Machine Learning in Multi-factor Investing

INAFU6783 - Prof. Kiernan

Jiaqi Fang

Introduction

Portfolio managers' approach for asset allocation, optimization, and risk management have been transformed with the entry of machine learning (ML) into financial decision-making (Robbins 2023). Multi-factor investing has long served as a foundational approach in modern portfolio management. It allows the identification of systematic risk and the use of risk premia. Grounded in seminal frameworks like Markowitz's Mean-Variance Optimization, the Capital Asset Pricing Model (CAPM), and the Fama-French Factor Models, these strategies provide a theoretical basis for understanding asset returns and constructing diversified portfolios (Markowitz 1952). However, as financial markets become increasingly complex and dynamic, the predictive accuracy and adaptability of traditional multi-factor models are often constrained by their reliance on static assumptions.

Static assumptions such as fixed factor weights and linear correlations between factors and returns are the foundation of traditional models. Even though these assumptions simplify implementation, they fail to capture the nuanced, nonlinear dynamics that often define financial markets. A more compelling alternative is provided by recent advancements in ML. By modeling intricate, non-linear interactions and enabling dynamic factor adjustments, ML techniques enhance the predictive capabilities of multi-factor models, optimizes portfolio construction, and facilitates adaptability in volatile market environments (Coqueret 2023). These advancements align with emerging research advocating for the integration of data-driven, adaptive methodologies in factor investing strategies.This paper explores the application of machine learning in multi-factor investing through both theoretical analysis and empirical demonstration using historical financial data. By comparing ML-enhanced strategies with traditional models, it explores how ML improves predictive power, portfolio optimization, and dynamic factor weighting. The findings underscore the transformative potential of machine learning in portfolio management, providing actionable insights for practitioners seeking to navigate the complexities of modern financial markets.

Theoretical Framework

Multi-factor investing is a systematic approach that seeks to explain and capture risk premia by using measurable characteristics of securities. It is grounded in theories of modern portfolio management and asset pricing. Markowitz's Mean-Variance Optimization, introduced in 1952, provided a mathematical framework for balancing risk and return (Markowitz 1952). By focusing on diversification, it enabled investors to construct portfolios that achieved the highest possible return for a given level of risk. This work laid the foundation for quantitative approaches to portfolio management.

The CAPM was developed in the 1960s and was expanded on Markowitz's ideas by introducing beta as a measure of systematic risk. CAPM posits that an asset's return is driven primarily by its exposure to market risk, represented by its beta relative to the overall market. While straightforward, CAPM struggled to account for anomalies observed in real-world data, such as the tendency of small-cap stocks and value stocks to outperform market expectations.

To address these shortcomings, the Fama-French Three-Factor Model was developed (FF 1993). It introduced size (small minus big) and value (high minus low) factors alongside market risk. These additional factors captured the excess returns associated with small-cap and value stocks. Subsequently, the Five-Factor Model added profitability (robust minus weak) and investment (conservative minus aggressive) factors. These extensions provided a more comprehensive framework for explaining variations in asset returns across diverse market conditions. (FF 2015)

Traditional models rely on static factor weights, assuming that these exposures remain constant over time (Robbins 2023). This rigidity makes them less effective in adapting to dynamic market conditions. Furthermore, their linear assumptions oversimplify the complex, nonlinear relationships in financial data. They also struggle to process large and alternative datasets, limiting their ability to capture new market trends and interactions.

ML addresses these limitations by modeling non-linear relationships and developing complex patterns in high-dimensional data. Algorithms like random forests and neural networks can identify important factors, dynamically adjust exposures, and enhance predictive accuracy. ML also allows portfolios to adapt factor weights based on real-time market conditions, which improves their resilience to volatility and changes in economic environments.

Literature Review

The integration of machine learning into multi-factor investing has been explored extensively in recent literature, offering insights into its transformative potential for portfolio management. In *Machine Learning for Factor Investing* by Guillaume Coqueret and Tony Guida, the authors highlight how machine learning enhances factor selection and prediction accuracy by capturing non-linear relationships among financial factors. They emphasize ML's ability to prioritize impactful factors within high-dimensional datasets, which has revolutionized traditional approaches to multi-factor modeling (Coqueret 2023).

Similarly, *Factor Investing with a Deep Multi-Factor Model* explores the role of deep learning in multi-factor investing. The authors demonstrate how neural networks, by dynamically adjusting factor exposures, outperform static models in volatile markets. By applying deep learning techniques, the study reveals superior performance in both return predictions and portfolio outcomes (Wei 2022).

Expanding on these findings, *Analyzing Multifactor Investing and Neural Networks* explores the predictive power of neural networks in forecasting asset returns. The study underscores neural networks' ability to model complex, non-linear interactions among factors, significantly improving the accuracy of return forecasts which is critical in understanding ML's capacity to optimize portfolio construction (Roy 2019).

Additionally, the paper *Multifactor Stock Selection Strategy Based on Machine Learning: Evidence from China* presents an application of random forests for factor prioritization. The study employs a random forest model to assess the importance of each factor by measuring its contribution to reducing prediction error in asset returns. Factors are ranked based on their mean decrease in impurity, which quantifies their significance in predicting portfolio performance. This structured approach provides a transparent mechanism for identifying and prioritizing impactful factors, ensuring that the most influential variables are emphasized in multi-factor models. This methodology directly informs this paper's empirical analysis, guiding the selection of factors to optimize portfolio construction and improve overall performance (Gao 2023).

Data

This study incorporates monthly returns for ten selected companies, representing the percentage change in their adjusted closing prices over time. Data for historical prices were collected from Yahoo Finance spanning November 2016 to November 2024 for ten publicly traded companies: *AAPL*, the monthly return for Apple; *GOOGL*, the monthly return for Alphabet (Google); *AMZN*, the monthly return for Amazon; *NVDA*, the monthly return for NVIDIA; *PFE* reflects the monthly return for Pfizer; *PG*, the monthly return for Procter & Gamble; *AXP*, the monthly return for American Express; *JPM*, the monthly return for JPMorgan Chase; *BAC*, the monthly return for Bank of America; and *JNJ*, the monthly return for Johnson & Johnson. The companies were selected to ensure representation across diverse industries and sectors, including technology (AAPL, GOOGL, AMZN, NVDA), healthcare (PFE, JNJ), consumer goods (PG), and financial services (AXP, JPM, BAC). This diversified selection aligns with the objective of examining factor models across varied market conditions and industries.

This study also uses the Fama-French Five Factors and the risk-free rate data collected from Kenneth French's data library, which are central to the multi-factor modeling approach (FF 2015). *MKT_RF* represents the market excess return, defined as the market return minus the risk-free rate, capturing the risk premium associated with market-wide movements. *SMB* is the size factor, which reflects the return difference between small-cap and large-cap stocks. *HML*, the value factor, measures the return difference between high and low book-to-market stocks, representing the value premium. *RMW*, the profitability factor, captures the differential between returns on stocks with robust versus weak profitability. *CMA*, the investment factor, represents the difference between returns on firms with conservative versus aggressive investment policies. *RF* represents the monthly risk-free rate from the Fama-French dataset, which serves as a baseline for evaluating excess returns. These variables form the basis for analyzing factor exposures, constructing portfolios, and applying machine learning methods in this study.

Data were processed to align stock returns and Fama-French factors by time to ensure consistency. Stock returns were adjusted for monthly risk-free rate (RF) to compute excess returns. Datasets were merged into a single matrix where rows correspond to months, and columns include stock returns and factor values.

Empirical Strategy

This study uses the Fama-French Five-Factor model to predict the excess return of stock *i* at time *t* using:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{i,MK\_RF} * MKT\_RF_t + \beta_{i,SMB} * SMB_t + \beta_{i,HML} * HML_t + \beta_{i,RMW} * RMW_t + \beta_{i,CMA} * CMA_t + \epsilon_{i,t}$$

Where α_i represents the stock-specific intercept capturing unexplained returns; β_i represents the sensitivity of stock *i* to each factor; and $\epsilon_{i,t}$ represents the idiosyncratic error term

Machine learning models utilized are Linear Regression and Random Forest.

In the Linear Regression model, the predict return of stock *i* is given by:

$$\widehat{R}_{i,t} = \alpha_i + \sum_{j=1}^{5} \beta_{i,j} * F_{j,t}, \ F_{j,t} = \{MKT\_RF, SMB, HML, RMW, CMA\}$$

The Random Forest model is given by:

$$\widehat{R}_{i,t} = f(F_{1,t}, F_{2,t}, ..., F_{5,t}) + \epsilon_{i,t},$$

where $f(\cdot)$ is a nonlinear function learned from the data that captures complex relationships among factors.

