Forecasting S&P 500 Sector ETFs Returns: A Machine Learning Approach to Sector Rotation

João Paulino

Dissertation written under the supervision of Professor Dan Tran

Dissertation submitted in partial fulfilment of requirements for the MSc in Finance, at the Universidade Católica Portuguesa, 01-06-25.

Abstract

This thesis develops a machine learning-based dynamic sector rotation strategy, sector ETFs are used as a representation of the United States 11 sectors and of the benchmark S&P500. We use a variety of machine learning models, including LASSO, XGBoost, and Random Forest, to forecast sector returns by utilizing financial market data, market sentiment metrics, currencies and macroeconomic indicators. The top and bottom-performing sectors are chosen to create long, short, and long-short strategies for portfolio construction based on these forecasts. Results show limited predictive power overall, but the strategies built offer modest improvements over the benchmark in certain market conditions.

Title: Forecasting S&P 500 Sector ETFs Returns: A Machine Learning Approach to Sector Rotation

Author: João Carlos Marques Paulino

Keywords: Sector Rotations, Exchange-Traded Funds (ETFs), Feature Importance, Returns Prediction

Resumo

Esta tese desenvolve uma estratégia de rotação dinâmica de setores baseada na aprendizagem automática. Os ETFs setoriais são utilizados como representação dos 11 setores dos Estados Unidos e do benchmark S&P 500. Utilizamos uma variedade de modelos de aprendizagem automática, incluindo o LASSO, o XGBoost e o Random Forest, para prever os retornos setoriais, utilizando dados do mercado financeiro, métricas de sentimento do mercado, mercado cambial e indicadores macroeconómicos. Os setores com melhor e pior desempenho são selecionados para criar estratégias de compra, venda e compra-venda para a construção de carteiras com base nestas previsões. Os resultados mostram um poder preditivo global limitado, mas as estratégias construídas oferecem melhorias modestas em relação ao benchmark em determinadas condições de mercado.

Título: Previsão dos Retornos dos ETFs Setoriais do S&P 500: Uma Abordagem de Machine Learning Para a Rotação de Setores

Autor: João Carlos Marques Paulino

Palavras-Chave: Rotação de Setores, Exchange-Traded Funds (ETFs), Importância das Variavéis, Previsão de Retornos

Table of Contents:

List of Figures

List of Tables

1. Introduction

One of the most challenging and valuable tasks in asset management is forecasting correctly changes in financial markets. Among the strategies developed to address market uncertainty, sector rotation emerges as a notable approach to enhance returns and mitigate risk by diversifying exposure across economic sectors. The fundamental notion is straightforward yet impactful: various sectors exhibit divergent performance patterns, which are contingent on the prevailing macroeconomic and financial conditions. The challenge lies in identifying which sectors are likely to outperform in the near future. This task is traditionally guided by business cycle models or investor perception. Simultaneously, the financial industry is undergoing a transition toward data-driven decision-making, propelled by enhanced access to economic, market, and sentiment data, as well as advancements in machine learning methodologies. The utilization of these models has the capacity to reveal nonlinear patterns and interactions that may be imperceptible to conventional linear methods. However, the integration of machine learning into sector rotation remains relatively under-explored, in this dissertation we will try to combine two tasks, forecasting Sector ETFs returns to implement a sector rotation strategy.

By combining interpretable machine learning models with a variety of predictive features gathered from four different domains, market sentiment indices, currency fluctuations, macroeconomic indicators, and derived financial metrics this dissertation tries to achieve a strong sector rotation strategy that can adjust to shifting market conditions and provide better risk-adjusted returns than a passive benchmark. In order to create a long/short rotation strategy, we use these inputs to train machine learning models that forecast the monthly returns of U.S. sector ETFs. We then go long on the three sectors with the highest predicted returns and short on the three with the lowest. Our method makes use of models like Lasso, Random Forest, and XGBoost, which are popular in finance because of their ability to balance interpretability and predictive power, in contrast to previous research that mostly concentrates on individual stocks, our study shifts the focus to sector ETFs, which are more practical for portfolio allocation. So, the main questions that this dissertation tries to answer are:

1. Can interpretable machine learning models provide reliable monthly forecasts of sector ETF returns using a broad set of macroeconomic, sentiment, currency, and market features?

2. Does a sector rotation strategy based on these forecasts deliver a significant improvement in performance over a passive investment in the S&P 500 ETF (SPY)?

To conduct the research, we start with a literature review on the main subjects, sector rotation and the application of machine learning for the prediction of returns. Then the data used, their sources and their details are explained. In methodology we give details on how we did the time-series cross validation, the models specifications and how we constructed the portfolio strategy. Lastly, we present the results, evaluate them and answer our research questions followed by the conclusion with the summary key findings, the limitations of this dissertation and the future path that this work can take to improve even more.

2. Literature Review

Sector rotation strategies have been employed for a long time to optimize risk-adjusted returns and handle shifting economic environments. The fundamental notion is that various sectors respond distinctively to business cycles and macroeconomic shifts. Chong and Phillips (2015) claim that cyclical industries, such as technology and consumer discretionary, thrive during periods of economic expansion, but defensive industries, including utilities and healthcare, generally perform better during economic contractions. Conventional sector rotation methods rely on investor judgments or macroeconomic concepts, sometimes associated with lagging indications or subjective interpretations. Several investigations have tried to elucidate these trends. Stangl, Jacobsen, and Visaltanachoti (2009) employed economic data, including GDP growth, interest rates, and inflation, to correlate sector returns with phases of the business cycle. According to their findings, sector-based timing can beat passive benchmarks if it is in line with macro indicators.

In acordance to this last research, other studies suggest that macroeconomic indicators can offer meaningful signals when integrated thoughtfully. Phua, Chung, and Tjia (2010) proposed a sector rotation framework based on economic indicators, particularly industrial production growth and the term spread between long- and short-term interest rates. Their approach outperformed the S&P 500 over the 1990–2008 period, achieving a higher Sharpe ratio and smaller drawdowns during recessions. They identified that the most significant gains occurred during economic turning points, phases in which traditional sector timing models struggle a lot to capture those turning points signals. Yet, the authors also acknowledged that real-time implementation is constrained by data lags and market frictions, which reduced the net benefit by approximately 2% annually, this is an important point since macroeconomic data, in general, has some lag.

In a comparable way, Cohen and Polk (1998) suggested that time-varying sector risk premia represents investors' economic expectations, which static asset pricing models may not always account for. The application of sector rotation in practice is still up for debate, despite its theoretical appeal. The effectiveness of traditional rotation strategies, especially those based on static business-cycle timing, was criticized in recent work by Wang and Xu (2020). Only marginal outperformance was found in their study, which included simulations using perfect foresight of economic phases. Even that advantage vanished when transaction costs or small timing errors were taken into account. During early expansion periods, which are generally thought to favor cyclical sectors, the strategy actually performed worse than the overall market. The authors came to the conclusion that simpler approaches frequently produced better risk-adjusted returns, such as sticking to a diversified index or switching to cash during recessions. Their results challenged the notion that sector rotation generally adds value, particularly when it is implemented using very strict guidelines.

The conventional models, based on research, have two main limitations. First of all, they are usually slow to adjust to fast changes in the market. Second, they sometimes lack the capacity to capture complex or nonlinear relationships between asset returns and economic drivers. This has caused increasing interest in using machine learning techniques to sector rotation since they promise more flexible and data-driven policies.

Machine learning methods provide a significant advantage dealing with the limitations that previous research identified when dealing with this type of strategies. This models offer the ability to detect complex, nonlinear interactions among a wide set of features, enabling more adaptive responses to evolving economic conditions, in our case macroeconomic indicators, market-based signals, sentiment indices, and currency trends are going to be used to create a good forecasting base, the relationships between this different set of variables sometimes is not clear , so machine learning methods may help a lot to correctly predict the sector returns. This represents a departure from traditional approaches, which typically rely on a small number of lagging variables and fixed threshold rules.

The work of the authors Gu, Kelly, and Xiu (2020), plays an important role in the recent literature of machine learning for finance, their findings are very relevant for our research. They compared the performance of different models, including deep neural networks, random forest and gradient boosting against the perfomsnce of linear models in predicting stock returns. The results confirm that tree based models outperform the classic linear regressions in sample and out of sample, beacause of their capacity to capture these non-linear relationships, another relevant finding is the feature importance the most important predictors of future returns are variations on momentum, liquidity, and volatility, which we are going to confirm if this condition is met or not, also using this type of indicators. Bryzgalova, Pelger, & Zhu (2022) also reached the same conclusions regarding the tree models, specially random forests, In our research we will perform simpler methods like Lasso and tree-based models such as XGBoost, and Random Forest and verify if they can still offer significant improvements in financial forecasting and with that improve the domain of sector ETF rotation.

In the last few decades Exchange-Traded Funds (ETFs) have gained relevance due to the advantage that this type of asset provides. They are instruments with high liquidity that trade continuously on exchanges and at low transaction costs (Hegde and McDermott (2004) and Agapova, A. (2011)), it allows investors to take exposure to specific indices, asset classes, geographies, sectors etc., since they track their indices with considerable precision as found by Gastineau (2004) and Charupat and Miu (2013). This has expanded the possibilities for sector rotation strategies by enabling investors to change their exposure between sectors more quickly, easily, and affordably in response to shifting market conditions. Investors can strategically rebalance sector allocations with ETFs without having to deal with the difficulties of picking individual stocks or paying high transaction fees.

Taken together, the literature on machine learning and ETF implementation supports a hybrid approach that this dissertation analyzes: using interpretable machine learning models to predict monthly sector ETF returns based on macroeconomic, sentiment, currency, and market features, and running a dynamic sector rotation strategy using liquid, reasonably priced ETF instruments.

3. Data

These ETFs have shown themselves to be representative of the 11 different sectors in the US, and all of them have high liquidity and low trading costs, for the Communication Services and Real Estate sectors it was decided to choose both from Vanguard due to the history of the ETFs, the other alternatives lacked in that aspect.

Table 1: ETFs description

ETF Ticker	Sector	Issuer	Inception Year
VOX	Communication Services	Vanguard	2004
VNQ	Real Estate	Vanguard	2001
XLY	Consumer Discretionary	State Street	1998
XLP	Consumer Staples	State Street	1998
XLE	Energy	State Street	1998
XLF	Financials	State Street	1998
XLV	Healthcare	State Street	1998
XLI	Industrials	State Street	1998
XLB	Materials	State Street	1998
XLK	Technology	State Street	1998
XLU	Utilities	State Street	1998
SPY	S&P500	State Street	1998

To accomplish the goal of this thesis, we collected four different data blocks, each of which was chosen to capture a distinctive type signal that we consider pertinent to our analysis. These data blocks provide a strong and varied input set for our predictive models by covering a range of financial and economic aspects. Starting at the end of October 2004 and ending at the end of October 2024, the data collection period spans 20 years.

3.1. Derived Features

We computed some derived features from historical price and volume. As a proxy for the risk involved with each ETF, volatility was calculated as the rolling standard deviation of monthly returns over a 12-month period. In financial modeling and portfolio construction, volatility is an essential risk metric. Higher volatility industries carry more uncertainty even though they might have higher potential returns. Instead of just pursuing returns, the model can better balance risk-adjusted performance by incorporating volatility.

A number of momentum- and volume/price-based features were developed to supplement the volatility measure and improve the model's capacity to evaluate relative performance. The momentum rank (rank), which ranks sector ETFs based on their cumulative returns over the past 12 months, captures the relative strength of each one, momentum refers to the tendency of assets that have historically performed well to continue doing so in the near to medium term. The ETFs were ranked according to their 12-month cumulative returns rather than raw momentum values. This cross-sectional ranking supports the model in identifying relative winners and losers within each period and offers a consistent means of comparing momentum across all ETFs, irrespective of scale or volatility variations.

Two features linked to trading volume were also built. The relative change in an ETF's trading activity from one period to the next is measured by the percentage change in volume (pct_vol), which may indicate changes in investor focus or liquidity conditions. Longer-term trends in investor interest are captured by the 12-month percentage change in volume (pct_12m_vol), which may be a sign of structural changes in market positioning or sentiment. When combined, these characteristics enable the model to take into account both price dynamics and market behavior, enabling a more thorough assessment of the risk-return profile of each ETF.

