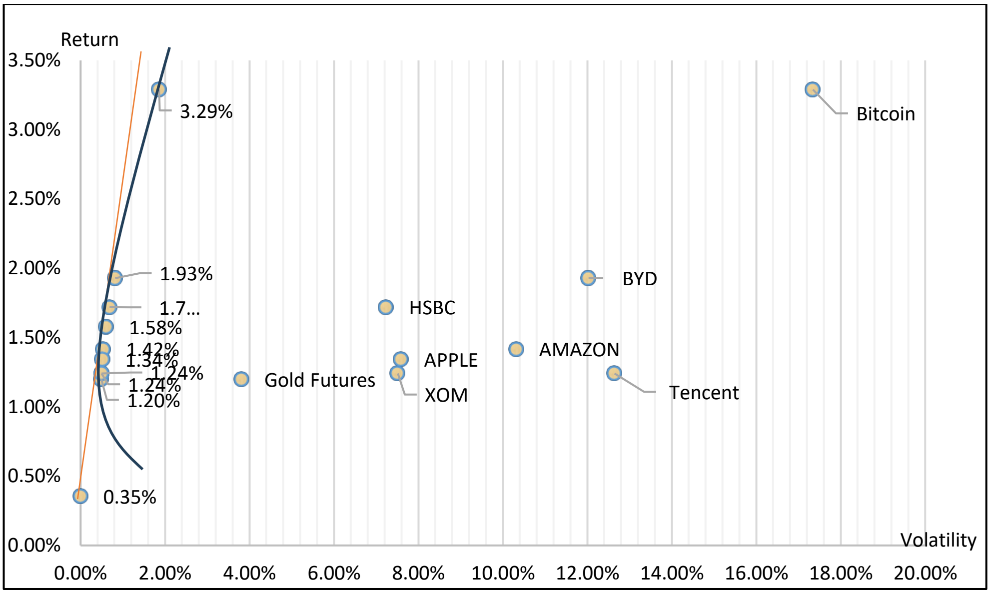


**Fig. 1.** Efficient Frontier and Optimal Portfolios under MPT

Our portfolio scatter plot demonstrates a poor fit with the efficient frontier, indicating a lack of portfolio diversity and a low correlation among asset prices over the specified period.

Furthermore, we could use the Excel solver to determine the optimal weight of each asset in the MPT and get the results.

The weights of each asset in the optimal portfolio indicate short-selling positions of 16% in Amazon and 15.39% in Tencent, along with long positions of 32% in HSBC, 66% in Gold Futures, 4% in XOM, 15.08% in Apple, 9.53% in Bitcoin, and 3.61% in BYD, as shown in Table 1. The beta here measures the systematic risk of an individual asset relative to the overall market; Bitcoin, with a beta value of 2.4, suggests higher volatility than the market, while Tencent, with a beta of 0.73, indicates more excellent stability. The SML represents the security market line, displaying the expected return of an asset based on its beta under the Capital Asset Pricing Model. Since all holdings in Table 1 share the same SML value of 0.012217, this means that under CAPM, each asset's required return based on its risk is calculated using a uniform risk-free rate and market premium.

**Table 1.** Portfolio Summary for MPT

| | Covariance | Beta | SML | weights | Expected return |
|---|---|---|---|---|---|
| HSBC(HK) | 4.1277E-05 | 1.117490658 | 0.012217 | 32% | 1.5756% |
| Gold Futures | 2.5569E-05 | 0.692227954 | 0.012217 | 66% | variance |
| XOM(US) | 2.6816E-05 | 0.725988445 | 0.012217 | 4% | 3.69372E-05 |

| | | | | | |
|---|---|---|---|---|---|
| Amazon(US) | 3.21276E-05 | 0.86978848 | 0.012217 | -16% | std dev |
| Apple(US) | 2.98629E-05 | 0.808477498 | 0.012217 | 15.08% | 0.006077602 |
| Bitcoin | 8.87578E-05 | 2.402935253 | 0.012217 | 9.53% | |
| Tencent(HK) | 2.68207E-05 | 0.726116031 | 0.012217 | -15.39% | Sharpe ratio |
| BYD(HK) | 4.76012E-05 | 1.288703582 | 0.012217 | 3.61% | 2.0102 |

Meanwhile, the E(Rp) equals 1.5756%, and the portfolio variance is 0.0037%. The high shape ratio (2.01) indicated that investors receive high returns compared to the extremely low volatility they experienced, which is uncommon in real-world scenarios. Moreover, it also suggested a short-term market anomaly limited to future prediction.