Portfolios were constructed using predicted returns. Where the expected returns and the variance are defined as:

$$E(R_p) = \sum_{i=1}^{n} w_i * \widehat{R}_i$$

$$Var(R_p) = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i * w_j * cov(R_i, R_j)$$

Here, w_i represents the weight of stock *i* in the portfolio; $\widehat{R}_i$ represents predicted return of stock *i;* and $cov(R_i, R_j)$ represents the covariance of stocks *i* and *j*

The Sharpe ratio is given by:

$$S = \frac{E(R_p) - R_f}{\sqrt{Var(R_p)}}$$

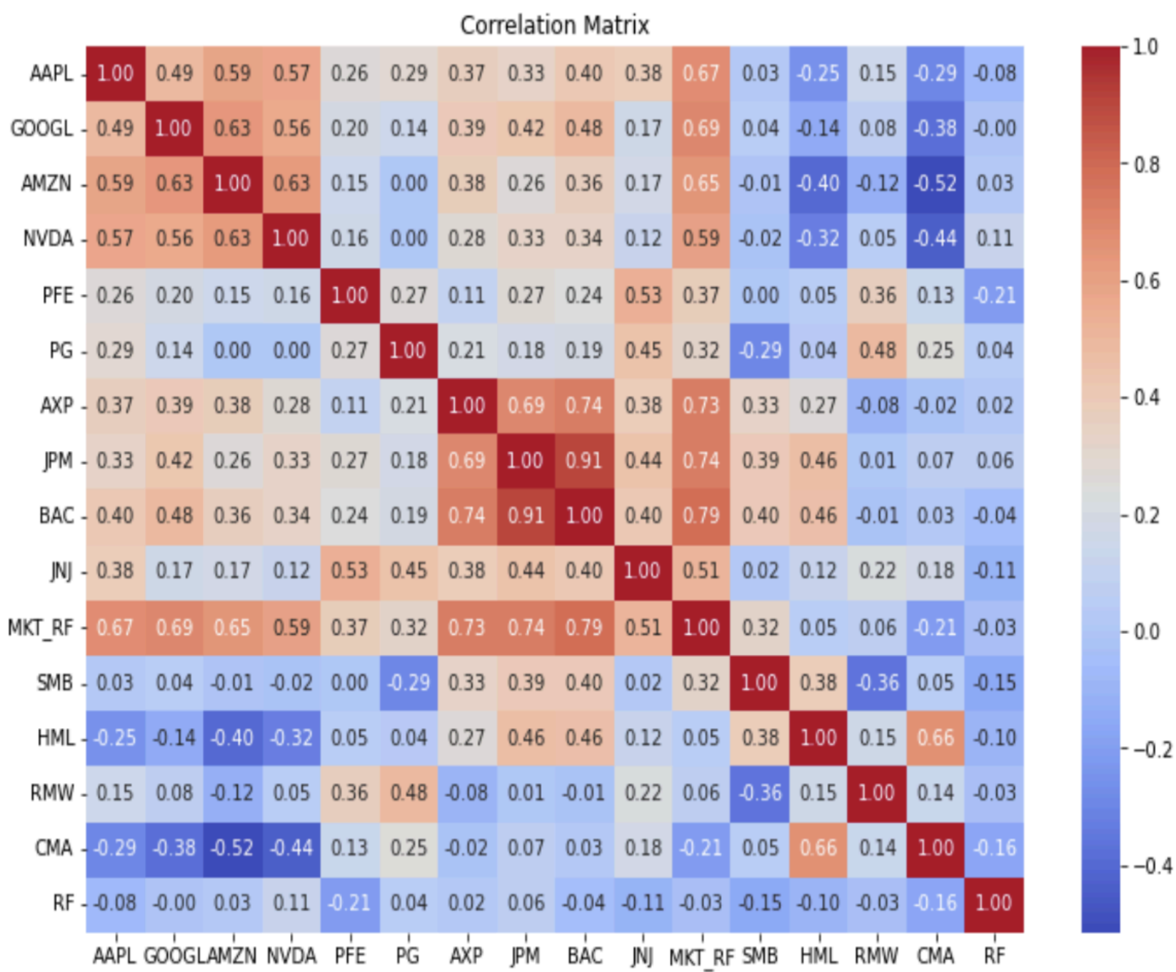
Summary Statistics

Variable	Obs	Mean	Std.Dev.	Min	Max
AAPL	95	0.02672	0.08006	-0.19108	0.20088
GOOGL	95	0.01865	0.06807	-0.19110	0.15975
AMZN	95	0.02118	0.08644	-0.25345	0.25171
NVDA	95	0.05233	0.13774	-0.36808	0.34775
PFE	95	0.00514	0.06829	-0.14859	0.22523
PG	95	0.01150	0.04682	-0.09325	0.12467
AXP	95	0.01952	0.07287	-0.18577	0.28705
JPM	95	0.01584	0.06972	-0.18215	0.20216
BAC	95	0.01308	0.08104	-0.20823	0.18693
JNJ	95	0.00753	0.04713	-0.12188	0.13939
MKT_RF	95	1.12747	4.81762	-13.39000	13.65000
SMB	95	-0.16505	3.13184	-8.24000	8.28000
HML	95	-0.18095	4.11030	-13.88000	12.80000
RMW	95	0.48663	2.19397	-4.79000	7.27000
CMA	95	-0.04526	2.58714	-7.20000	7.74000
RF	95	0.16274	0.15691	0.00000	0.48000

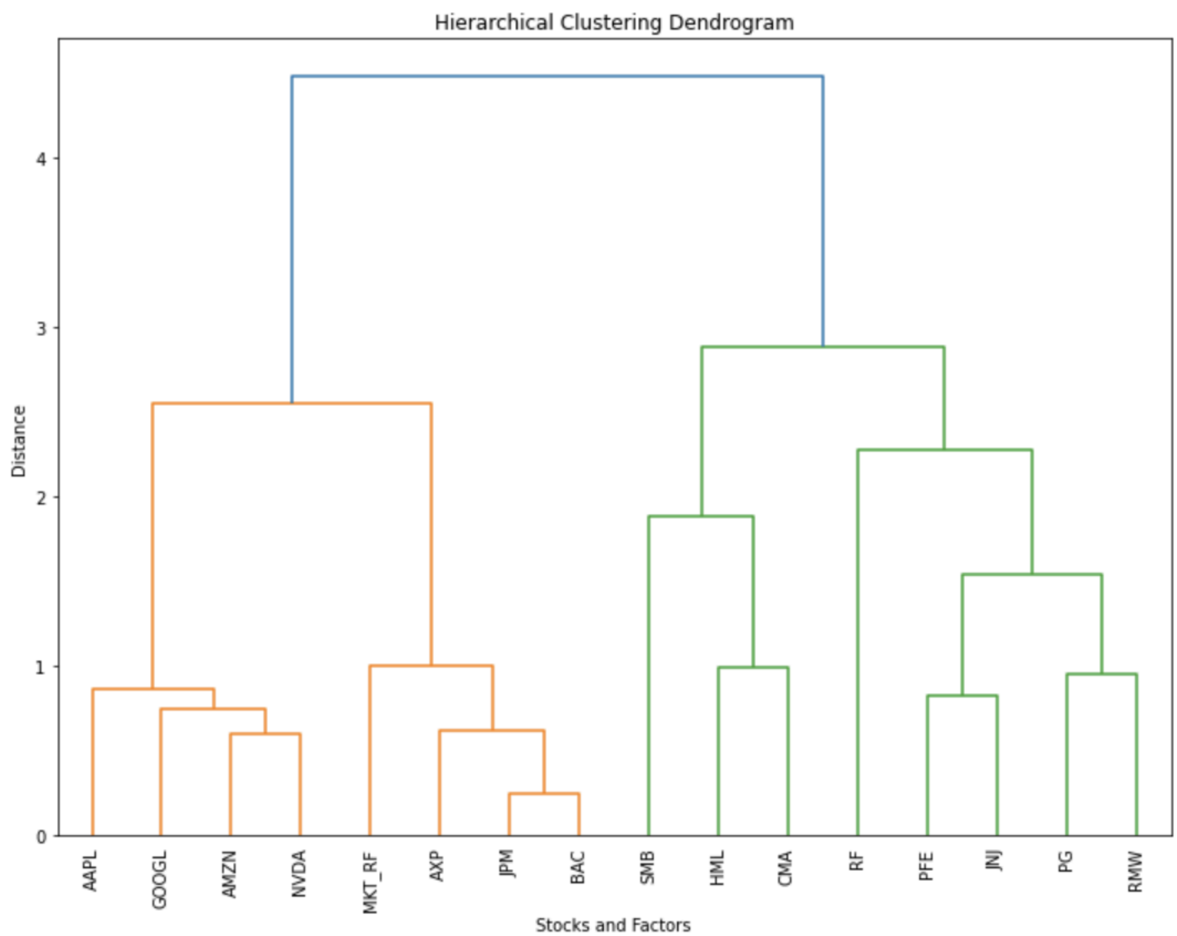
***Units of Measurement:**

All variables are measured in percentage (%)