3.2. Foreign Exchange (FX)

Exchange rate fluctuations play an important role in global trade, influencing the competitiveness of exports and imports, corporate earnings, and investment flows. Given the economic position in the global panorama of United States, the FX movements of its top trading partners can have significant implications for various sectors. Large movements in exchange rates can affect trade balances, input costs, and multinational corporations' profitability, making FX data a valuable variable for modeling sector performance. For this analysis, we have collected daily FX rates for the currencies of the United States major commercial partners:

- Canada (CAD/USD): As the US's largest trading partner, exchange rate movements impact energy, manufacturing, and agricultural trade flows.

- United Kingdom & Eurozone (GBP/USD, EUR/USD): FX fluctuations in these economies affect transatlantic commerce, financial services, and multinational operations between the US, UK, and EU.
- China & Japan (CNY/USD, JPY/USD): FX movements in these major Asian economies affect global supply chains, manufacturing costs, and trade relations with the US.

By incorporating FX data into our analysis, we aim to assess how currency fluctuations impact on the sector's performance, particularly the trade-sensitive industries. Sectors with significant global exposure may be more responsive to FX movements, making this a factor to consider in a dynamic sector allocation strategy.

Table 2: Currencies description

Variable Name	Definition	Frequency	Source	Economic Meaning
CAD	CAD/US Dollar FX Spot Rate	Daily	LSEG Workspace	Largest trading partner.
EUR	EURO/US Dollar FX Spot Rate	Daily	LSEG Workspace	Movements affect transatlantic commerce, financial services, and trade flows with the US and EU.
JPY	JPY/US Dollar FX Spot Rate	Daily	LSEG Workspace	Influences trade with US manufacturing sectors.
CNY	CNY/US Dollar FX Spot Rate	Daily	LSEG Workspace	

3.3. Sentiment Indicators

Investor sentiment plays a crucial role in financial markets, in asset prices beyond traditional fundamental and technical factors. Sentiment indicators capture market participants' expectations, risk perception, and overall confidence in the economy, making them valuable for predictive modeling, particularly in the short term. Since these indicators often incorporate forward-looking elements, they can serve as early signals of shifts in market dynamics.

Incorporating sentiment indicators into machine learning models may enhance their ability to forecast sector ETF performance by providing insights into risk appetite, economic confidence, and behavioral biases. For this analysis, we include three key sentiment indicators:

- VIX (CBOE Volatility Index): Often referred to as the "fear gauge," the VIX measures market expectations of volatility in the S&P 500 over the next 30 days. A rising VIX generally signals increased uncertainty and risk aversion, leading to defensive positioning in markets. Conversely, a declining VIX suggests a more stable, risk-on environment.
- EPU (Economic Policy Uncertainty Index): This index quantifies the degree of uncertainty related to government policy decisions, regulatory risks, and economic outlook. High levels of policy uncertainty can lead to market hesitancy, less investment and more volatility.
- Michigan Consumer Sentiment Index: This indicator measures consumer confidence in economic conditions, which directly impacts household spending and overall economic growth. A strong consumer sentiment often correlates with bullish market trends, while declining sentiment may signal economic slowdowns.

3.4. Macroeconomic Data

Besides the data already exposed, we bring in a set of macroeconomic variables with the purpose of getting a broader market perspective, the sectors react very differently from each other to changes in these variables.

Interest Rates and Yield Curve**:** were retrieved for a wide range of maturities (3-month to 20-year) the US Treasury Yields. It is expected that these can tell us about market expectations, in the short-term rates normally reflect decisions made by the Federal Reserve and long-term rates capture growth and inflation expectations. The shape of the yield curve (e.g., the 10-year minus 2-year spread) has historically been viewed as a signal for economic cycles, if this spread is negative investors expect in the short run the economy to slow down so they want higher compensation to hold short term bonds when compared to long term bonds. This section is important because sectors more heavily in debt feel the impact of these in different ways.

Housing Market Indicator**:** Regarding the retail sector we collected the Case- Shiller Home Price Index. With this indicator we hope to gather meaningful information about the trends of the sector, the price variation of houses reflect the actual health of real estate.

Energy and Commodity Prices: Since energy costs have a big impact on various sectors like consumer discretionary, we also take into account the Gas Spot Price and Price of Crude Oil (WTI). Unexpected increases in oil prices have the potential to slow economic growth by raising input costs for businesses and decreasing consumer disposable income, whereas declines in oil prices frequently offer a welcome relief that can increase consumer spending and business profitability. In addition to crude oil price was gathered prices of Refinitiv Spot Gold as a crucial marker of market sentiment and economic concern. Gold has long been thought of as a safe-haven investment. Investors often rush to gold during times of market turbulence or economic uncertainty, driving up the price. This dynamic can offer important insights into investors' risk gauge and help in signaling changes in risk perception throughout the market.

Table 3: Macroeconomic variables description

Variable Name	Data Source	Type	Definition	Frequency
3-Month Treasury Yield	FRED	Interest Rate	Yield on 3-month US Treasury securities	Monthly
2-Year Treasury Yield	FRED	Interest Rate	Yield on 2-year US Treasury securities	Monthly
10-Year Treasury Yield	FRED	Interest Rate	Yield on 10-year US Treasury securities	Monthly
20-Year Treasury Yield	FRED	Interest Rate	Yield on 20-year US Treasury securities	Monthly
Crude Oil Price (WTI)	EIA	Commodity Price	Price per barrel of West Texas Intermediate (WTI) crude oil	Daily
Healthcare Services Price Index	FRED	Price Index	Measures healthcare service price changes over time	Monthly

10Y-2Y Yield Spread	FRED	Yield Spread	Difference between the 10-year and 2-year Treasury yields	Monthly
Case-Shiller Home Price Index	S&P CoreLogic	Price Index	Tracks changes in the value of residential real estate	Monthly
Refinitiv Spot Gold	LSEG Workspace	Price Index	Tracks gold prices	Daily
Natural Gas	FRED	Gas Spot Price	Tracks Gas Spot Price	Daily

4. Methodology

4.1. Time-Series Cross-Validation (TSCV)

Predictive modeling with time series has some complexity, the standard k-fold that randomly shuffles data raises some problems in this situation. When dividing the data into training, validation and test set, we should be careful to prevent data leakage from the future. We want to preserve the temporal order of the variables to avoid look ahead bias, so in this case we used an expanding window approach.

Each iteration consists of three temporally ordered segments: a training set, a validation set, and a one-month out-of-sample (OOS) test set. The training window begins with an initial length of 60 months and expands forward over time, while the validation window spans a fixed 12-month period directly following the training data. The test set comprises a single month and is positioned immediately after the validation window. The windowing process advances by one month in each iteration, effectively simulating a rolling forecast in which models are updated and re-evaluated on a monthly basis.

$$\textbf{Train} = \{x_0, x_1, \dots, x_{ti}\}$$

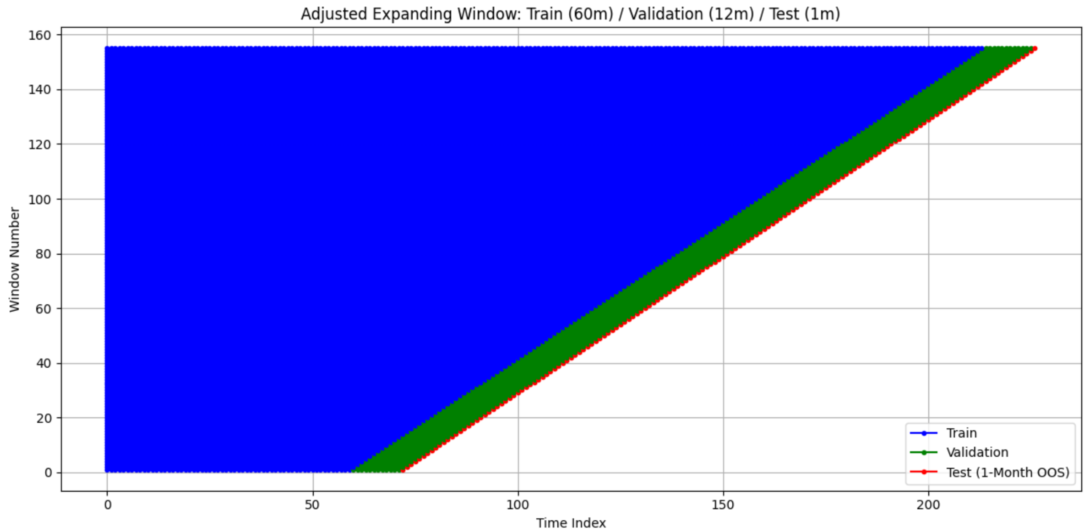
$$\textbf{Validation} = \{x_{ti+1}, \dots, x_{ti+12}\}$$

$$\textbf{Test} = \{x_{ti+13}\}$$

$$\mathbf{N} = T - (W_{train} + W_{val} + W_{test}) + 1$$

So, this structure yields a total of $N=T-73$, which is equal to 155 rolling splits being T the number of observations, this enables a comprehensive and temporally consistent model evaluation. The use of an expanding training window allows the model to learn from an increasing volume of historical data, while the fixed-length validation window supports hyperparameter tuning in a manner that avoids data leakage. The one-month test window ensures that performance metrics reflect a realistic out-of-sample forecasting ability. A graphical illustration of this methodology is presented in the graph below, where the training, validation, and test periods are depicted in blue, green, and red respectively.

Figure 1: Expanding window



4.2. Baseline Model

As a benchmark for our models, we have a simple naïve forecasting model where we use the average of past returns to compute the return of the next month. Specifically, for each test period, the forecast is given by the arithmetic mean of returns observed during the preceding 12-month validation window. Below you have the mathematical formula.

$$\hat{r}_{t+1} = \frac{1}{W_{Val}} \sum_{i=1}^{W_{Val}} r_{t-i+1}$$

With this we want to verify if more complex models really add value to the prediction, as shown by Welch and Goyal (2008) many models do not outperform a simple historical average in OOS, financial data sometimes is noisy which produces a low noise to signal ratio and more complex models may perform very well in sample and then fail to generalize in OOS, that's called overfitting. So is important to have different levels of complexity between models.

4.3. LASSO Regression

Least Absolute Shrinkage and Selection Operator, also known as Lasso, is a linear model that has a regularization ℓ1 term to control model complexity by converging some coefficients to zero, depending on how strong the signal of them is, since we have a lot of variables is good to shrink the universe to the ones that have significance.

The model cost function is the following:

$$\boldsymbol{Cost} = \lambda \sum_{j=1}^{p} |\beta_j|$$

Where j are the coefficients and is the penalization term, the higher the more severe will be the punishment for coefficient with less significance. The tuning parameter λ is chosen via **cross-validation**, basically is chosen in each fold in the validation set, in this way we optimize the predictive ability of the model while controlling for overfitting.

4.4. Ensemble Methods

The predictive modeling strategy used in this study is based on ensemble learning, a subclass of machine learning methods in which several model that are often referred to as "weak learners" are deliberately combined to generate a prediction that is more reliable and accurate. Because they can handle complex, non-linear relationships in the data, reducing overfitting, and enhancing generalization of performance, ensemble methods are especially effective in financial applications. Given the complex interactions between our different types of variables, ensemble learning offers a strong basis for capturing these complex relationships. The two main families of ensemble methods are bagging and boosting. Although they achieve this through essentially different mechanisms, both strategies seek to mitigate the variance and/or bias of individual predictive models. In order to compare the predictive power and robustness of each algorithm in predicting sector ETF returns, this study uses one representative algorithm from each category: XGBoost for boosting and Random Forest for bagging.

The decision tree, is a non-parametric supervised learning technique for regression and classification, is at the heart of both algorithms. Because decision trees can model the effects of interactions between variables, do not assume linear relationships, and are inherently capable of handling both numerical and categorical features, they are ideally suited for financial time series forecasts. Single decision trees, on the other hand, are notoriously unstable and prone to overfitting. By building a sequence of decision trees in a manner that balances bias and variance, ensemble techniques such as bagging and boosting overcome those limitations.

By averaging the predictions of several decision trees trained on various bootstrap samples of the data, a parallel ensemble technique known as "bagging," short for "bootstrap aggregating," reduces variance. A randomly selected subset of the data is used to train each decision tree in Random Forest, and only a random subset of features is taken into account at each split within a tree. By decorrelating the individual trees, this extra randomness strengthens generalization and the overall model's resilience. Additionally, Random Forests are well-known for their integrated feature importance metrics, which offer important information about the relative contributions of various predictors. Ensemble approaches, especially those based on decision trees, were chosen because of their interpretability and capacity to represent complex and non-linear relationships. These models provide an adequate balance between flexibility and explanatory power in the context of financial forecasting, particularly when the relationship among the variables is not clear.