Nevertheless, many institutions or investors are restricted or reluctant to short-sell, and option trades are highly professional. Short selling also incurs borrowing costs and exposes investors to unlimited losses, functioning as a high-leverage financial derivative. Moreover, some assets may face liquidity issues that complicate short selling due to borrowing constraints.

If we add another no-short-selling constraint to the model, our results change significantly.

As shown in Table 2, the portfolio's expected return decreased to 1.48%, and the Sharpe ratio declined by 20%. However, the portfolios became concentrated in three securities: buying HSBC (23%), XOM (14%), and Apple (5.71%). Additionally, they took long positions on Gold Futures and spent 7.05% on Bitcoin, which reduced diversification, portfolio returns, and the Sharpe ratio. Moreover, the beta value of each asset increases dramatically, leading to different SML values, indicating a higher systematic risk in our portfolio.

**Table 2.** Portfolio Summary for MPT under no-shorting constraint

| | Covariance | Beta | SML | weights | Expected return |
|---|---|---|---|---|---|
| HSBC(HK) | 4.55454E-05 | 1.233047033 | 0.011072 | 23% | 1.4796% |
| Gold Futures | 2.82132E-05 | 0.763813489 | 0.011072 | 51% | variance |
| XOM(US) | 2.95891E-05 | 0.801064343 | 0.011072 | 14% | 3.75544E-05 |
| Amazon (US) | 4.98255E-05 | 1.348922145 | 0.007878 | 0% | std dev |
| Apple(US) | 3.29511E-05 | 0.892084215 | 0.011072 | 5.71% | 0.006128165 |
| Bitcoin | 9.79368E-05 | 2.651435341 | 0.011072 | 7.05% | |
| Tencent (HK) | 5.86507E-05 | 1.587847221 | 0.005587 | 0.00% | Sharpe ratio |
| BYD(HK) | 5.33021E-05 | 1.443042907 | 0.010911 | 0.00% | 1.8370 |

There are several drawbacks and limitations to traditional MPT. One disadvantage is the instability of input parameters. Specifically, MPT relies on historical data, and small changes or estimation errors in these parameters can significantly alter the optimal weights, indicating this instability. For instance, recent work defends the relevance of optimization frameworks, arguing that naive allocation strategies like 1/N fail to capture portfolio efficiency [5]. Moreover, the Markowitz model assumes that asset returns are normally distributed, overlooking skewness, kurtosis, and fat tails in returns, which could underestimate the risk in actual financial returns. Additionally, as investment decisions and markets are dynamic and continuously evolving in practice, MPT, viewed as a static and single-period model, would lead to uncertain changes in the optimal portfolio.

# 3 Machine learning in portfolio optimization

Integrating Machine Learning (ML) in finance has dramatically transformed the industry. ML facilitates the analysis of intricate, nonlinear patterns in financial data, successfully overcoming the constraints of conventional economic models [6]. This advancement is especially noticeable in portfolio optimization, where blending ML methods with the Markowitz Mean-Variance Model enhances decision-making strategies. For instance, ML has been applied to predict leverage dynamics better, outperforming traditional linear models by identifying a broader set of leverage determinants [7]. Furthermore, Integrating ML techniques into this framework also addresses these challenges by providing more robust estimations and capturing nonlinear dependencies among assets.

## 3.1 Empirical findings

There have been some empirical findings on machine learning strategies recently. An innovative strategy integrates convolutional neural networks (CNN) with bidirectional long-short-term memory (BiLSTM) models to forecast stock returns [8]. The Markowitz model utilizes these forecasts, resulting in portfolios that exceed the performance of conventional methods concerning risk-adjusted returns.