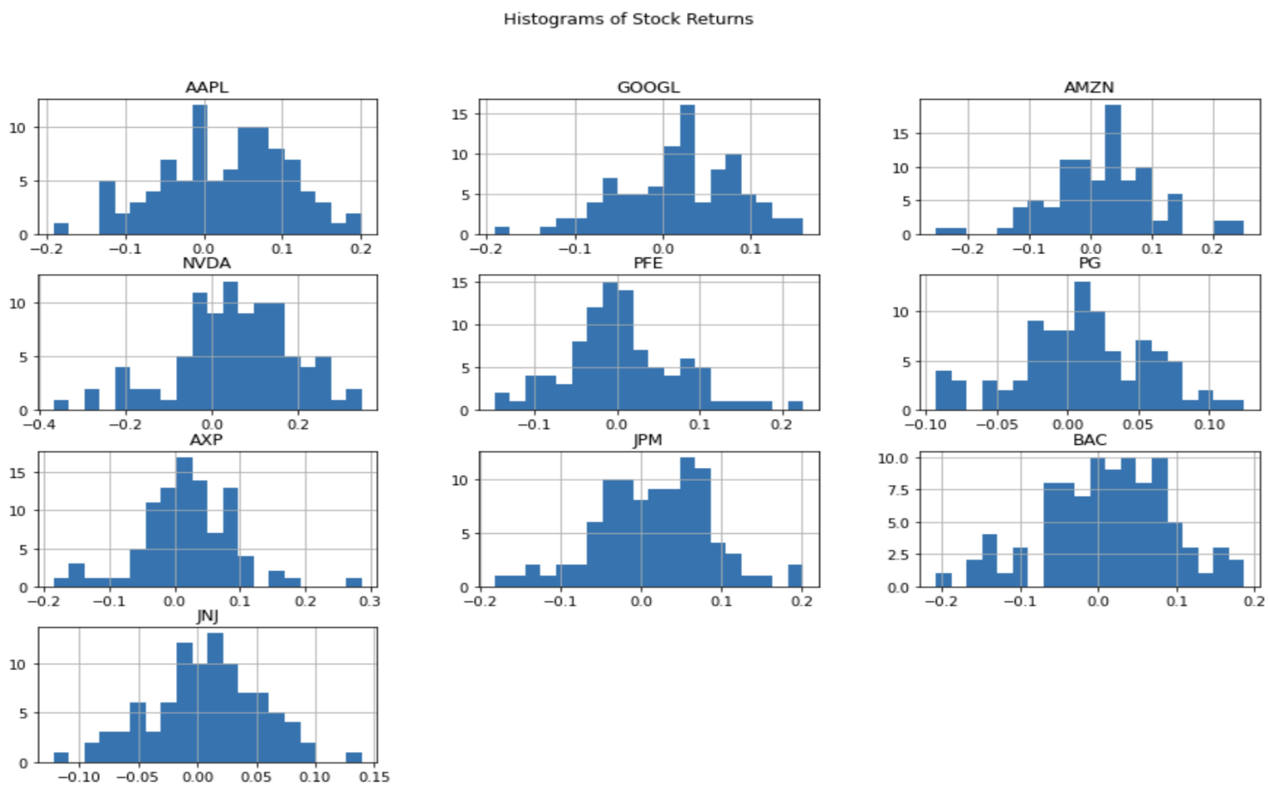
From the table, the mean monthly returns for the stocks range from 0.0051 (Pfizer) to 0.0523 (NVIDIA), which shows the varying levels of average performance during the observed period. Standard deviations reveal differences in volatility, with NVIDIA (0.1377) showing the highest variability and Procter & Gamble (0.0468) the lowest. Higher volatility indicates greater uncertainty in returns. Extreme values in returns, such as the maximum return of 0.3477 (NVIDIA) and the minimum return of -0.3681 (NVIDIA), highlight significant price swings for some companies during the analysis period. MKT_RF (market excess return) shows a mean of 1.1275 and a relatively high standard deviation of 4.8176, indicating fluctuations in the broader market's performance. SMB (size factor) and HML (value factor) have negative means, suggesting the underperformance of small-cap and value stocks relative to their counterparts during the observed period.



The correlation Matrix shows that stock returns reveal positive correlations, particularly among companies in similar industries or sectors. For example, JPMorgan Chase (JPM) and Bank of America (BAC) have a strong positive correlation of 0.91, reflecting their similar market dynamics as financial institutions. The market excess return (MKT_RF) positively correlates with most stock returns, emphasizing the influence of broader market movements on individual stock performance. In line with theoretical assumptions in factor investing, factor correlations show significant interdependence, such as the positive relationship between HML (value) and RMW (profitability).



The hierarchical clustering dendrogram shows the correlation structure among the stocks and factors in the dataset. Stocks and factors are grouped based on their similarity, with shorter branch lengths indicating higher correlations. The dendrogram indicates distinct clusters, such as AAPL, GOOGL, AMZN, and NVDA, which are closely linked due to their shared technology-driven market dynamics. Similarly, financial institutions like JPM and BAC cluster together, reflecting their correlated performance as they both are financial stocks. On the factor side, SMB and HML have close relationships, aligning with their theoretical interdependence in explaining value and size effects in stock returns. The broader separation between stocks and factors implies their distinct but interconnected roles in portfolio construction. This clustering not only highlights the underlying relationships but also underscores the importance of diversification and factor selection in portfolio optimization.



The histograms show the distribution of monthly returns for each stock. Returns for most stocks appear to follow a roughly symmetric distribution, with slight skewness in some cases. NVIDIA has a slightly right-skewed distribution due to its occasional high positive returns. Procter & Gamble (PG) and Johnson & Johnson (JNJ) display narrower spreads, consistent with their lower volatility, characteristic of stable, defensive stocks.

Training Models

The study evaluates three models for predicting stock returns: Linear Regression, Random Forest, and an Optimized Random Forest. These models were applied to predict monthly returns for ten selected stocks. The performance of the models is assessed using Mean

Squared Error (MSE) and R^2, the coefficient of determination, which reflect prediction error and explained variance, respectively.

Stocks/Models/Metrics	Linear Regression		Random Forest		Optimized Random Forest	
	MSE	R-Squared	MSE	R-Squared	MSE	R-Squared
AAPL	0.0031	0.6159	0.0036	0.5591	0.0034	0.5789
GOOGL	0.0018	0.5866	0.0019	0.5533	0.0020	0.5278
AMZN	0.0027	0.5930	0.0049	0.2604	0.0044	0.3370
NVDA	0.0135	-0.2405	0.0168	-0.5466	0.0145	-0.3332
PFE	0.0025	-0.5303	0.0029	-0.7819	0.0025	-0.5844
PG	0.0007	0.5570	0.0008	0.5018	0.0008	0.4817
AXP	0.0026	0.6133	0.0022	0.6752	0.0024	0.6511
JPM	0.0027	0.4745	0.0014	0.7219	0.0013	0.7436
BAC	0.0028	0.6635	0.0021	0.7526	0.0018	0.7868
JNJ	0.0014	0.0849	0.0014	0.0460	0.0013	0.1273

Linear Regression assumes a linear relationship between the factors and the returns. The result shows that the best performance is BAC, with MSE = 0.0028 and R^2 = 0.6635, and AXP with MSE = 0.0026 and R^2 = 0.6133. The high R^2 value indicates robust predictive accuracy for these stocks. The poorest performance stocks include NVDA, with MSE = 0.0135 and R^2 = -0.2405 and PFE with MSE = 0.0025 and R^2 = -0.5303, which shows negative R^2 values, suggesting the model struggled to capture their return patterns. The results suggest that Linear regression works well for stocks with relatively stable return patterns but underperforms for those who have non-linear relationships or high volatility.

Random Forest introduces non-linear interactions and feature importance into the analysis, and is better when capturing market dynamics. The result shows that the best performance is BAC, with MSE = 0.0021 and R^2 = 0.7526, and JPM with MSE = 0.0014 and R^2 = 0.7219. The high R^2 value indicates that Random Forest significantly outperforms linear regression. The poorest performance stocks include NVDA, with MSE = 0.0168 and R^2 = -0.5466 and PFE with MSE = 0.0029 and R^2 = -0.7819, which shows even worse R^2 values, possibly due to their high volatility or complex return dynamics. The Random Forest model demonstrates its strength in predicting stock returns, particularly for stocks with moderate to low volatility. However, stocks with extreme price movements remain a challenge. Therefore, an optimization for the Random Forest model is performed.