4.5. Portfolio Construction

A systematic, rules-based approach is used to build portfolios in order to put into practice the model outputs to investable strategies. Each model generates return forecasts for the whole sector ETF universe on a monthly basis. The ETFs are then ranked using these projections, which serve as the foundation for portfolio construction. There are three different portfolio strategies used. The first is a long-only portfolio, which invests equally in the three ETFs that are projected to produce the highest returns over the next month. This strategy is replicated in the second, which is a short-only portfolio that takes equal-weighted short positions in the three ETFs with the lowest expected returns. Lastly, by concurrently going long the top three ETFs and short the bottom three, with equal weighting on both sides, a long-short portfolio is created.

We implement this strategy across five predictive models: a naïve benchmark, LASSO regression, XGBoost, and Random Forest. In addition, we include a combined model, which averages the predicted returns from the three machine learning models (LASSO, XGBoost, and Random Forest) to form an ensemble forecast. This ensemble aims to mitigate model-specific biases and improve robustness. Performance is benchmarked against the SPY ETF, which represents a passive investment in the S&P 500 index. By comparing each strategy's performance to SPY, we assess the added value of model-driven sector rotation relative to a static, broad-market exposure.

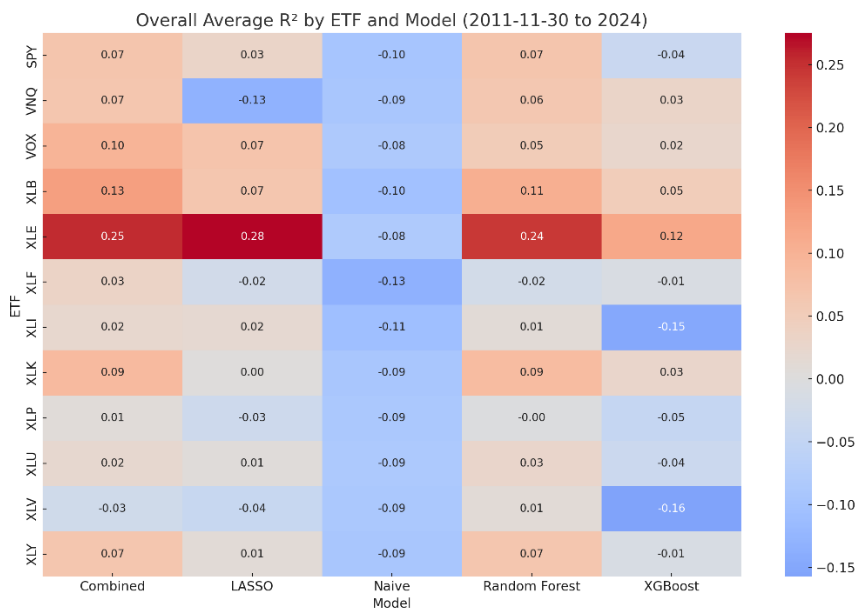
5. Results

5.1. Models Performance and Efficacy

We evaluate the predictive performance of the models employed, XGBoost, Random Forest, LASSO, Combined model and a naive benchmark across multiple aspects and periods. We analyze both statistical forecasting accuracy, using out-of-sample R^2, and practical efficacy, particularly in ranking sectors correctly. Special attention is given to how model performance varies across sectors and market regimes (pre-2020 vs. post-2020), offering insights into robustness and adaptability. The goal is to establish not only which models predict well, but under what conditions they succeed or fail the strategy implemented.

To quantify each model's predictive ability, we computed the out-of-sample R^2 using the rolling window approach, as we already explain before. Overall, the R^2 values are relatively low, for some sector even negatives, due to inherent noise and low signal-to-noise ratios.

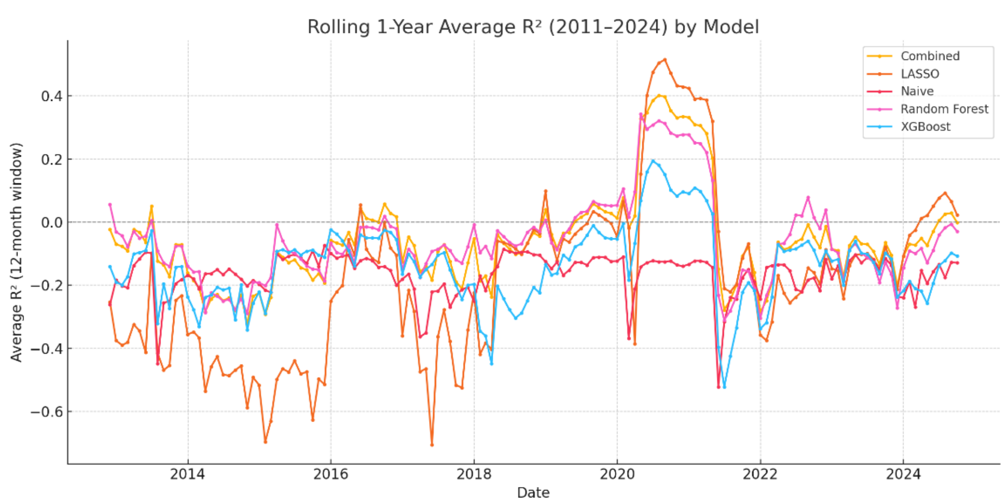
Figure 2: Models Average R^2



Regardless of that, there are notable variations in predictive performance across models and sector ETFs, which we will explore and highlight in detail throughout this section. Starting with the best model Random Forest, it frequently obtained the most stable and generally positive R^2 values across sectors. With especially great performance in cyclical sectors like XLE (Energy) and XLB (Materials), where it achieves R^2 values of 0.27 and 0.16 respectively, its average R^2 across all ETFs is 0.07. LASSO, despite being a linear model, performs surprisingly well. It achieves comparable performance to Random Forest in several sectors. In particular, it scores 0.34 R^2 in XLE and 0.18 in XLB, probably indicating that many relationships between macroeconomic indicators and sector returns are at least partially linear. LASSO also performs consistently well in defensive sectors such as XLP (Consumer Staples) and XLU (Utilities), where complex nonlinear interactions may be less pronounced. XGBoost, while powerful in theory, underdelivers in this context. Its average R^2 is lower and often negative in certain sectors. Although it shows in some sector potential, achieving $R^2 = 0.15$ in XLE and 0.08 in VOX, its performance is generally erratic across time and sectors. This suggests either overfitting in small samples or a mismatch between the model's capacity and the simplicity of the underlying signal.

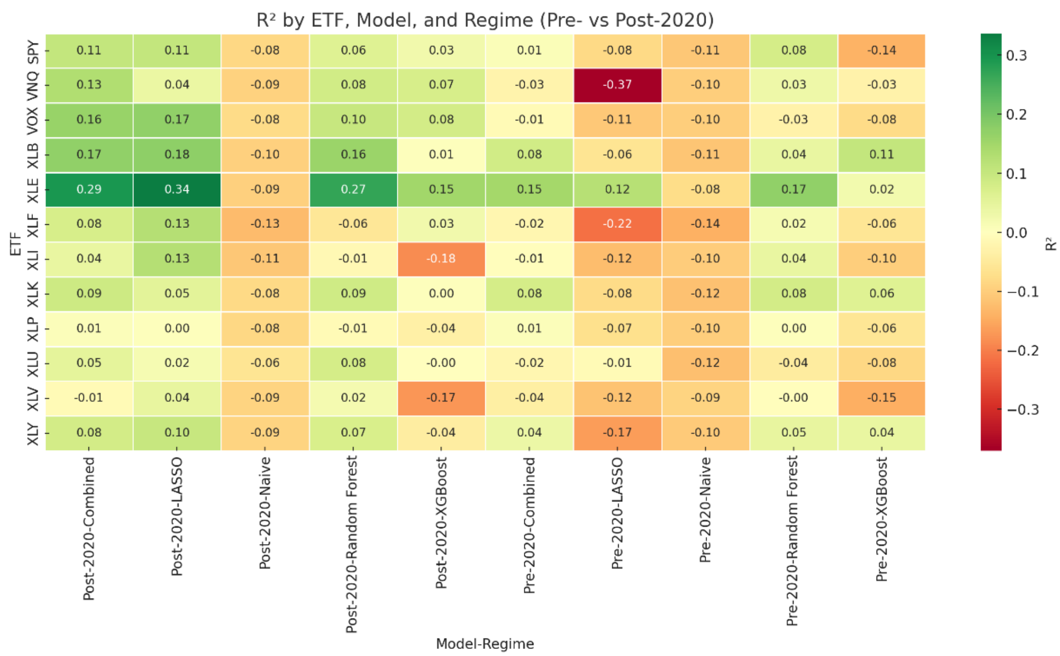
The Combined Model, that is basically an average of the predictions from Random Forest, LASSO, and XGBoost, achieves a more stable performance than XGBoost alone and occasionally reaches the top single models perfomance. For instance, it has a 0.15 R^2 in XLE and performs above 0.10 in several other sectors. Its advantage lies in reducing idiosyncratic model variance and capturing diverse types of signals (linear and nonlinear). Lastly, the Naive model, that uses the previous 12 months return average as the forecast, performs the worst by far. Its average R^2 is negative across most sectors. This result is intuitive: one-month return autocorrelation is typically low or negative in equity sectors, and naïve extrapolation is ineffective, this type of result was expected.

Figure 3: One Year Rolling R^2



Plotting the rolling 1-year average R^2 helped us to evaluate the predictive power change over time for every model. Following 2020, there was a clear change that forced more research on how different market regimes affect model performance. In our analysis we break down performance into two regimes: Pre-2020 (before COVID-19) and Post-2020 (from January 2020 onward). This allows us to assess how model efficacy shifts across different macroeconomic environments. The heatmap below visualizes each model's out-of-sample R^2 across all ETFs, segmented by regime.

Figure 4: R^2 by ETF, Model and Regime



Warmer colors indicate lower or negative explanatory power, while cooler colors highlight stronger model performance.

Random Forest stands out again for the consistency, achieving R^2 gains in 11 of the 12 ETFs, with an average post-2020 increase of +0.064. Its nonparametric, ensemble structure allows it to capture nonlinear interactions between the different variables and sector dynamics, which may become more relevant in turbulent periods such as COVID. LASSO shows the largest average gain, improving by +0.073, despite its linear structure. This suggests that in volatile regimes, linear effects become stronger or easier to isolate. Notably, LASSO's large improvement is concentrated in cyclical sectors XLE (Energy), XLB(Materials), and XLK(Technology) suggesting that structural macro shifts have linear predictive components that were previously muted. For instance, Energy (XLE) sees a jump from near-zero R^2 pre-2020 to **0.27+** post-2020 across all models.