Recent research also indicated that incorporating prediction and optimization through decision-focused learning modifies prediction models to enhance decision-making outcomes in portfolio optimization [9]. This strategy yields more efficient portfolios by directly aligning predictive modeling with optimization objectives.

## 3.2 Implementing machine learning

### *3.2.1 Regression model*

In this essay, we discussed how ML models can adapt to real-time data, capture linear or nonlinear relationships, and enhance risk estimation, thereby addressing some limitations of traditional MPT. The first framework we utilize is called Supervised Learning, where we employ various regression models to predict future stock returns based on historical data.

To investigate the relationship between various assets and explore how these relationships could influence the prediction of future stock prices, we set Amazon's monthly closing price as our response variable. We built a linear regression model in R [10].

We split the dataset into training and testing sets (80% for training and 20% for testing), stratifying according to Amazon’s prices. This splitting helps select indices systematically, ensuring that approximately 80% of data points become training data and 20% become testing data. This approach maintains a consistent distribution of Amazon's prices between both subsets, reducing bias and ensuring the model is tested on representative data, which makes predictions more reliable. Next, we perform the linear regression model on other stock prices in R and get the result as shown in Table 3.

**Table 3.** Linear regression statistics in R

| Variable | Estimate | Std. Error | t Value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 89.93529 | 24.85111 | 3.619 | 0.00137 |
| Apple | 0.37503 | 0.143307 | 2.617 | 0.01511 |
| Bitcoin | 0.001232 | 0.000237 | 5.207 | 2.46E-05 |
| BYD | 0.034268 | 0.075621 | 0.453 | 0.65451 |

| | | | | |
|---|---|---|---|---|
| Exxon | -0.48833 | 0.272058 | -1.795 | 0.08527 |
| Gold | -0.01534 | 0.026044 | -0.589 | 0.56128 |
| HSBC | 0.778961 | 0.571271 | 1.364 | 0.18536 |
| Tencent | -0.10135 | 0.065362 | -1.551 | 0.13409 |
| Adjusted R-squared | | 0.9371 | | |
| F-statistic | 66.97 on 7 and 24 DF | | | |

Based on the results, we could then set a linear equation for the Amazon stock price.

$$Amazon.price = 89.935 + 0.375 * Apple.price + 0.001 * Bitcoin.price + 0.034 * BYD.price - 0.488 * XOM.price - 0.015 * Gold.price + 0.779 * HSBC.price - 0.101 * Tencent.price \quad (8)$$

However, since the p-values of BYD, Gold futures, HSBC, and Tencent are more significant than 0.1, the four variables in our model are statistically insignificant and may incur severe bias. Therefore, after excluding these inadequate variables, we conclude that, under the linear assumption, Apple's stock price could positively influence Amazon's stock price of Amazon (price, with an increase of $1 in Apple’s price being associated with approximately a $0.38 increase in Amazon’s price. Additionally, Bitcoin’s price has a positive relationship with Amazon’s, but only to a small extent. The rise in Exxon’s prices could lead to a decline in Amazon’s price, indicating that investors might consider inverse diversification opportunities between Exxon and Amazon.

When analyzing the model’s performance, we examined the adjusted R-squared and F-statistics, which are 0.9371 and 66.97, respectively. The adjusted R-squared indicates an excellent fit for the model, explaining approximately 93.71% of the variation in Amazon’s stock prices. Furthermore, since we have high F-statistics, the joint statistical significance of all predictor variables is confirmed, strongly suggesting that the model is reliable for predicting Amazon’s price.