The Optimized Random Forest model utilized hyperparameter tuning with the primary objectives of minimizing mean squared error (MSE), enhancing R^2, and reducing overfitting to improve predictive performance across the portfolio. The result shows that the best performance is BAC, with MSE = 0.0018 and R^2 = 0.7868, and JPM with MSE = 0.0013 and R^2 = 0.7436. Both optimum MSE and R^2 improved compared to Linear regression and the original Random Forest. Also, stocks like AMZN, with MSE = 0.0044 and R^2 = 0.3370 and JNJ, with MSE = 0.0013 and R^2 = 0.1273 showed improvements over the untuned Random Forest, despite being marginal in some cases. The poorest performance stocks continue to be NVDA, with MSE = 0.0145 and R^2 = -0.3332 and PFE with MSE = 0.0025 and R^2 = -0.5844. The Hyperparameter tuned Random Forest model improved in predictive ability, particularly for stocks with intermediate predictability, but still remains less effective for stocks with extreme volatility or noise.

Portfolio Construction

Portfolios were constructed using the Markowitz Mean-Variance Optimization framework, applying the predicted returns from the linear regression and random forest as inputs. This process aims to find the optimal allocation of assets that maximizes the Sharpe Ratio.

Linear Regression Optimal Portfolio Weights:

AAPL: 0.0678

GOOGL: 0.0000

AMZN: 0.0000

NVDA: 0.2841

PFE: 0.0000

PG: 0.4661

AXP: 0.1820

JPM: 0.0000

BAC: 0.0000

JNJ: 0.0000

Expected Portfolio Return: 0.0256

Portfolio Volatility: 0.0547

Sharpe Ratio: 0.4383

The model chooses PG to occupy the most weight, 46.61%. This high weight suggests that PG provides a strong risk-adjusted return in the optimization process. NVDA has a weight of 28.41%, this allocation indicates that the linear regression model predicts high returns for NVDA relative to its risk. It may also have a diversification effect, reducing overall portfolio risk. AXP has 18.2%, which is another substantial contributor to the portfolio, likely reflecting its relatively attractive expected return and risk profile. AAPL only have 6.78%, implying moderate predictive returns or a higher correlation with other stocks, limiting diversification benefits. Other stocks are not included in the portfolio, which could be due to their predicted returns being insufficient relative to their risk, or they have high correlations with included stocks, reducing their diversification contribution.

This portfolio has an expected monthly return of 2.56%. Its volatility is 5.74%, which is relatively low. While the return is moderate, it may reflect a balance between achievable returns and minimized risk. The Sharpe ratio of 0.4348 infers a moderate reward for each unit of risk taken, suggesting this linear regression portfolio performs reasonably well but has room for improvement.

The portfolio is highly concentrated in PG (46.61%), NVDA (28.41%), and AXP (18.20%). While this maximizes returns for the given volatility, it may leave the portfolio exposed to idiosyncratic risks from these stocks. And the exclusion of GOOGL, AMZN, PFE, and other stocks implies that the linear regression model's limitations in incorporating complex relationships or dynamic market conditions. Overall, the results indicate that the linear regression model provides a moderately efficient portfolio, but its simplicity and reliance on linear relationships limit its ability to capture the complexities of financial markets.

Random Forest Optimal Portfolio Weights:

Optimal Portfolio Weights:

AAPL: 0.2546

GOOGL: 0.0000

AMZN: 0.0351

NVDA: 0.2389

PFE: 0.0000

PG: 0.0209

AXP: 0.4504

JPM: 0.0000

BAC: 0.0000

JNJ: 0.0000

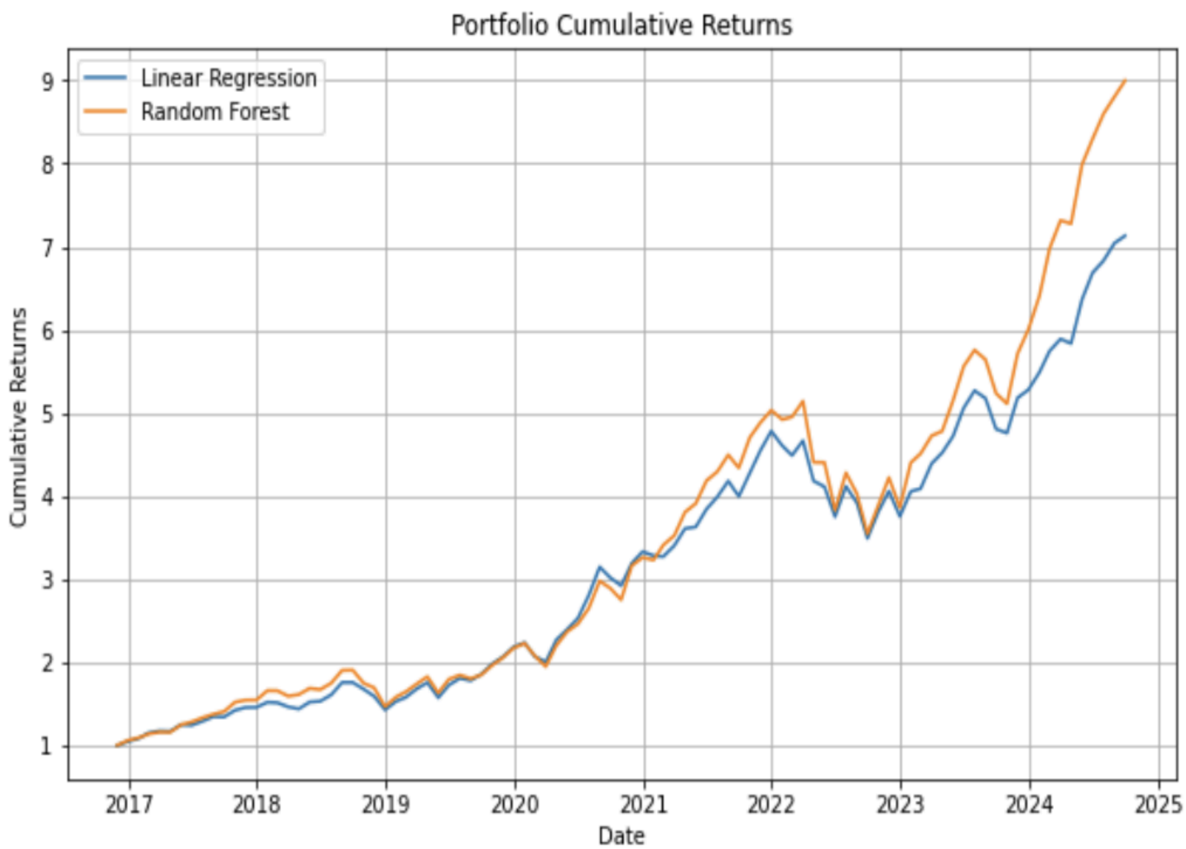
Expected Portfolio Return: 0.0358

Portfolio Volatility: 0.0689

Sharpe Ratio: 0.4966

This Random Forest model allocates AXP with 45.04%, AAPL with 25.46%, NVDA 23.89%, AMZN 3.51%, and PG 2.09%, with all other stocks 0%. The monthly expected return is 3.58%, which exceeds the expected return from the Linear Regression portfolio (2.56%). This indicates the Random Forest model's ability to predict higher returns by leveraging non-linear patterns in the data. However, the volatility is 6.89%, which is higher than that of the Linear Regression portfolio (5.47%), reflecting the larger allocations to high-risk stocks like AAPL and NVDA. The sharpe ratio is 0.4966, which also improved compared to the Linear Regression model, indicating better risk-adjusted performance. In spite of the increased volatility and concentrated weights, the Random Forest model demonstrates its strength in improving portfolio outcomes compared to Linear Regression by identifying non-linear relationships and dynamic factor impacts. In spite of the increased volatility and concentrated weights, it achieves superior returns and risk-adjusted performance compared to traditional methods, highlighting the advantages of ML models in factor investing.