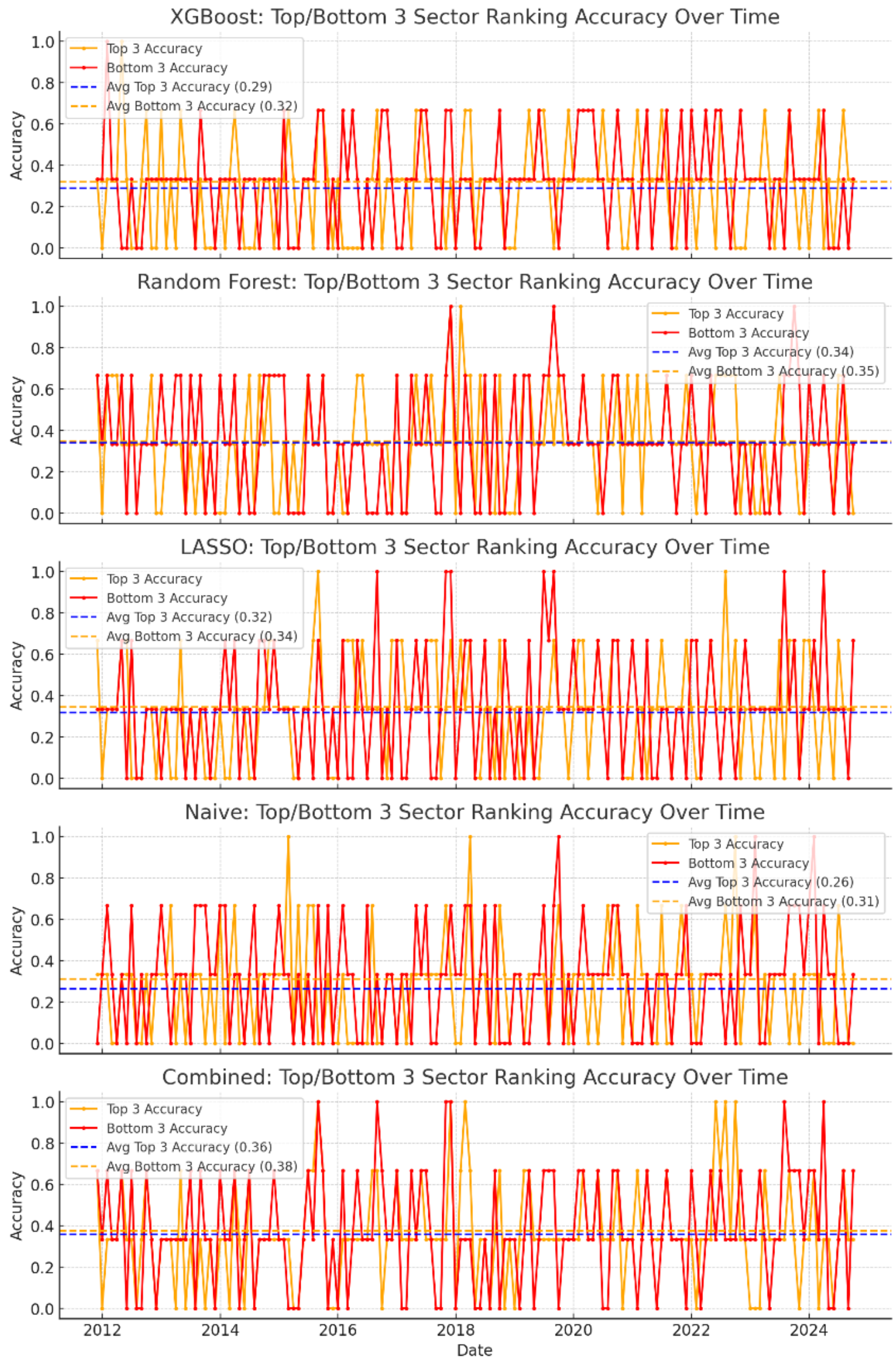
Being the Combined model an average model, is natural that the result after 2020 is better since it benefits from the strengths of each. Its average gain of +0.066 post-2020 confirms that ensemble diversity helps stabilize forecasts in volatile conditions. Both XGBoost and the Naïve model show a weak and inconsistent performance, while this results were expected for the naïve model, for XGBoost it may be related to some overfitting in sample that makes it not able to extrapolate good performance OOS. Even tough that it had a very small improvement in post-2020 , we can affirm that both models are not able to perform well in whatever regime.

Table 4: Average R^2 Post-2020

Model	Avg. R^2 Gain Post-2020
LASSO	+0.073
Combined	+0.066
Random Forest	+0.064
XGBoost	+0.017
Naive	+0.000

Apart from estimating precise return magnitudes, a main goal of sector rotation is the proper sector ranking. Choosing the best- and worst-performing sectors usually concerns investors more than reducing prediction errors in general. We evaluate how precisely each model detects the top 3 and bottom 3 performing sector ETFs (excluding SPY, the benchmark) at each monthly rebalancing point to assess this dimension of model efficacy. We define ranking accuracy as the number of correct matches (0 to 3) divided by 3, resulting in a score between 0.0 and 1.0 for each month and each group (top and bottom). This gives us two accuracy time series per model: one for the top 3 and one for the bottom 3. We then compute a rolling average over time to assess consistency and visualize average performance over the full window.

Figure 5: Models Accuracy Overtime



The combined model leads in both dimensions with the highest average accuracy in both categories. Its top-3 accuracy is 0.36, and bottom-3 accuracy reaches 0.38. This confirms the ensemble's robustness and diversification benefits: while individual models may misfire, their average prediction converges toward correct rankings more consistently. While LASSO, a linear and more constrained model, shows almost as effective, recording 0.32 and 0.34 accuracy in the top and bottom 3 rankings, respectively, Random Forest achieves the highest average ranking accuracy among the standalone models with a top-3 accuracy of 0.34 and bottom-3 accuracy of 0.35. In more steady, defensive sectors like Utilities (XLU) and Consumer Staples (XLP), where return behavior is more in line with consistent macroeconomic drivers such interest rates, inflation, and consumption trends, LASSO performance is good, since this model's strength is its capacity to preserve interpretability and lower noise by penalizing pointless features, so producing stable, if more conservative, sector rankings. Together, the two models show a complementary pattern: LASSO offers strong performance in more linear, stable regimes while Random Forest thrives in more dynamic and high-volatile environments by using complex interactions. Though theoretically is the most flexible model, XGBoost exhibits a lot of fluctuation over the period; ranking performance is modest: 0.29 for top-3 and 0.32 for bottom-3. Its tendency to overfit in some windows could cause sharp but misplaced confidence, so compromising ranking precision. Still, it beats the naive model. That has the lowest top-3 accuracy of 0.26, the naive approach shows that recent winners are not consistently staying as winners one month ahead. Surprisingly, its bottom-3 accuracy (0.31) is more in line with those of machine learning models, implying some persistence in poor sector performance that even a simple signal could capitalize from.

The time series of accuracies expose significant dynamics even if the average accuracies offer a high-level comparison, the models accuracy often increases during high-dispersion periods (e.g., the COVID-19 crash/recovery and the 2022 inflation cycle). This shows up in the spikes for 2020 and 2022 across almost every model. On low-volatility, periods like 2016–2018, on the other hand, ranking accuracy falls, which reflects the less chance to separate sector leadership.

5.2. Feature Importance Analysis

To better understand the drivers of forecast accuracy, we examine the average feature importance across all models and sectors. Feature importance was computed using model-specific methodologies: absolute coefficient magnitudes for LASSO, mean impurity reduction for Random Forest, and gain-based importance for XGBoost. Each method reflects the degree to which a feature contributes to prediction accuracy. We also examine trough SHAP (SHapley Additive exPlanations) each feature, it offers a local view of feature impact, quantifying how each feature value pushes a prediction higher or lower relative to the mean. Unlike global importance, SHAP accounts for interaction effects and distributional variance, especially critical in non-linear models like XGBoost or Random Forest.

Particularly in non-linear models like Random Forest and XGBoost, the VIX Volatility Index and rolling 12-month volatility repeatedly ranked highest among all the predictors. Often showing in the top three across sectors (e.g., SPY, XLK, XLF), these characteristics indicate that sector performance predictions are mostly driven by market uncertainty and recent return variability.

Comparisons of SHAP values pre- versus post-2020 expose significant changes in feature behavior. Reflecting increased macro sensitivity in the COVID/post-COVID era, UMCSent (consumer sentiment) and EPU (Economic Policy Uncertainty) gained impact post-2020 for Random Forest and XGBoost. Conversely, historically steady indicators like 10y yields and yield curve spreads (10y-2y) grew rather less useful in recent years, particularly under LASSO's coefficient-based lens. Though linear, the LASSO model reinforced the dominance of VIX, pct_wti, and rolling volatility with some reduction in absolute coefficients post-2020, implying that while these variables remain important, their direct linear impact may have somewhat weakened or become more complex. Sector-level breakdowns support this trend. In tech (XLK) and industrials (XLI), XGBoost and RF models underlined UMCSENT and VIX as dominating; in defensives like XLU and XLP, cumret and rolling volatility maintained steady importance. The wide consistency of these characteristics over models and time spans supports their function as strong predictors in tactical sector rotation. With their relevance increasing post-2020 in tree-based models, macro-sentiment indicators (VIX, UMCSent, EPU) and recent volatility measures are overall central to predictive performance. This emphasizes the advantage of SHAP in capturing localized, changing relationships inside models as well as structural changes in market behavior.

Figure 6: Average Feature Importance (Random Forest)

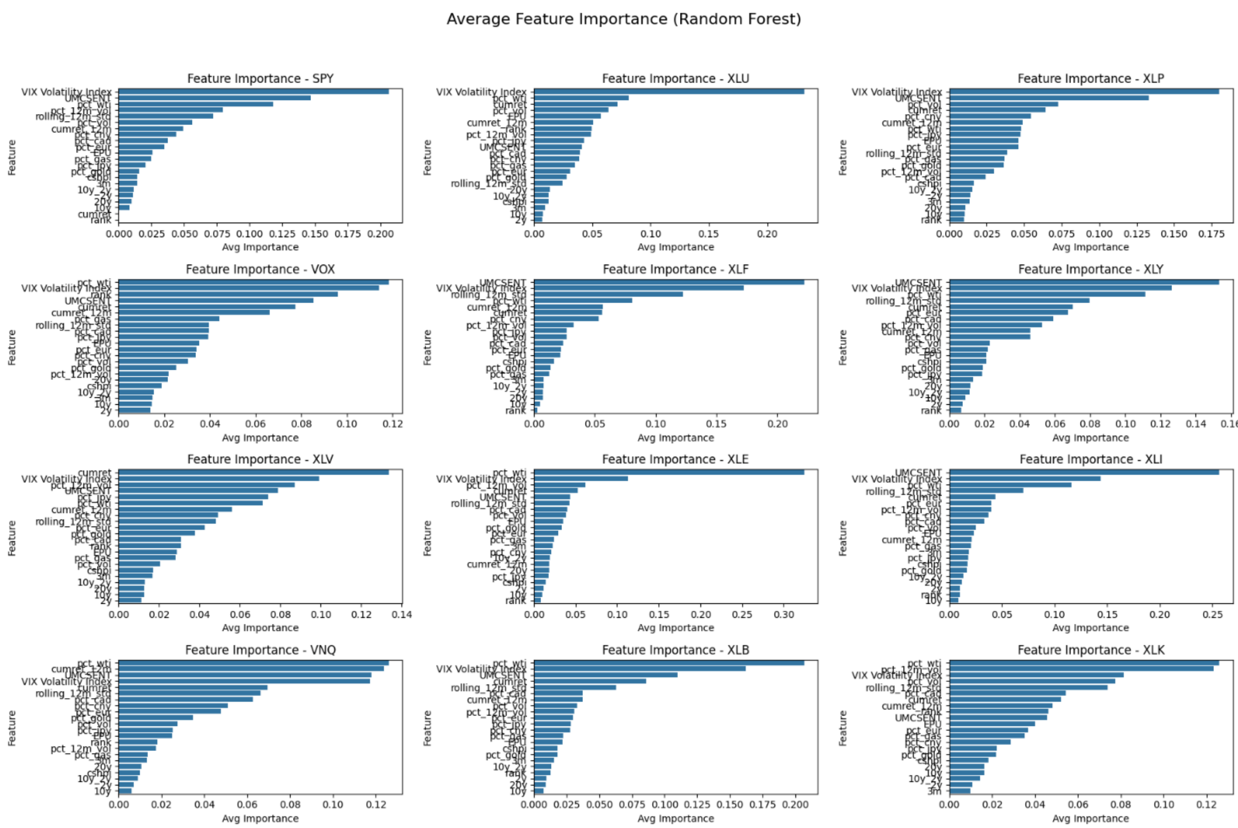


Figure 7: Average Feature Importance (XGBoost)