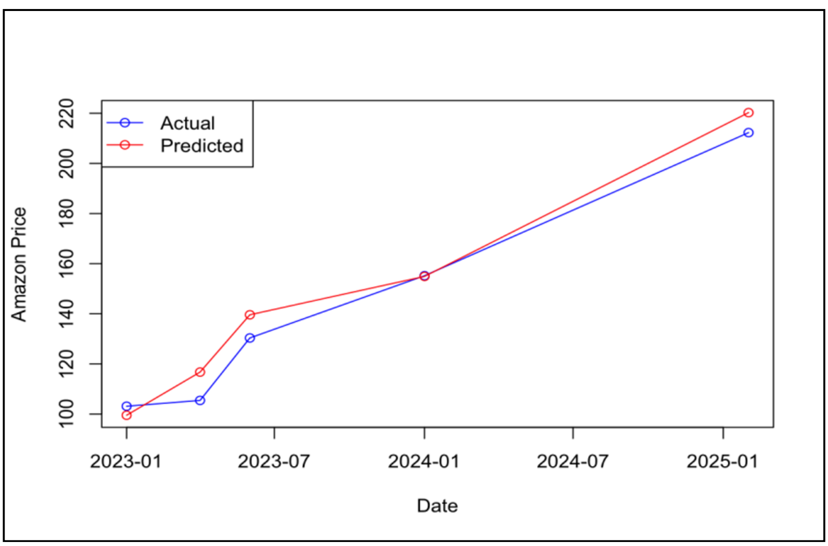


**Fig. 2.** Actual VS Predicted Amazon Prices

The Amazon price is predicted using historical monthly data from February 1, 2022, to February 1, 2025, based on the prediction rule of the linear regression model in R. We examine a chart (Figure 2) that shows the prediction line compared to the actual price over the specified dates.

We derive a new portfolio summary after removing insignificant variables from our MPT model, which have their weights assigned to zero. We apply the Markowitz optimizations again to achieve the highest Sharpe ratio using only four variables and produce a new portfolio summary as presented in Table 4.

**Table 4.** Portfolio summary for MPT optimization when eliminating insignificant factors based on the prediction of Amazon's price.

| | Covariance | Beta | SML | weights | Expected return |
|---|---|---|---|---|---|
| HSBC(HK) | 8.86353E-06 | 0.239961771 | 0.056895 | 0% | 2.0206% |
| Gold Futures | 1.64995E-05 | 0.446690345 | 0.018933 | 0% | variance |
| XOM(US) | 9.66448E-05 | 2.616459096 | 0.00339 | 40% | 0.000181608 |
| Amazon(US) | 0.000115788 | 3.134714868 | 0.00339 | -23% | std dev |
| Apple(US) | 0.000107626 | 2.913750272 | 0.00339 | 45.34% | 0.013476202 |
| Bitcoin | 0.000319883 | 8.660175171 | 0.00339 | 37.83% | |
| Tencent(HK) | -2.6024E-05 | -0.704547452 | -0.01259 | 0.00% | Sharpe ratio |
| BYD(HK) | -2.3816E-06 | -0.064477978 | -0.24418 | 0.00% | 1.2368 |

Under the linear regression assumption, the new portfolio's expected return now increases to 2.02%, but the Sharpe ratio drops to 1.237. Our portfolio strategy involves short-selling Amazon and taking long positions in XOM, Apple, and Bitcoin. Therefore, we can conclude that if we set the Amazon price as a pivot in the portfolio, Markowitz's optimization among relevant stock choices would lead to a lower Sharpe ratio, reflecting a less financially efficient model.

Instead, we could develop our first machine-learning strategy and refer to it as the ML-predicted model based on linear regression analysis. We apply the weighting rule corresponding to a portfolio where assets that more accurately predict Amazon's returns are assigned higher weights. Meanwhile, since Amazon stock is the response variable, we incorporate it into the ML-predicted model. The following step is to normalize the absolute values, which are then scaled so that all weights total one based on each prediction force.