Backtesting Results



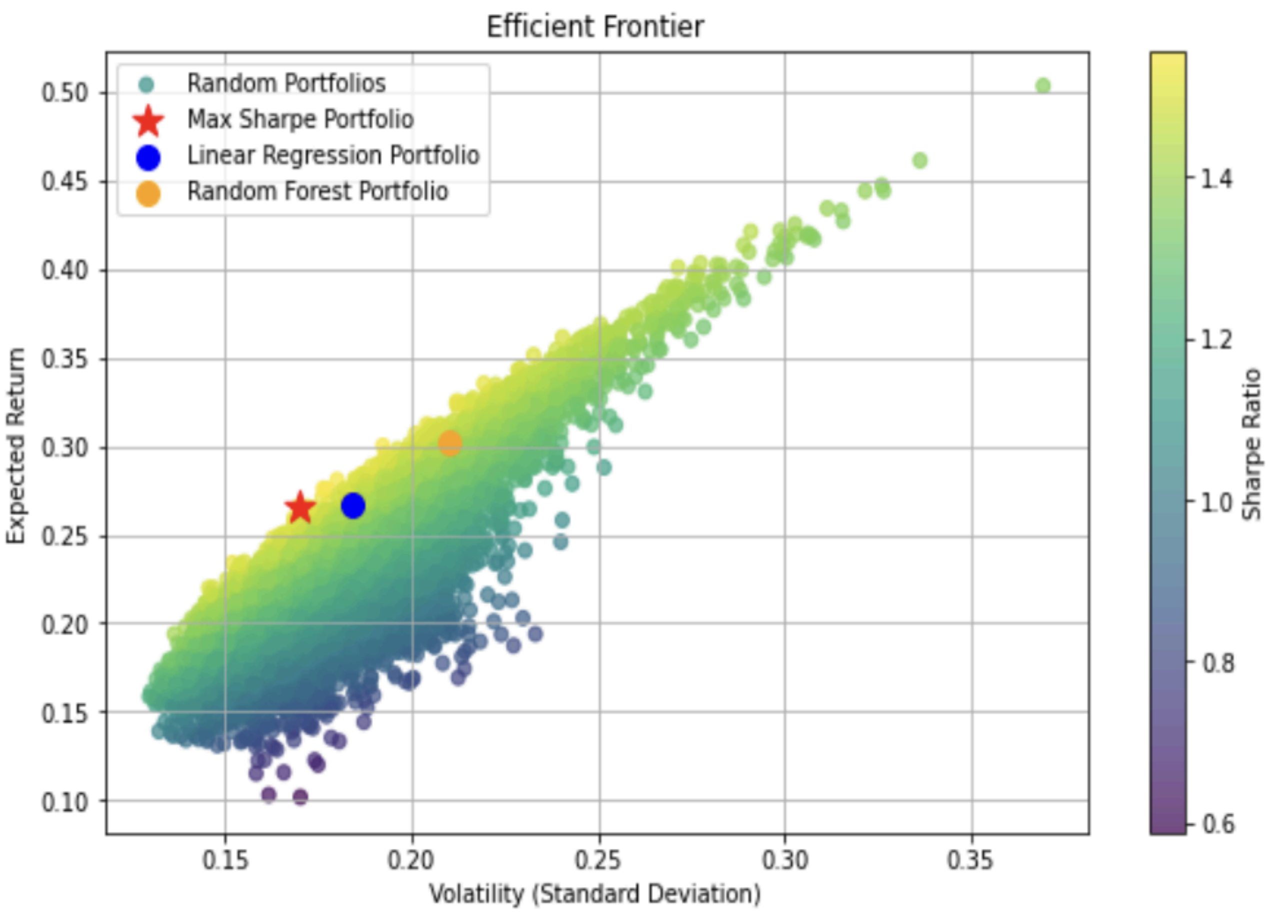
	Linear Regression	Random Forest
Annualized Return	26.75%	30.28%
Annualized Volatility	18.39%	21.02%
Sharpe Ratio	1.44550	1.43250
Max Drawdown	-26.98%	-31.04%

The Linear Regression portfolio generates a robust annualized return of 26.75%, possibly mainly driven by its significant allocation to lower-risk, stable stocks such as PG and AXP. Its annualized volatility is 18.39%, the portfolio reflects a conservative risk allocation, avoiding high-risk stocks and maintaining stability. Its Sharpe Ratio of 1.4455 shows a strong risk-adjusted performance, balancing returns and volatility. The portfolio's maximum drawdown of -26.98% implied its resilience during market downturns as it experiences smaller losses from peak to trough compared to more aggressive portfolios. The Random Forest portfolio generated a higher annualized return of 30.28%, primarily due to its allocation to high-growth stocks like NVDA and AXP. However, its annualized volatility is higher too, which is 21.02%, driven by its concentration in riskier assets such as NVDA and AAPL. The Sharpe Ratio of 1.4325 remains commendable, reflecting the portfolio's strong risk-adjusted returns despite slightly lower efficiency than the Linear Regression portfolio. The maximum drawdown is -31.04%, illustrating the trade-off for pursuing higher returns, with the portfolio exposed to greater losses during downturns due to its higher risk exposure.

The Portfolio Cumulative Returns line plot shows that Random Forest portfolio outperforms the Linear Regression portfolio in terms of cumulative returns, particularly between 2021 to 2024, where the machine-learning-driven model capitalizes on market dynamics.

However, this outperformance comes with higher drawdowns, emphasizing the inherent trade-off between risk and return.

Results showed that the Linear Regression portfolio is ideal for investors prioritizing stable, risk-adjusted returns with less exposure to extreme volatility. Since its more resilient performance during downturns, as evidenced by a lower max drawdown. Random Forest portfolio is suited for risk-lover investors seeking higher absolute returns. It can capture market nuances through non-linear relationships, resulting in better growth over the backtesting period.



The Efficient Frontier is used to showcase the trade-off between risk (volatility) and return. The random portfolios are represented by scattered points, color-coded based on their Sharpe Ratios. The portfolio with the maximum Sharpe ratio was represented by a red star, showing the optimal risk-return trade-off across all portfolios.

The Linear Regression portfolio, represented by the blue dot, is positioned closer to the maximum Sharpe Ratio portfolio, reflecting its strong balance of return and risk. The Random Forest portfolio, represented by the orange dot, while having a higher expected return, also assumes higher volatility, indicating a slightly more aggressive investment approach.

This plot emphasizes the comparative insights from earlier metrics. It demonstrates how machine learning-driven strategies, like the Random Forest portfolio, can achieve competitive returns while adhering to varying risk preferences.

Dynamic Factor Weighting

Dynamic factor weighting introduces an adaptable approach to portfolio construction by using the dynamic relationships between factors and stock returns. Dynamic factor weights are computed as a function of time-varying factor values using factor exposures (betas) estimated through linear regression (Coqueret 2023). The equation:

$$\beta_{i,j,t} = \frac{\partial R_{i,t}}{\partial F_{j,t}},$$

models stock returns $R_{i,t}$ as a function of factor exposures $\beta_{i,j}$ and factor values $F_{j,t}$. These betas represent the sensitivity of each stock to the factors. When betas are estimated, the factor weights are mapped to stock weights as follows:

$$w_{i,t} = \frac{|\beta_{i,t}|}{\sum_{k=1}^{n} |\beta_{k,t}|}.$$

where the absolute values of the betas are normalized to ensure the stock weights sum to one for each time period *t*.

This study incorporates a monthly rebalancing approach, where the factor exposures are recalculated at the end of each month using rolling regression on the past 12 months of data. The updated betas are then used to compute new stock weights which allow the portfolio to dynamically adjust to changes in market conditions and factor relationships. By contrast, static weighting approaches use fixed weights over the entire investment horizon, potentially missing out on shifts in market dynamics.

Random forest is not integrated with dynamic factor weighting in this study because it does not produce interpretable coefficients like betas, which are essential for mapping factor weights to stock weights. Instead, the dynamic factor weighting emphasizes the interpretability and transparency provided by linear regression in constructing portfolios that respond more flexibly to evolving market conditions.

The process begins with the extraction of excess returns for the selected stocks and their regression on the Fama-French Five factors to derive beta coefficients. Beta coefficients quantify each stock's sensitivity to the respective factors. Combining factor values with the beta matrix enables the computation of dynamic stock weights that adjust over time. The portfolio returns are then calculated by weighting the stock returns with the dynamic stock weights. Finally, it will generate cumulative returns and performance metrics such as annualized return, volatility, Sharpe Ratio, and maximum drawdown.

Original Dynamic Factor Weighting Output:

Annualized Return: 31.78%

Annualized Volatility: 23.07%

Sharpe Ratio: 1.3705

Max Drawdown: -34.49%

The annualized return reached 31.78%, outperforming both Linear Regression (26.75%) and Random Forest (30.28%) portfolios. This can imply the efficacy of dynamic factor weighting in capturing time-sensitive market opportunities. The volatility is 23.07%, reflecting its adaptiveness, which allows for increased exposure to high-risk, high-return opportunities when market conditions are favorable. The Sharpe Ratio is 1.3705, which is slightly lower than the Sharpe Ratio of the Linear Regression portfolio (1.4455). It can still demonstrate that dynamic factor weighting balances the trade-off between risk and return effectively. The max drawdown of -34.49% is higher than both Linear Regression (-26.98%) and Random Forest (-31.04%) portfolios which indicates that while dynamic factor weighting enhances returns, it also aligns with increased vulnerability during market downturns.