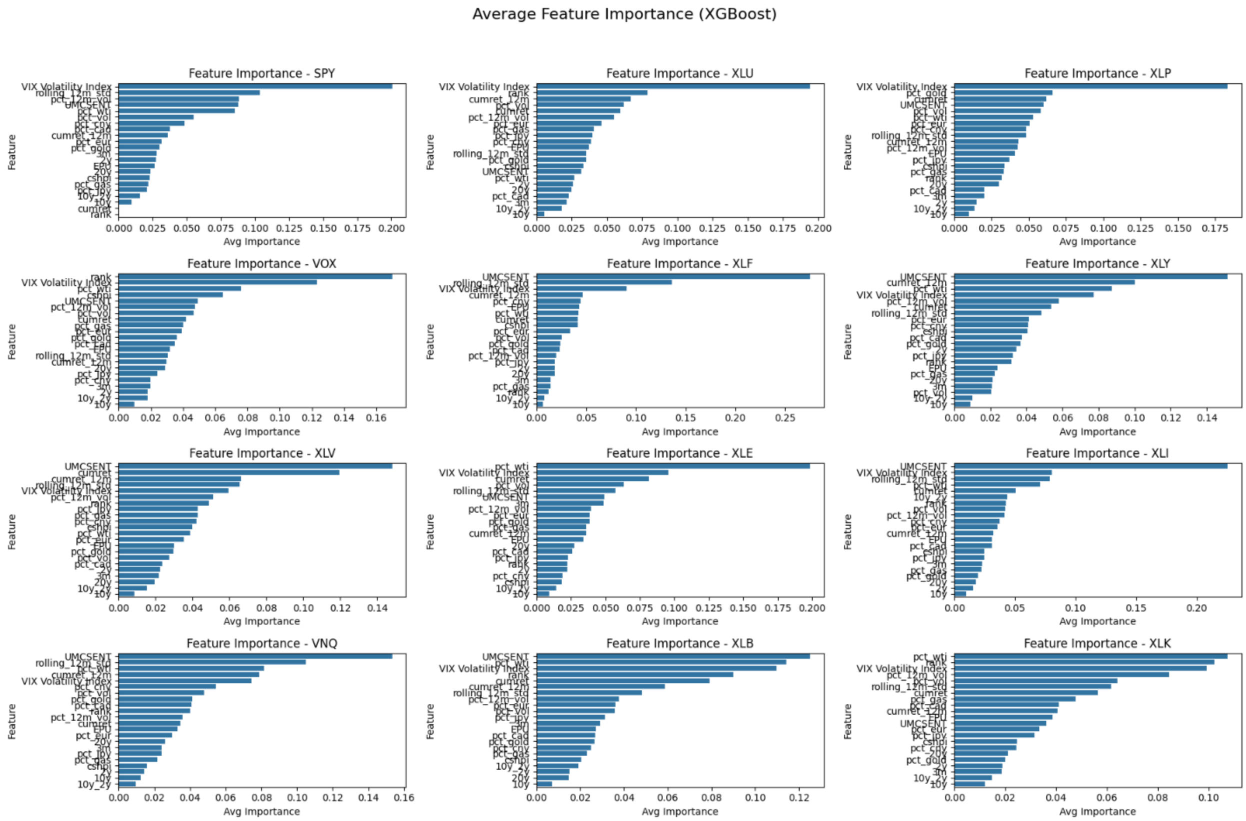


Figure 8: Average Feature Importance (LASSO)

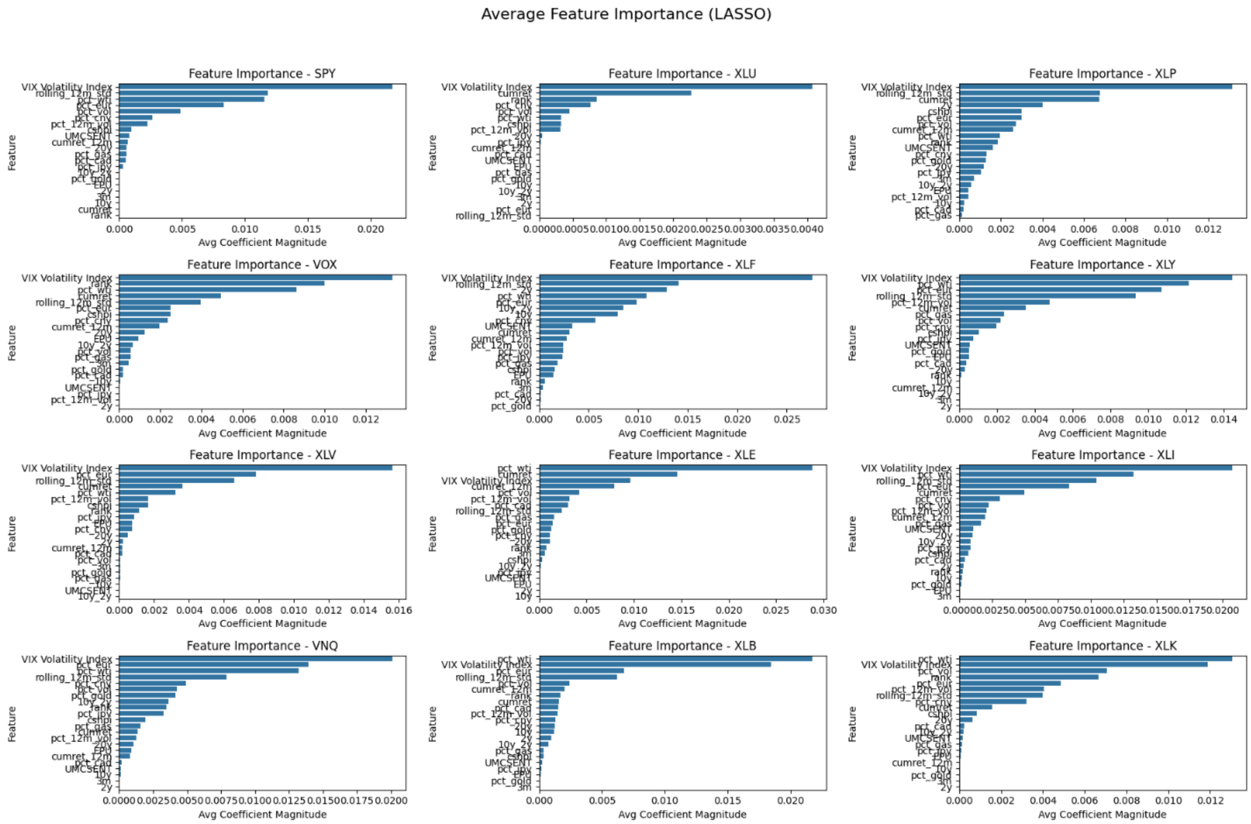


Figure 9: Average SHAP (Random Forest) Pre vs Post 2020

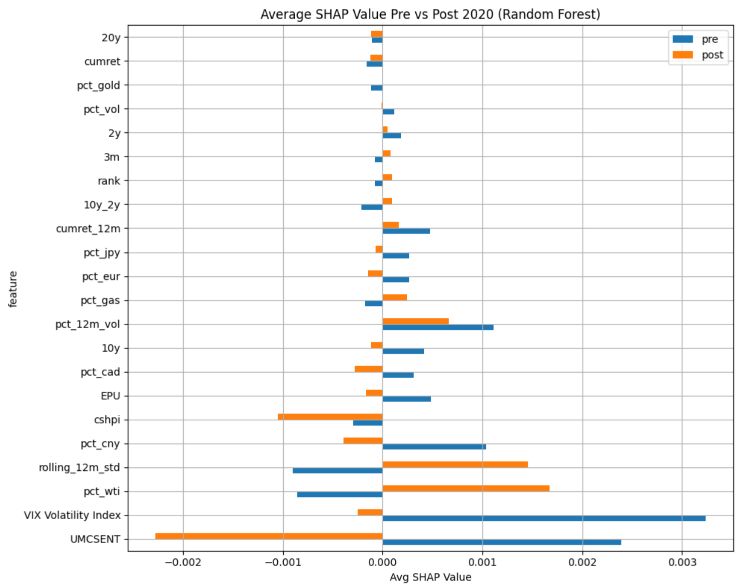


Figure 10: Average SHAP (XGBoost) Pre vs Post 2020

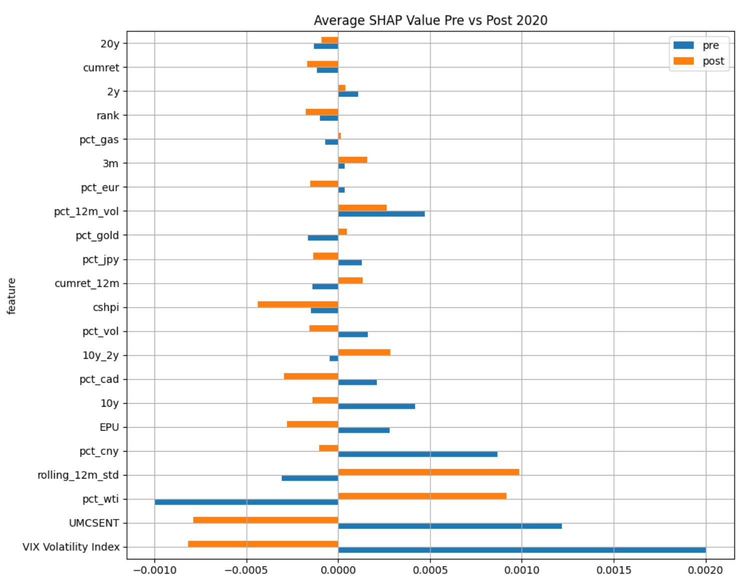
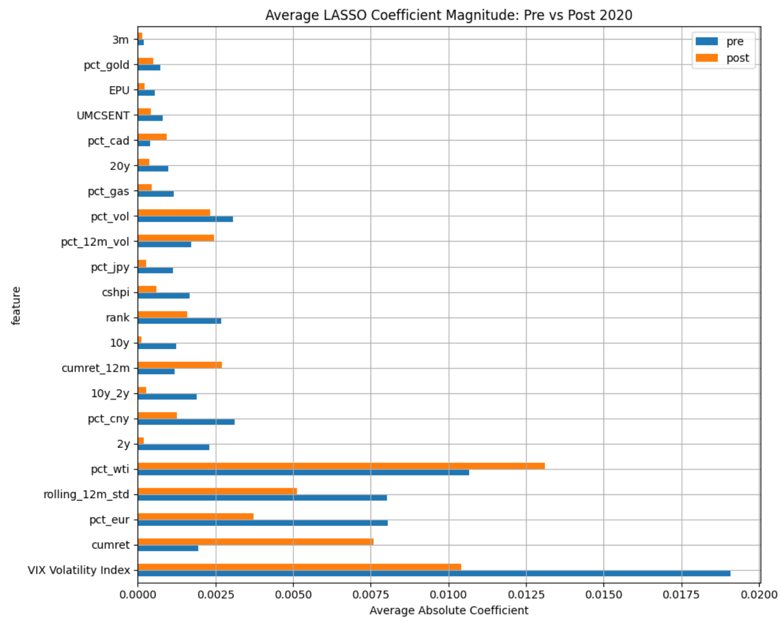


Figure 11: Average LASSO Coefficient Magnitude Pre vs Post 2020



5.3. Portfolio Performance

Now we are going to see the performance of the machine learning-based sector allocation strategies by contrasting the five models in three different portfolio constructions: long (top three sectors), short (bottom three sectors), and long-short (spread between top and bottom three). Metrics are annualized for comparability, and all strategies are benchmarked against the SPY ETF between November 2011 and October 2024. It is important to state that it was assumed no transaction costs, since we have a buy and sell strategy the cost is almost null, ETFs are costly to hold not to trade.

Table 5: Portfolios Perfomance

Portfolios	Ann. Return	Ann. Volatility	Sharpe Ratio	Max Return	Min Return	VaR 5%
SPY Benchmark	14.41%	14.45%	1	12.70%	-13.00%	-6.25%
XGBoost Long	17.15%	14.71%	1.17	13.84%	-11.50%	-6.29%
XGBoost Short	4.08%	16.06%	0.25	17.37%	-18.24%	-7.26%
XGBoost Long_short	12.61%	8.28%	1.52	10.35%	-4.80%	-2.40%
Random Forest Long	22.26%	15.76%	1.41	17.55%	-12.73%	-5.91%
Random Forest Short	1.60%	15.10%	0.11	10.15%	-19.81%	-7.16%
Random Forest Long_short	20.37%	10.57%	1.93	10.39%	-5.07%	-3.06%
LASSO Long	20.06%	15.50%	1.29	19.90%	-10.39%	-5.35%
LASSO Short	1.82%	15.17%	0.12	12.23%	-21.12%	-6.98%
LASSO Long_short	17.94%	11.20%	1.6	12.30%	-11.34%	-3.13%
Naive Long	11.64%	13.86%	0.84	11.97%	-11.51%	-5.60%
Naive Short	10.08%	17.79%	0.57	18.17%	-23.21%	-7.48%
Naive Long_short	1.42%	12.56%	0.11	11.70%	-13.61%	-6.56%
Combined Long	23.85%	15.42%	1.55	18.25%	-11.50%	-4.63%
Combined Short	-0.32%	14.92%	-0.02	9.18%	-21.12%	-7.25%
Combined Long_short	24.24%	11.29%	2.15	10.63%	-5.07%	-3.32%

Key performance metrics are summarized in the table above . The Combined Long-Short portfolio produced the highest annualized return (24.24%) and overall Sharpe ratio (2.15). With strong positive outliers, like a maximum monthly return of 10.35%, a relatively low volatility (11.29%) with a conservador downside risk (VaR 5% of –3.32%).

Long-short strategies constantly outperform long-only portfolios, followed by short-only strategies that underperform both in returns and risk-adjusted measures; the performance results across all portfolios show a clear hierarchy in efficacy. Table 5 shows more apparent these trends.