**Table 5.** Portfolio summary for ML-predicted model

| | Covariance | Beta | SML | weights | Expected return |
|---|---|---|---|---|---|
| HSBC(HK) | 8.86353E-06 | 0.239961771 | 0.056895 | 43% | 1.4842% |
| Gold Futures | 1.64995E-05 | 0.446690345 | 0.018933 | 1% | variance |
| XOM(US) | 9.66448E-05 | 2.616459096 | 0.00339 | 27% | 6.75152E-05 |
| Amazon(US) | 0.000115788 | 3.134714868 | 0.00339 | 0% | std dev |
| Apple(US) | 0.000107626 | 2.913750272 | 0.00339 | 20.90% | 0.008216764 |
| Bitcoin | 0.000319883 | 8.660175171 | 0.00339 | 0.10% | |
| Tencent(HK) | -2.6024E-05 | -0.704547452 | -0.01259 | 5.60% | Sharpe ratio |
| BYD(HK) | -2.3816E-06 | -0.064477978 | -0.24418 | 1.90% | 1.3757 |

The portfolio summary in Table 5 shows that HSBC holds 43%, Gold Futures 1%, XOM 27%, Apple 20.9%, Bitcoin 0.1%, Tencent 5.6%, and BYD comprises 1.9%. Currently, we have an expected return of 1.48%, which is lower than the 2.02% in MPT, and a Sharpe ratio of 1.3757, which exceeds the 1.2368 in MPT, enhancing the model's performance due to the previous goal of maximizing the Sharpe ratio.

### *3.2.2 Unsupervised learning model*

Secondly, we could apply clustering algorithms and use an unsupervised learning model to identify patterns or group similar assets, which would aid our diversification strategies. After clustering our portfolio data by the monthly dates during the period, we reached three clustering groups, as illustrated in the graph. Each color in the clustering set represents a distinct market state or behavior (e.g., bullish, bearish, volatile periods).

We will discuss principal component analysis (PCA), a statistical technique that simplifies data by identifying principal components and linear combinations of the original variables that capture the most significant variance [10].

After performing clustering and PCA in R, we obtained the blue arrows highlighting the loading of each asset and three clusters in Figure 3. PC1 and PC2 together explained 83.81% of the point variability, meaning that these two dimensions capture most of the relevant patterns in the data. Practically, the purple cluster (comprising observations 11–29) forms a tightly packed group far from the origin. This likely represents a stable regime, indicating similar asset movements and possibly trending markets. The second cluster in blue (comprising observations 1–10) may represent bullish or momentum periods, mainly when high returns are observed. Thirdly, the red cluster exhibits more dispersed patterns, indicating volatile market conditions with diverging asset behaviors. Since Tencent's arrow overlapped with the red cluster, it likely reflects periods where Tencent had a strong influence — possibly due to significant swings.

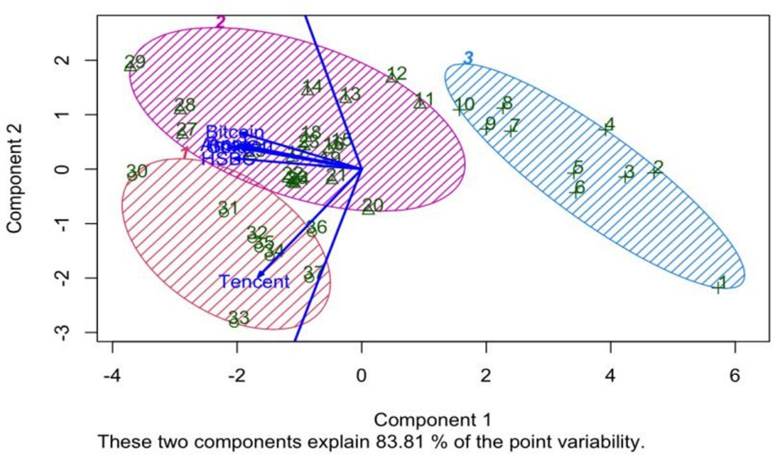


**Fig. 3.** K-mean clustering plot with two PCA components of dates

If we classify the data explicitly by asset names, we have another simple PCA graph, as shown in Figure 4. Amazon, Bitcoin, Apple, Gold, and HSBC are tightly grouped, Tencent and BYD are on their own, and Exxon appears in a distinct group. This also matches the patterns of loadings in the previous graph.