The portfolio can benefit from new trends and overcome the drawbacks of static models by dynamically adjusting factor exposures. However, the associated increase in volatility and drawdown shows the need for cautious implementation and robust risk management practices.

Optimized Dynamic Factor Weighting Output:

Annualized Return: 21.46%

Annualized Volatility: 17.05%

Sharpe Ratio: 1.2588

Max Drawdown: -22.81%

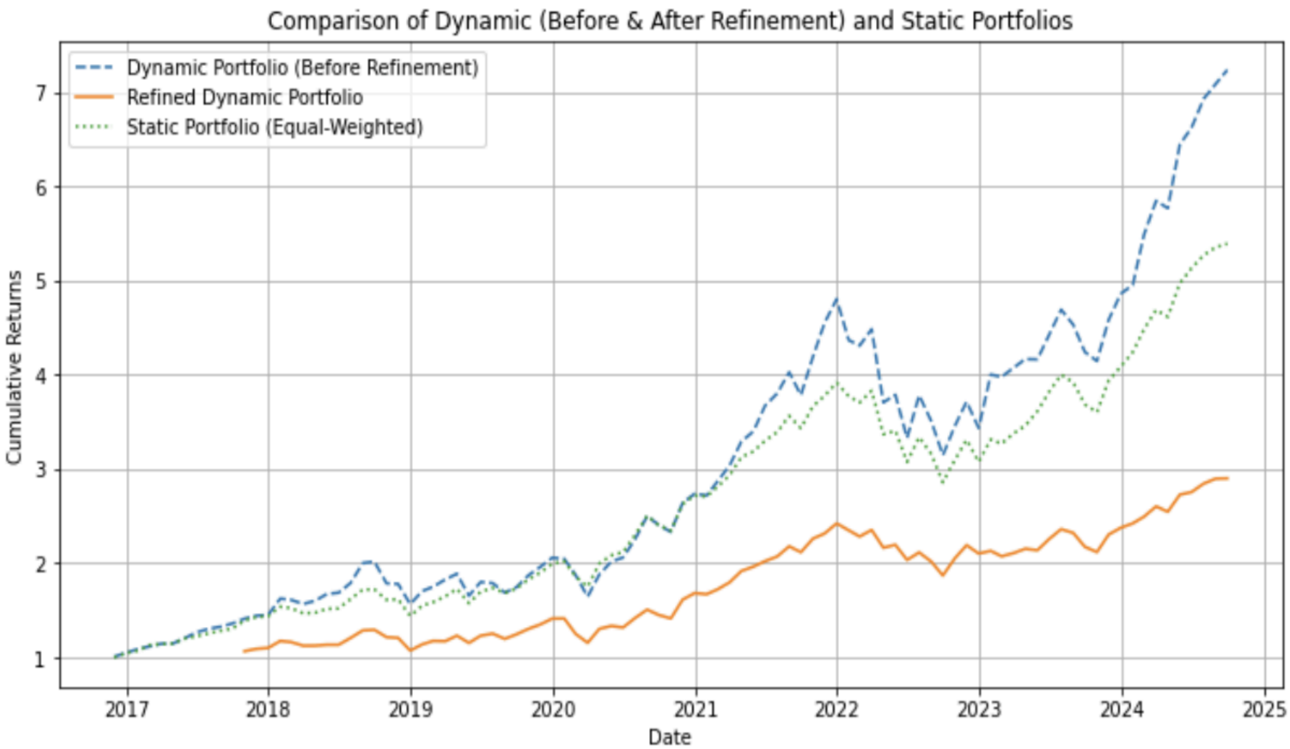
The optimized dynamic factor weighting builds upon the original dynamic factor weighting by incorporating refinements to the beta matrix and stock weights. These improvements ensure a more precise correlation between stock returns and factor exposures, which in turn improves portfolio construction and performance. Specifically, the process maps time-varying factor weights to stock weights using the beta matrix and ensures that the resulting weights are normalized to maintain a well-diversified portfolio. These refined stock weights are used to recalculate the portfolio returns and recalculate the key performance metrics.

The annualized return of 21.46%, shows a strong annualized return, but is lower than nearly 10% of the original dynamic portfolio. This reduction implies a more balanced and conservative allocation aimed at reducing excessive risk. An annualized volatility of 17.05% is a noticeable decrease compared to the original dynamic portfolio with volatility of 23.07%, indicating improved risk control. This is accomplished by limiting exposure to extreme factors and ensuring that stock weights are distributed more smoothly. A Sharpe Ratio of 1.2588 remains commendable, the slight decrease from the original dynamic portfolio may be due to the trade-off between optimizing returns and controlling volatility. It’s worth noting that the maximum drawdown, -22.81%, is significantly reduced compared to the original dynamic portfolio, -34.49%, underscoring the refined portfolio's advanced resilience during bad market conditions.

The optimized dynamic factor weighting method demonstrates a more balanced approach to portfolio management. Even though the returns are lower than the original dynamic portfolio, the decrease in volatility and drawdown shows an improvement in risk management. This optimization would underscores the potential of ML to refine traditional factor investing strategies, allowing for better adaptability and alignment with evolving market conditions.

Dynamic vs. Static Factor Weighting

Portfolio Type	Annualized Return	Volatility	Sharpe Ratio	Max Drawdown
Original Dynamic Portfolio	31.78%	23.07%	1.37750	-34.49%
Optimized Dynamic Portfolio	21.46%	17.05%	1.25880	-22.81%
Static Portfolio	25.56%	17.42%	1.46760	-26.97%



The original dynamic portfolio succeeds in generating high returns but has the highest volatility and drawdown risks. The optimized dynamic portfolio, while offering slightly lower returns, delivers a more stable and resilient performance. The static portfolio provides a consistent balance between return and risk, achieving the highest Sharpe ratio. The static portfolio's equal weighting serves as a baseline, but comparing it to the Linear Regression, Random Forest, and Optimized Random Forest portfolios reveals that these models offer diverse trade-offs in terms of return, volatility, and risk-adjusted performance. Random Forest achieving the highest return and Linear Regression providing more stable risk management.

Overall from the plot, the dynamic portfolio has the best performance, with the highest cumulative returns, and the optimized dynamic portfolio has the lowest cumulative return, but it has the lowest volatility and max drawdown. The results support the potential of dynamic factor weighting to enhance portfolio outcomes and the effectiveness of optimization in improving stability and risk management.

Results

The conventional linear regression method was greatly outperformed by machine learning models, especially the Random Forest algorithm, in terms of predictive performance. While linear regression exhibited reasonable predictive capabilities, as evidenced by R^2 values above 0.5 for stocks like AAPL, AMZN, and BAC, it struggled with non-linear relationships and complex interactions in financial data. On the contrary, the Random Forest model captured these complexities more effectively, achieving higher R^2 values for most stocks, such as JPM (0.7219) and BAC (0.7526). This improvement underscores the ability of ML models to uncover more complex and dynamic patterns in historical data, thereby enhancing predictive accuracy.