With an annualized return of 24.24%, a Sharpe ratio of 2.15, and a modest annualized volatility of 11.29%, the Combined Long-Short portfolio turned out to be the top-performing strategy. This suggests a quite effective utilization of risk to produce constant relative gains. With a Sharpe of 1.93 and a return of 20.37%, the LASSO Long-Short approach also showed good results. Long-short approaches showed better performance overall among all models since they combined quite high average returns with lower volatility and narrower downside tails. Their lowest monthly losses (min returns) were far softer than those seen in short-only portfolios; their max monthly returns varied from over 9% to 10.35%. Attractive returns were also produced by long-only techniques. For example, the Combined Long and Random Forest Long portfolios yielded respectively 23.85% and 22.26% annually. But these portfolios showed larger volatility (15–16%) and more noticeable tail risks, the min monthly returns were beyond –12% in some situations and VaR (5%) values either near or below –5%. These numbers show their exposure to overall market fluctuations, which increases performance in optimistic times but provides less downside protection in stressed conditions. By contrast, short-only portfolios often underperformed all the models. For example, the following: With a Sharpe ratio of 0.11, the Random Forest Short portfolio reported an annualized return of just 1.60%. Although they obtained double-digit max monthly gains in rare cases, the XGBoost Short and LASSO Short portfolios had similarly low Sharpe ratio 0.25 and 0.12, respectively. Their worst monthly returns, ranging from –18% to almost –20%, and VaR 5% values around –7%, greatly worse than long or long-short portfolios, therefore highlighting the poor performance of short strategies. These results capture key structural and behavioral traits of the equities markets during this period: We observe a persistent market uptrend in our time window of results, that makes shorting sectors unviable, even if they are expected to underperform relatively with the benchmark, and in a volatile or transitional period, poorly performing sectors sometimes show strong comeback. Strategies that short recent underperformers are vulnerable to sharp reversals, especially when sector leadership changes abruptly this results in disproportionately large losses in short portfolios, a limitation of our strategy is that is monthly rebalanced. Conversely, long-short strategies concentrate on using dispersion in relative sector performance to hedge market orientation. This results in better risk-adjusted metrics. For instance, every long-short portfolio (except for Naive) exceeded the Sharpe ratio 0.997 of the benchmark SPY , so emphasizing their ability to capture alpha in excess of market beta.

The cumulative return plots below show clear divergence starting in 2020, with some models effectively adjusting to the new regime while others become worse. This suggests to separate analysis of these two periods because there is a possibility that the predictive efficacy and risk characteristics of the models are regime dependent. These variations across return, volatility, and risk-adjusted measures are then quantified with a comparison of pre and post 2020 results.

Figure 12: Cumulative Long-Short Strategy Returns

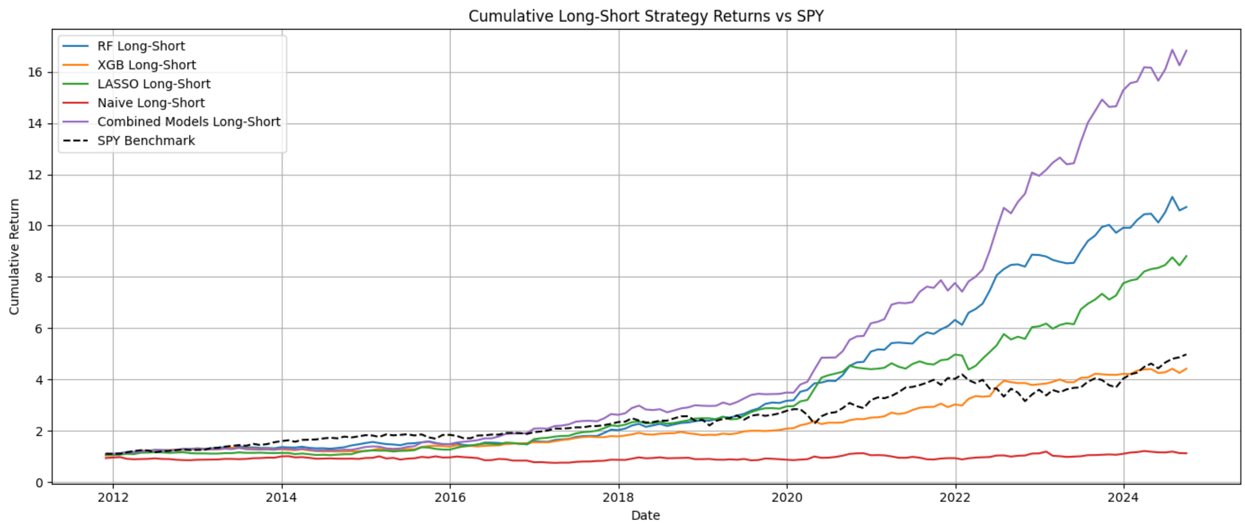


Figure 13: Cumulative Long-Only Strategy Returns

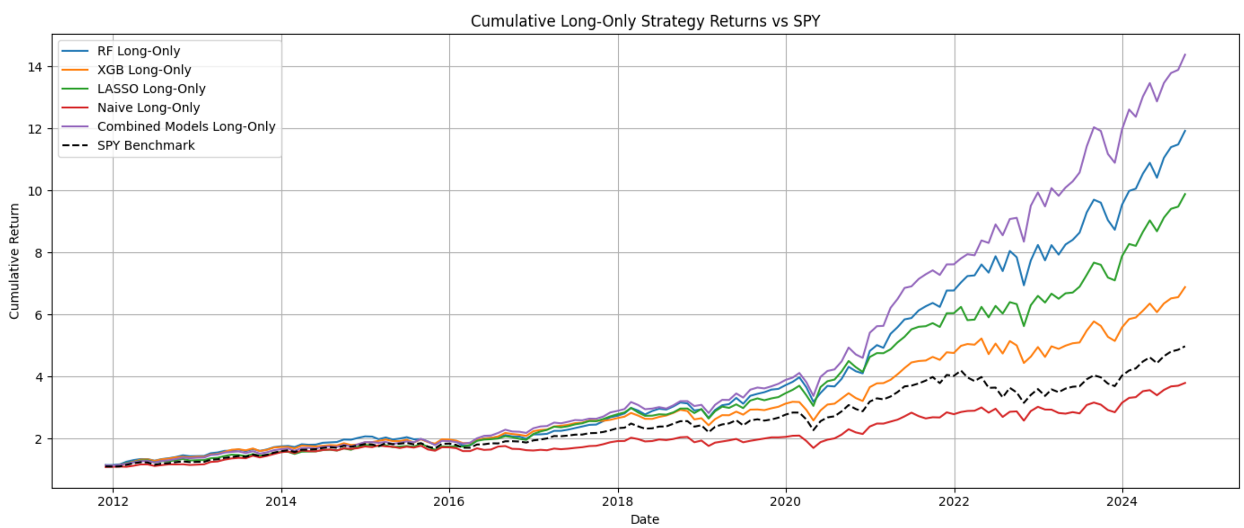
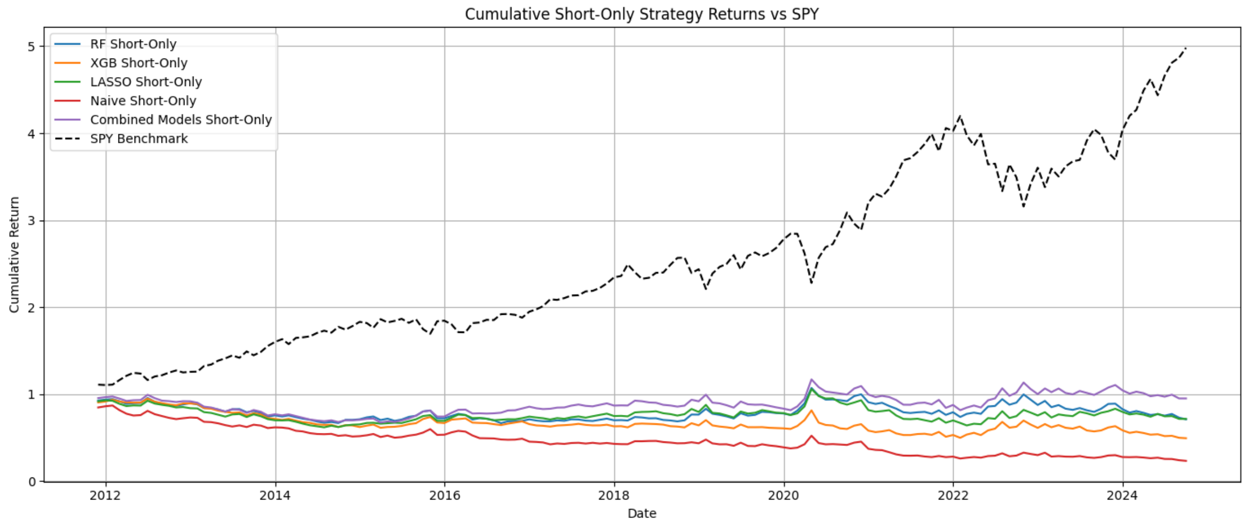


Figure 14: Cumulative Short-Only Strategy Returns



During both times, the Random Forest model performed well and consistently. Prior to 2020, its long-only strategy yielded an annual return of 18.45% with a Sharpe ratio of 1.41, and its long-short strategy yielded a return of 15.70% with a Sharpe of 1.61. These outcomes were attained in a setting with steady sector trends and minimal volatility. With a sharply higher Sharpe ratio of 2.65, the model's long-only performance increased to 30.12% and its long-short strategy soared to 30.03% in the post-2020 regime. This illustrates how the model can still extract directional information in the face of erratic market conditions. The short-only strategy, on the other hand, continued to be unreliable, returning -0.06% in the second period. XGBoost displayed comparable trends. Although the Sharpe ratios stayed low at 1.34 and 0.77, respectively, it produced a 15.78% return from its long-only strategy and a 9.67% return from its long-short approach prior to 2020. With a return of -5.34%, the short strategy was noticeably unsuccessful. The long-short strategy increased to 17.73% with a better Sharpe ratio of 1.74 after 2020, while the long-only strategy improved to 20.03%. Short exposure, on the other hand, continued to be a weak point, returning -1.95% and not adding value when market stress was high. Additionally, the LASSO model adjusted well to the post-2020 environment. Its long-only strategy yielded 17.37% in the pre-2020 period, while the long-short configuration yielded 14.69% with a Sharpe ratio of 1.52. Its short-only version performed poorly, with a return of -2.31%, similar to other models. LASSO's long-short strategy almost doubled its return to 27.40% in the more volatile post-2020 environment, and its Sharpe improved to 2.01. Additionally, the long-only approach improved, hitting 26.87% with a Sharpe of 1.40. However, with a flat return of 0.13%, the short side continued to be largely ineffective.

Throughout both time periods, the model that combines predictions from Random Forest, XGBoost, and LASSO consistently produced the best results. Prior to 2020, the long-only strategy yielded a 19.00% return with a Sharpe of 1.52, while the long-short strategy yielded a 17.12% return with a Sharpe of 1.69. The short component was still slightly negative. The group's performance significantly improved after 2020. The long-short strategy produced an impressive 40.40% return with the highest Sharpe ratio of 3.00 of any strategy, while the long-only strategy yielded a return of 34.00%. With a return of 4.89%, even the short-only component turned positive. This demonstrates how useful ensemble learning is for capturing a variety of signals and successfully adjusting to various market regimes. The SPY benchmark, on the other hand, produced comparatively consistent absolute returns during the two periods, 14.06% prior to 2020 and 15.01% following. However, because of a significant rise in volatility (from 11.48% to 18.60%), its Sharpe ratio decreased from 1.22 to 0.81. These results support a number of significant conclusions. First, in both stable and volatile regimes, long-short strategies provided the most reliable and consistent performance; after 2020, their capacity to hedge out market beta proved particularly useful. Second, long-only strategies lost stability in high-volatility environments where directional exposure meant higher risk, even though they were still successful in trend-driven markets. Third, short-only strategies typically carried a significant downside risk and did not add value, especially in tumultuous or quickly reversing conditions.