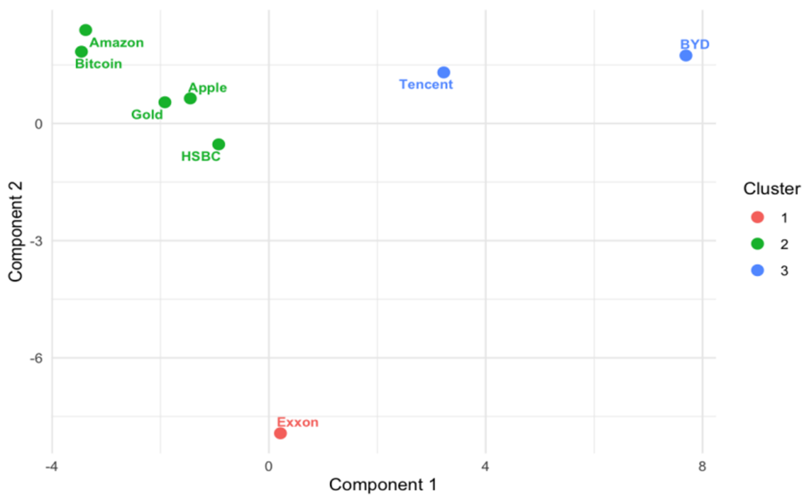


**Fig. 4.** K-mean clustering plot with two PCA components of assets

Surprisingly, instead of comparing portfolios overall, we could see how they perform under different market conditions (in other clusters), which is more realistic and insightful. Suppose we apply the MPT weights from the portfolio summary for MPT to the clustering model after using the group in R. In that case, we reach Table 6, showing the mean return and Sharpe ratio in different market conditions.

**Table 6.** Portfolio summary under MPT in different clusters.

| Cluster | Mean Return | Standard Deviation | Sharpe Ratio |
|---|---|---|---|
| 1 | -0.9675111 | 0.229567 | -4.2301858 |
| 2 | 1.3040432 | 0.417813 | 3.1125039 |
| 3 | -0.2789654 | 0.392826 | -0.7193151 |

In cluster two, where the market condition is bullish, we could use the MPT to optimize the Sharpe ratio, yielding an extremely high value of 311%. However, there are superior optimization methods under different market conditions compared to the MPT, which can lead to losses.

Since we had previously constructed the linear regression model and made predictions, we can utilize those results to apply the ML-predicted model here. After scaling and using Amazon's prediction for the entire dataset, we reached an interesting outcome, as shown in Table 7.

**Table 7.** Portfolio summary under ML-predicted model in different clusters

| Cluster | Mean Return | Standard Deviation | Sharpe Ratio |
|---|---|---|---|
| 1 | 0.123107 | 0.22795 | 0.524268 |
| 2 | 1.146612 | 0.783284 | 1.459256 |
| 3 | 0.097979 | 0.188259 | 0.501324 |

In Cluster 2, the ML weights are notably high, revealing that the model forecasts more significant returns from Amazon in this context. This indicates that the ML strategy should enhance its exposure to Amazon during this period. Conversely, in Cluster 3, the ML weights fall to zero, signaling a bleak outlook for Amazon. This suggests no investment or potential defensive repositioning.

Thus, we can create a chart (Figure 5) comparing ML with traditional strategies. Our data and models suggest that a hybrid approach is most effective, utilizing Markowitz in stable clusters, such as cluster 2, while considering ML-predicted exposure or limiting exposure in volatile or dangerous regimes, for example, in cluster 3.