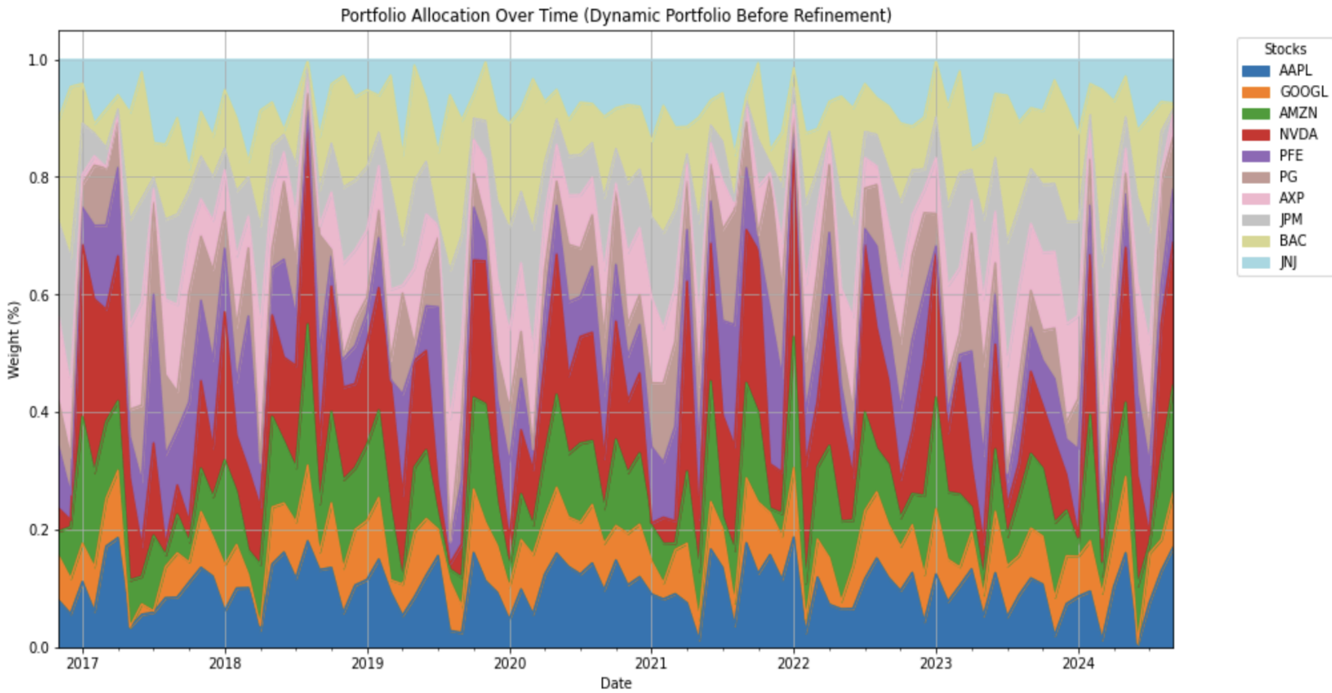
The optimized Random Forest further refined these predictions, though with marginal gains, suggesting that hyperparameter tuning contributes to performance, but has diminishing returns when compared to the original Random Forest model. These results collectively implied that ML-driven strategies provide superior predictive power by leveraging their capacity to model non-linear relationships and adapt to data complexities.

Machine learning models not only improved predictive accuracy but also enhanced portfolio construction and optimization outcomes. Using Markowitz's Mean-Variance Optimization framework, portfolios constructed based on ML-predicted returns exhibited superior risk-adjusted performance metrics (Markowitz 1952). The Linear Regression Portfolio achieved a Sharpe Ratio of 0.4383, with an annualized return of 26.75% and volatility of 18.39%. And the Random Forest Portfolio improved these metrics with a Sharpe Ratio of 0.4966, an annualized return of 30.28%, and volatility of 21.02%. The superior performance of the Random Forest Portfolio highlights how well ML-based predictions work when used in portfolio optimization. By accurately predicting expected returns, the model allocated higher weights to high-performing stocks like NVDA and AXP, resulting in higher returns. The improved Sharpe Ratio further demonstrates the portfolio's efficient trade-off between risk and return.

Dynamic factor weighting's ability to adjust to shifting market conditions and improve portfolio performance was demonstrated by this study. The original dynamic portfolio, constructed using factor-driven stock weights, achieved an annualized return of 31.78%, a Sharpe Ratio of 1.3775, and a max drawdown of -34.49%. However, the high volatility of 23.07% highlighted its sensitivity to market fluctuations.

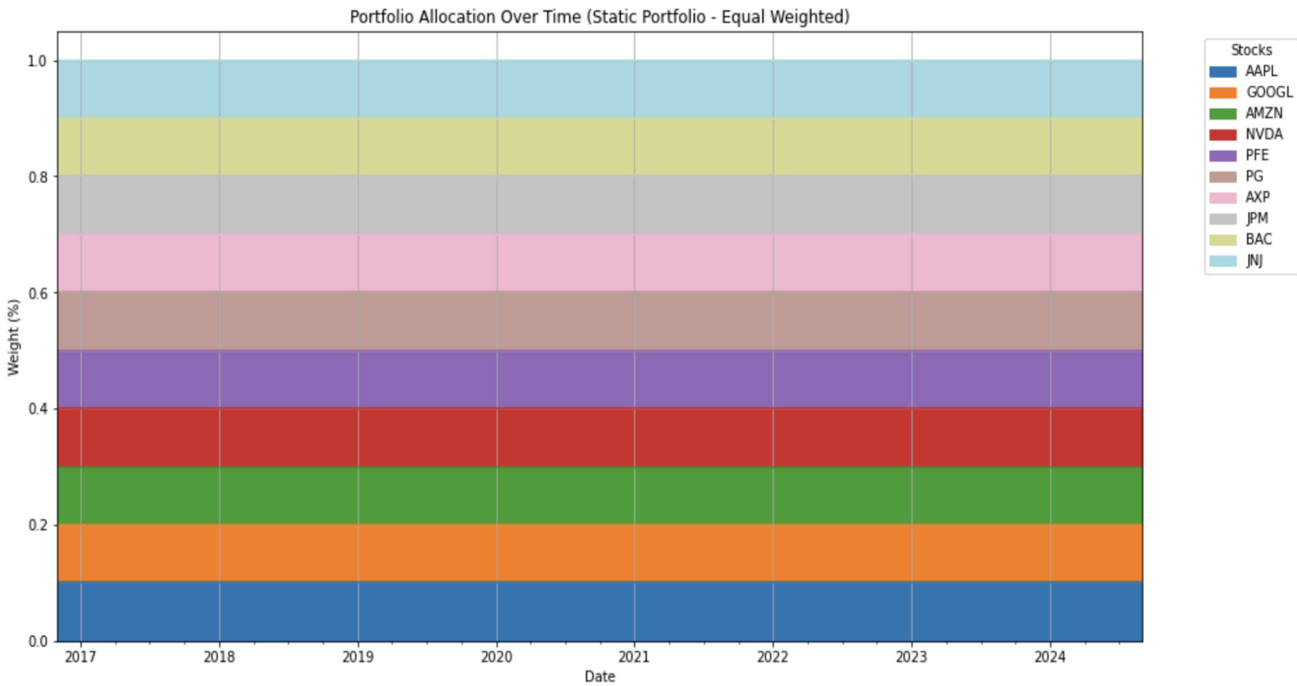
Then the optimized dynamic portfolio was created by aligning beta coefficients and imposing constraints on factor weights, which demonstrated improved stability and risk management with and annualized return of 21.46%, volatility of 17.05%, Sharpe Ratio of 1.2588 and max drawdown of -22.81%.

The optimized dynamic portfolio smoothed excessive allocation shifts while maintaining adaptability to time-varying factor exposures. And the comparison with the static portfolio indicates that dynamic factor weighting better captured evolving market dynamics, achieving superior performance metrics while maintaining greater flexibility.



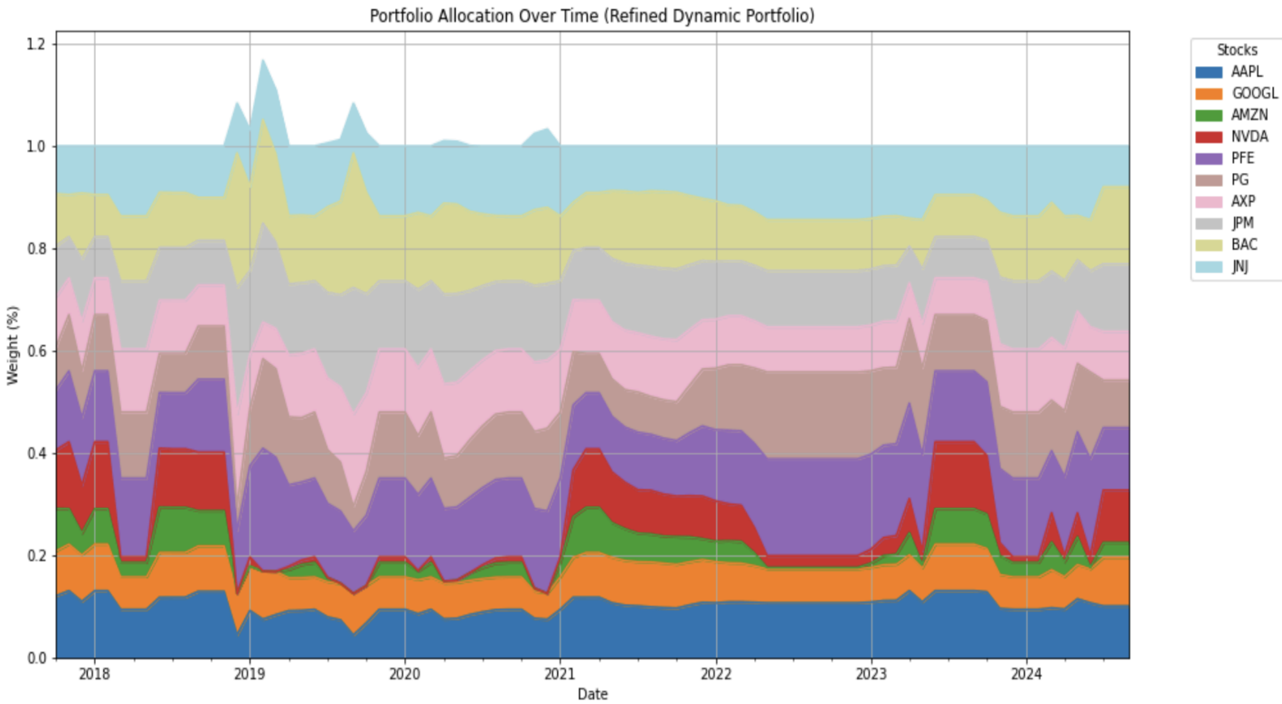
***Risk: 23.07% Annualized Return:31.78%**