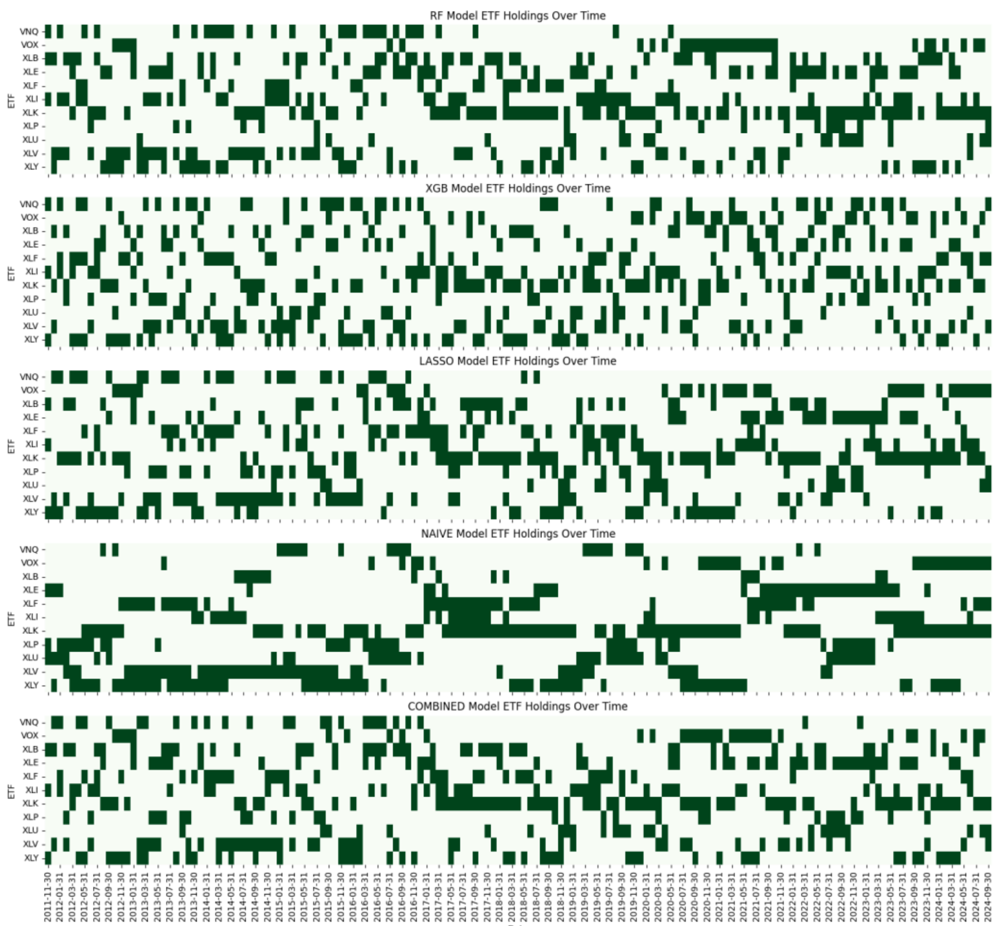
While some of the strategy delivered strong results, part of its success is due to the favorable time window chosen, most sector ETFs performed well from 2011 to 2024, especially post-2020. However, a notable issue is the high concentration in certain sectors, particularly technology (XLK). Across all models, XLK was the most frequently selected ETF, driven by its exceptional performance during this period. Because the strategy selects only three ETFs per month, it obviously concentrates risk. This approach can amplify returns when top sectors outperform, but it also increases exposure to sector-specific drawdowns. This vulnerability is highlighted by XLK's constant appearance in all models, which implies a lack of diversification even though it made a substantial contribution to historical returns. Overall, the market conditions favored a few dominant sectors, so this time window is partially responsible for the strategy's effectiveness. This calls into question the model's resilience in various market conditions, highlighting the necessity of addressing concentration risk in upcoming improvements.

Even though the performance metrics displayed, especially Sharpe ratios and annual returns, are robust across a number of machine learning-based strategies, particularly in long-short configurations, concern must be taken when interpreting these findings. Though they may seem impressive, high Sharpe ratios, like the 2.15 the Combined Long-Short portfolio attained, rarely hold up in out-of-sample situations. Their appearance in backtests frequently indicates that the model may be over-tuned to the particular sample, unintentionally taking advantage of noise, or benefiting from favorable historical conditions. The shift from predictive outputs to portfolio construction is one of the main issues. Instead of directly optimizing for Sharpe ratios or portfolio returns, the models rank sectors according to expected performance, from which portfolio weights are subsequently calculated. Errors may compound in this multi-stage pipeline, which includes feature engineering, prediction, ranking, and allocation. Instead of stable, generalizable predictive power, a seemingly strong portfolio result could be the result of fortunate coincidences in sector dynamics or ranking process (such as the arbitrary selection of the top or bottom three sectors).

Furthermore, because opposing exposures tend to cancel out market-wide movements, long-short strategies naturally lower volatility. Therefore, lower volatility rather than better predictive returns may be the reason for a higher Sharpe ratio in these portfolios. We might mistakenly believe that noise filtering is a skill because of this mechanical suppression of risk. Long-short positioning's risk-reduction benefits are frequently exaggerated, particularly if the net performance is largely dependent on a small number of important sector calls. Regime dependence is another drawback. The test period covers both the pre- and post-2020 market environments, which vary greatly in terms of macroeconomic uncertainty, monetary policy, and volatility. The average metrics may conceal underlying instability if a model adjusts well to one period but poorly to another. It is challenging to determine whether these tactics are opportunistic or robust in the absence of clear validation on unseen data or under market stress scenarios.

Finally, the apparent edge may also be inflated by the lack of transaction costs, slippage, and liquidity constraints. Despite the fact that ETFs are typically liquid, real-world frictions may still arise from the monthly rebalancing. Together, the findings imply that machine learning techniques may be useful for sector rotation; however, we should continue to doubt their efficacy and consistency until they are confirmed in larger samples, under different market circumstances, and with respect to implementation limitations.

Figure 15: ETF Held by Period and Model



ETFs monthly top 3 performers by model

6. Conclusion

This thesis looked into whether using a varied set of macroeconomic, sentiment, currency, and market-derived characteristics, interpretable machine learning models might improve tactical sector rotation by forecasting sector ETF returns. We showed that machine learning models, XGBoost, Random Forest, and LASSO among others, can with some modest efficiency capture both linear and non-linear relationships in financial data. Particularly in turbulent times like the post-2020 regime, the results show that ensemble tree-based models consistently exceeded linear benchmarks in both explanatory power and ranking accuracy. These results support the first research question: interpretable machine learning models can indeed produce with some consistency monthly predictions of sector performance. During the test period, the resulting portfolios outperformed the passive S&P 500 (SPY) benchmark and showed strong risk-adjusted performance, especially the long and long-short strategies. This supports our second research question and suggests that machine learning-informed rotation strategies can generate value. However, this result must be interpreted cautiously. The evaluation period coincided with a stablility in equity markets at first and then with a strong post-pandemic recovery during which most sectors performed well. This suggests that favorable market conditions rather than model accuracy may be responsible for a portion of the outperformance. Additionally, the strategy's real-world applicability is limited due to the significant idiosyncratic and sector-specific risk introduced by the concentration in just the top three sector ETFs. There are additional practical and methodological limitations. The analysis assumes zero transaction costs, no slippage, and perfect liquidity, assumptions that would not hold for institutional investors.

Implementing a highly concentrated allocation in just three sector ETFs would be difficult in practice for asset management firms, especially those that manage large, diversified portfolios since they are subject to regulatory restrictions. Industry-level ETFs, which provide finer granularity and lower concentration risk, may be a more practical and diversified extension of this work. Despite these limitations, the thesis adds to the increasing amount of evidence showing that asset allocation can benefit from data-driven, interpretable machine learning tools. A strong framework for managing complicated market dynamics is provided by forecasting models that incorporate sentiment indices, volatility metrics, macroeconomic and market data. To further confirm and extend these findings, future studies should investigate more detailed ETF universes, extend the time horizon, account for transaction costs, and test robustness across various regions and economic regimes.

Appendix

Figure 16: Max and Min Monthly Return by Strategy

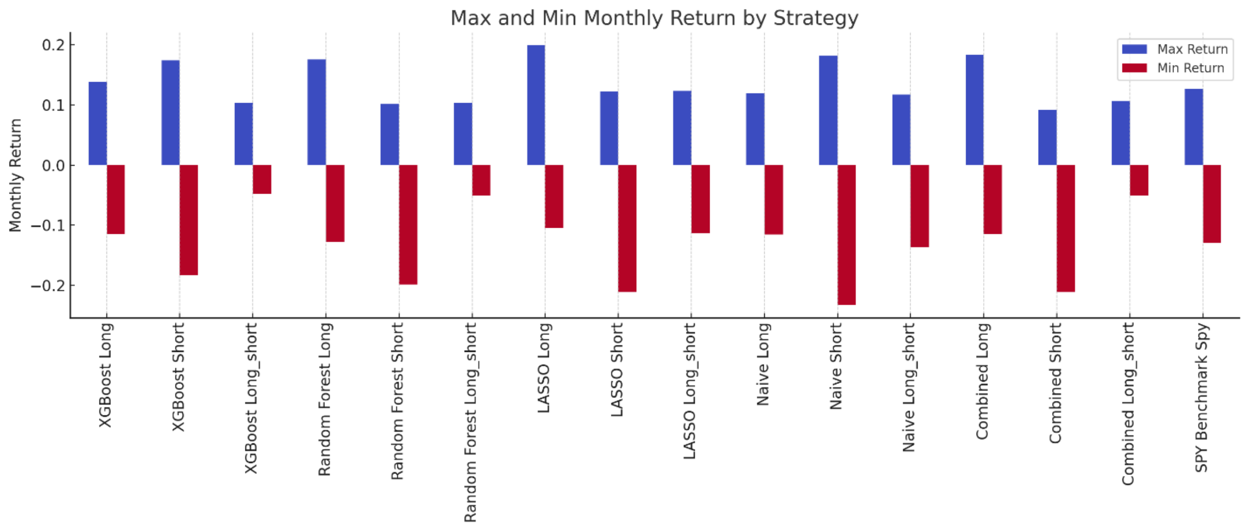


Table 6: Strategies Performance post 2020

Model	Strategy	Ann. Return	Ann. Volatility	Sharpe	Max Return	Min Return	VaR 5%
RF	long	0.301158	0.196557	1.532166	0.175504	-0.1273	-0.06504
RF	short	-0.00064	0.196086	-0.00327	0.198093	-0.09087	-0.07255
RF	long_short	0.300342	0.113178	2.65371	0.103949	-0.04791	-0.03037
XGB	long	0.200342	0.185284	1.081272	0.138351	-0.11497	-0.08393
XGB	short	-0.01953	0.219843	-0.08883	0.159286	-0.17374	-0.09771
XGB	long_short	0.177252	0.102144	1.735321	0.106797	-0.04796	-0.03508
LASSO	long	0.26873	0.192629	1.395066	0.198956	-0.11192	-0.07158
LASSO	short	0.001302	0.2057	0.006331	0.232093	-0.12227	-0.07648
LASSO	long_short	0.27035	0.139186	1.942368	0.132672	-0.11069	-0.03322
NAIVE	long	0.155443	0.180362	0.861839	0.119701	-0.11509	-0.08544

NAIVE	short	-0.07789	0.231839	-0.33598	0.232093	-0.18173	-0.11144
NAIVE	long_short	0.06648	0.143323	0.463846	0.117001	-0.13607	-0.05562
COMBINED	long	0.340015	0.195699	1.737441	0.182542	-0.11497	-0.06462
COMBINED	short	0.048933	0.206399	0.237079	0.234636	-0.09851	-0.07726
COMBINED	long_short	0.403972	0.134639	3.000398	0.119662	-0.05179	-0.03221
SPY	long	0.150139	0.185986	0.807258	0.126984	-0.12999	-0.08668

Table 7: Strategies Performance pre 2020

Model	Strategy	Ann. Return	Ann. Volatility	Sharpe	Max Return	Min Return	VaR 5%
RF	long	0.184493	0.130752	1.411006	0.153405	-0.09272	-0.04484
RF	short	-0.02353	0.118736	-0.19818	0.095055	-0.11127	-0.05455
RF	long_short	0.157007	0.097584	1.608949	0.094622	-0.05065	-0.03314
XGB	long	0.157783	0.117891	1.338381	0.135328	-0.09249	-0.04114
XGB	short	-0.05335	0.113304	-0.47086	0.094026	-0.09523	-0.05479
XGB	long_short	0.096747	0.068663	1.408998	0.072659	-0.03532	-0.02381
LASSO	long	0.17367	0.127778	1.359151	0.145546	-0.10394	-0.04914
LASSO	short	-0.02314	0.116627	-0.1984	0.101966	-0.10437	-0.04551
LASSO	long_short	0.146869	0.096554	1.521114	0.122952	-0.04102	-0.03019
NAIVE	long	0.099216	0.110568	0.89733	0.091511	-0.0895	-0.05129
NAIVE	short	-0.10059	0.138551	-0.72598	0.09769	-0.15333	-0.07085