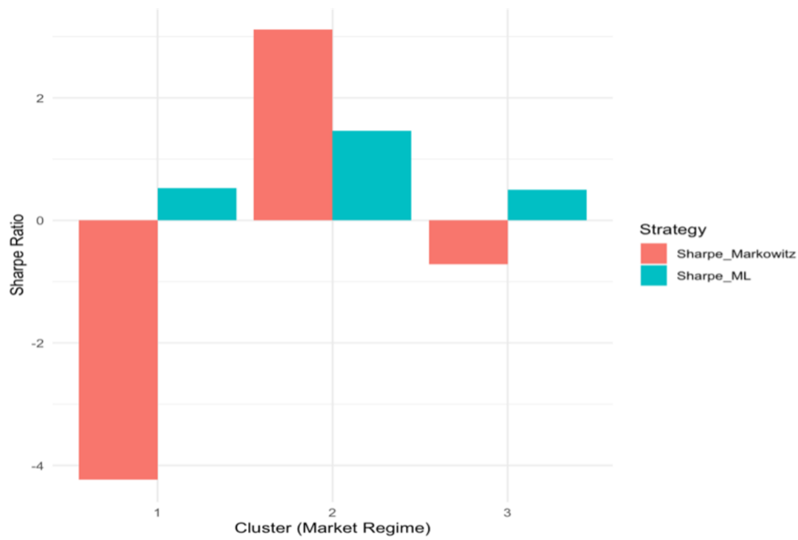


**Fig. 5.** Comparison of the Sharpe ratio for two strategies under three market regimes.

This prime example of adaptive portfolio management utilizes machine learning alongside traditional optimization methods.

## 4 Conclusion

Portfolio theory is pivotal, necessitating traditional models like Markowitz's Mean-Variance Optimization to evolve with the ever-changing financial markets. This research indicates that although the Markowitz framework offers essential insights into risk-return trade-offs and adequate diversification, it faces limitations due to its assumptions, including fixed correlations, normal distributions, and dependence on historical data.

Integrating Machine Learning (ML) into this framework helps address the shortcomings of traditional methods. Supervised learning improves return predictions; unsupervised clustering reveals hidden market regimes, and decision-focused learning connects prediction with optimization. Empirical evidence shows that ML-based strategies exceed traditional techniques, especially in volatile or non-stationary settings [11]. For example, ML-predicted portfolios exhibited greater adaptability and more accurate Sharpe ratios across various regimes identified through principal component analysis and clustering methods.

Combining machine learning (ML) with modern portfolio theory (MPT) facilitates adaptive portfolio management, merging data-driven insights with structural optimization. This study indicates that no single strategy is superior across every market condition. Instead, a customized approach—employing Markowitz optimization in stable environments and ML-based models in times of volatility—achieves better results. As the financial sector increasingly adopts automation and real-time analytics, this integration leads to a new investment strategy paradigm: robust, flexible, and forward-looking.

## References

1. H. Markowitz, Portfolio selection. J. Finance. **7**, 77–91 (1952)
2. F.J. Fabozzi, F. Gupta, H.M. Markowitz, The legacy of modern portfolio theory. J. Invest. **11**, 7–22 (2002)
3. Yahoo Finance, Historical market data (2025). Retrieved from: https://finance.yahoo.c om. Accessed 1 March 31, 2025
4. S.D. Stewart, J. Heisler, Excel-based modeling for investment management. J. Invest. Consult. **12**, 46–59 (2011)
5. M. Kritzman, S. Page, D. Turkington, In defense of optimization: The fallacy of 1/N. Financ. Anal. J. **66**, 31–39 (2010)
6. A. Vellido, R. Martí, E. Romero, Machine learning in finance: The case of unsupervised learning. Comput. Hum. Behav. **27**, 2–10 (2011). https://doi.org/10.1016/j.chb.2010.03.020
7. A. Chaweewanchon, R. Chaysiri, Markowitz mean-variance portfolio optimization with predictive stock selection using machine learning. Int. J. Financ. Stud. **10**, 64 (2022). https://doi.org/10.3390/ijfs10030064
8. X. Li, H. Xie, R. Wang, Y. Cai, J. Cao, F.Y. Wang, Empirical analysis: Financial portfolio management using deep reinforcement learning. Expert Syst. Appl. **117**, 291–307 (2020). https://doi.org/10.1016/j.eswa.2018.09.022
9. Anatomy of Machines for Markowitz: Decision-Focused Learning for Mean-Variance Portfolio Optimization. arXiv (2024). https://arxiv.org/abs/2401.00001
10. D. Ruppert, D.S. Matteson, Statistics and Data Analysis for Financial Engineering: With R Examples, 2nd ed. (Springer, 2015)
11. S. Gu, B. Kelly, D. Xiu, Empirical asset pricing via machine learning. Rev. Financ. Stud. **33**, 2223–2273 (2020). https://doi.org/10.1093/rfs/hhaa009