This graph shows that the original dynamic portfolio composition fluctuates significantly across the timeline, reflecting the influence of factor exposures and dynamic market conditions. It's frequently rebalancing, as allocations to specific stocks, such as NVDA and AXP, vary widely depending on changing factor dynamics. Stocks like PG and AAPL often maintain stable, yet fluctuating, allocations, indicating their consistent performance relative to factor weights. Periods of heightened volatility in market conditions, such as during economic shocks in 2020, result in sharp changes in allocations.



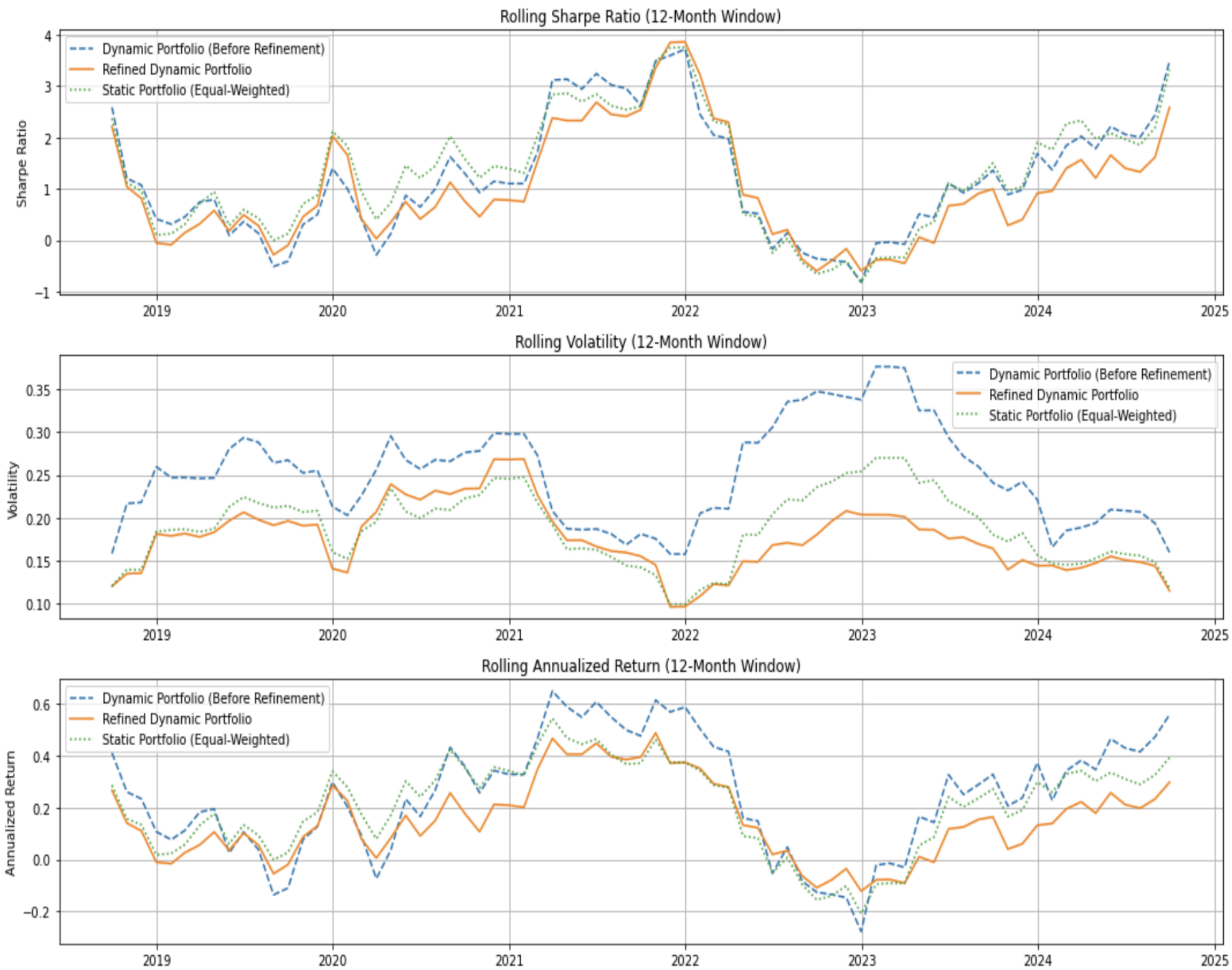
***Risk: 17.42% Annualized Return: 25.56%**

This graph shows the static portfolio. Unlike the dynamic portfolios, the allocations remain constant across time, with each stock holding a weight of 10%.



***Risk:17.05% Annualized Return: 21.46%**

The third graph represents the optimized dynamic portfolio. The allocation is smoother and less erratic compared to the original dynamic portfolio, this shows an improved stability while retaining its adaptiveness. Stocks such as AXP and NVDA show more balanced allocations, indicating their importance in capturing risk premia. Also, the optimized model appears to reduce extreme shifts in allocations during volatile periods, which demonstrates its ability to mitigate risks with dynamic responsiveness.



The rolling performance metrics shows that for Sharpe Ratio, the original dynamic portfolio and static portfolio have similar performances, and the optimized dynamic portfolio has the poorest performance. But when it comes to volatility, the optimized dynamic portfolio has the best performance, it consistently maintains lower volatility compared to the other two, demonstrating enhanced risk-adjusted performance. Lastly, the original dynamic portfolio outperforms the other two portfolios in annualized returns.

Discussion

This study shows the potential of machine learning to enhance traditional multi-factor investing by improving predictive accuracy and enabling dynamic portfolio adjustments. The Random Forest portfolios outperformed linear regression portfolios, showing superior returns and effective risk management through dynamic factor weighting. However, the trade-offs include increased volatility and drawdowns, emphasizing the importance of optimization and regular rebalancing. Although machine learning techniques require a lot of data and processing power, their adaptability in responding to evolving market situations makes them useful instruments for advancing portfolio management. Future research is needed to explore integrating alternative data or optimizing dynamic weighting techniques to enhance scalability and robustness.

References

Changshuai Cao, Yuchan Jin, Huang Huang, *Research on the Construction of Enterprise Financial Shared Service Center Based on Cloud Computing,* E3S Web Conf. 235 01041 (2021), DOI: 10.1051/e3sconf/202123501041

Coqueret, G., & Guida, T. (2023). *Machine Learning for Factor Investing: Python version. CRC Press.*

Eugene F. Fama, Kenneth R. French*,* Common risk factors in the returns on stocks and bonds*, Journal of Financial Economics,*Volume 33, Issue 1, 1993, Pages 3-56

Eugene F. Fama, Kenneth R. French, A five-factor asset pricing model, *Journal of Financial Economics*, Volume 116, Issue 1, 2015, Pages 1-22,

Markowitz, Harry. “Portfolio Selection.” *The Journal of Finance,* vol. 7, no. 1, 1952, pp. 77–91. JSTOR, https://doi.org/10.2307/2975974. Accessed 7 Dec. 2024.

Gao, J., Guo, H., & Xu, X. (2022). *Multifactor stock selection strategy based on machine learning: Evidence from China. Complexity*, *2022*(1). https://doi.org/10.1155/2022/7447229

Robbins, M. (2023). *Quantitative Asset Management Factor Investing and machine learning for institutional investing. McGraw Hill.*

Roy, S., & Jönsson, J. (2019). *Analysing multifactor investing & artificial neural network for modern stock market prediction (Dissertation).* Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:hj:diva-43875

Wei, Z., Dai, B., & Lin, D. (2022, October 22). *Factor investing with a deep multi-factor model.* https://arxiv.org/pdf/2210.12462