NAIVE	long_short	-0.01052	0.114299	-0.09206	0.07195	-0.08884	-0.06248
COMBINED	long	0.190018	0.124654	1.52436	0.145546	-0.08591	-0.04307
COMBINED	short	-0.01603	0.111968	-0.14315	0.083132	-0.09184	-0.05056
COMBINED	long_short	0.171216	0.101233	1.691311	0.103008	-0.05065	-0.04047
SPY	long	0.14059	0.114841	1.224214	0.109147	-0.09334	-0.05132

Table 8: Delta between pre and post 2020

Delta between pre and post 2020	Strategy	Ann. Return	Ann. Volatility	Sharpe	Max Return	Min Return	VaR 5%
Random Forest	long	11.67%	6.58%	0.12	2.21%	-3.46%	-2.02%
	short	2.29%	7.74%	0.19	10.30%	2.04%	-1.80%
	long_short	14.33%	1.56%	1.04	0.93%	0.27%	0.28%
XGBoost	long	4.26%	6.74%	-0.26	0.30%	-2.25%	-4.28%
	short	3.38%	10.65%	0.38	6.53%	-7.85%	-4.29%
	long_short	8.05%	3.35%	0.33	3.41%	-1.26%	-1.13%
LASSO	long	9.51%	6.49%	0.04	5.34%	-0.80%	-2.24%
	short	2.44%	8.91%	0.20	13.01%	-1.79%	-3.10%
	long_short	12.35%	4.26%	0.42	0.97%	-6.97%	-0.30%
NAIVE	long	5.62%	6.98%	-0.04	2.82%	-2.56%	-3.41%
	short	2.27%	9.33%	0.39	13.44%	-2.84%	-4.06%

	long_short	7.70%	2.90%	0.56	4.51%	-4.72%	0.69%
COMBINED	long	15.00%	7.10%	0.21	3.70%	-2.91%	-2.15%
	short	6.50%	9.44%	0.38	15.15%	-0.67%	-2.67%
	long_short	23.28%	3.34%	1.31	1.67%	-0.11%	0.83%
SPY	Benchmark	0.95%	7.11%	-0.42	1.78%	-3.66%	-3.54%

Figure 17: Actual vs Predicted 1 Month Returns for all ETFs (LASSO)

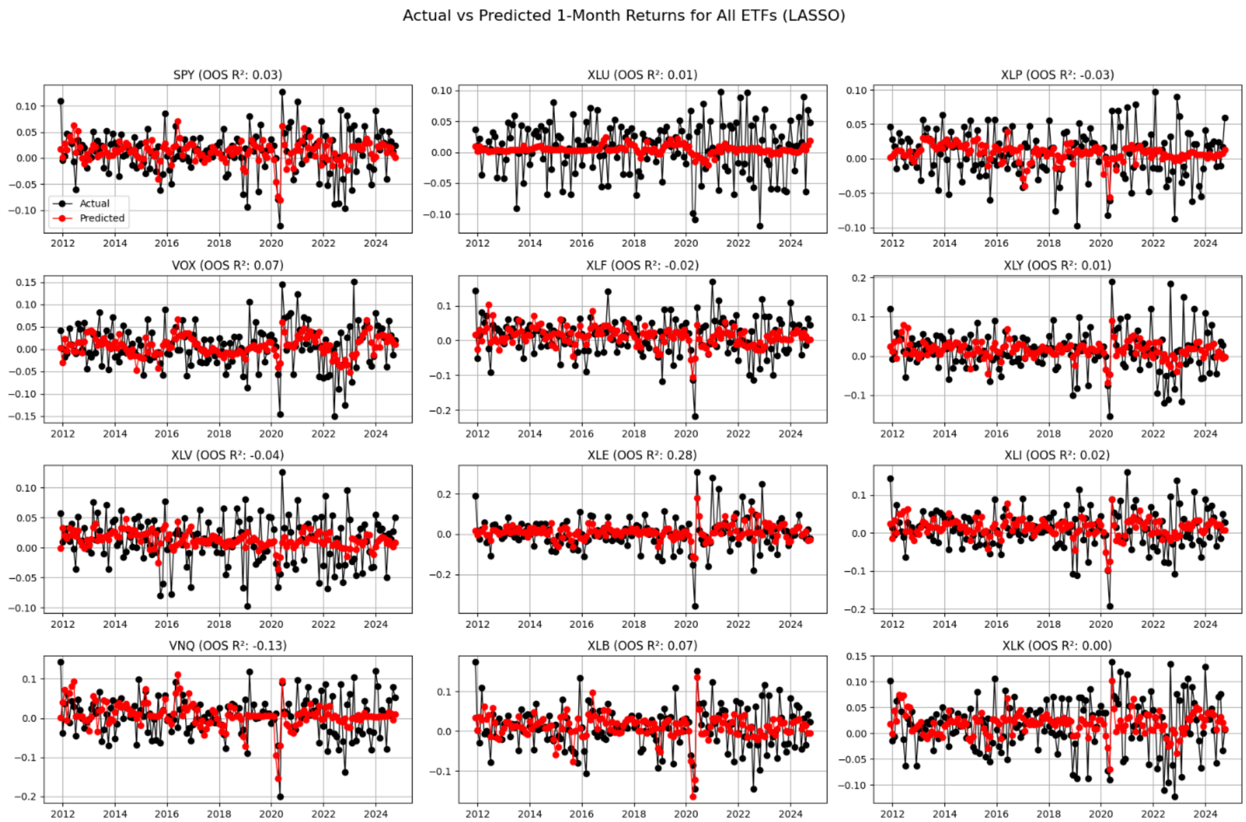


Figure 18: Actual vs Predicted 1-Month Returns for All ETFs (XGBoost with Early Stopping)

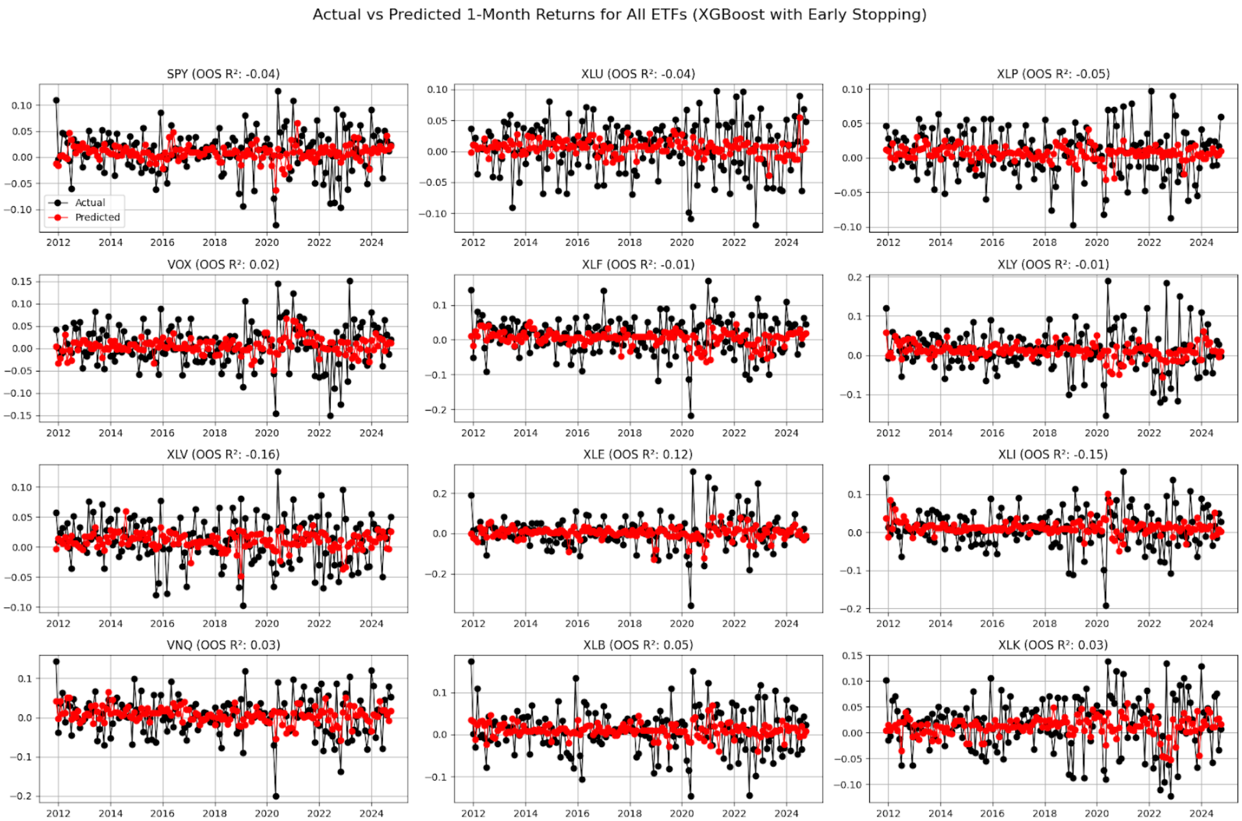
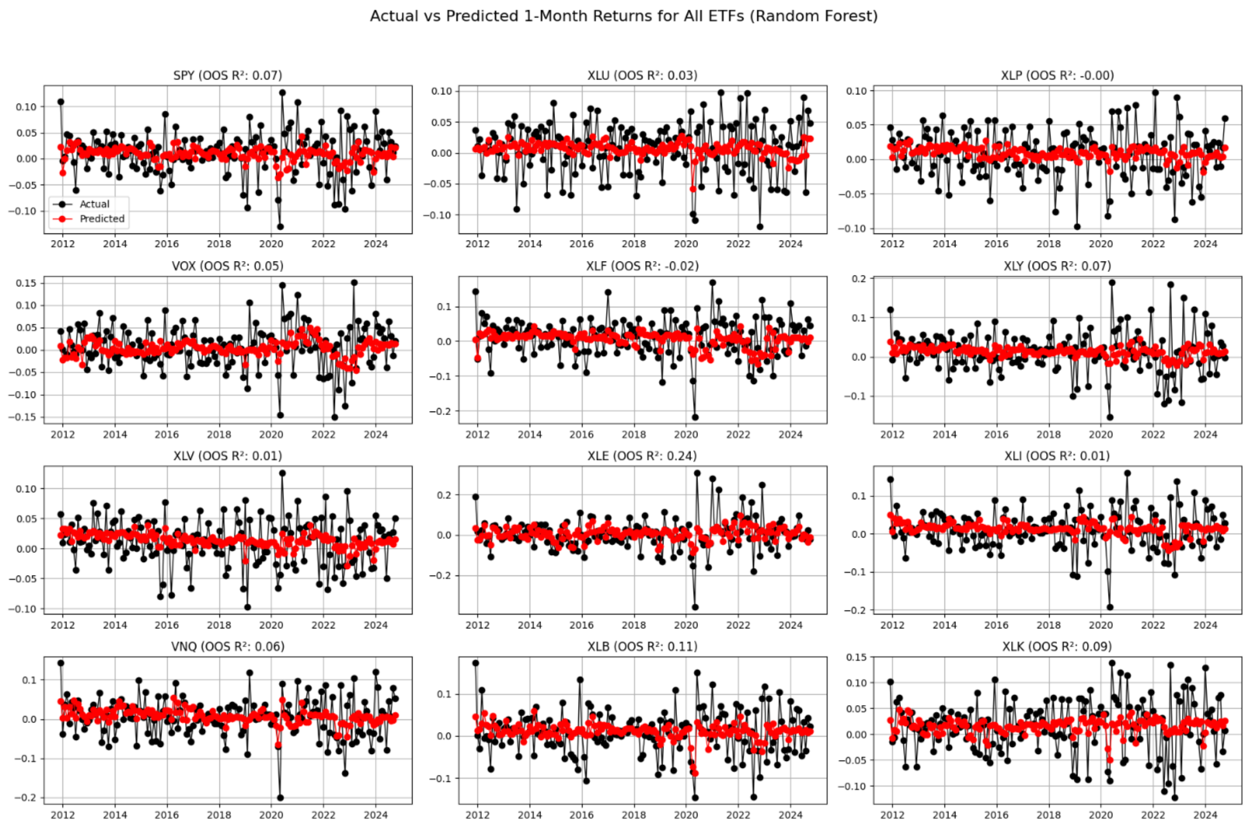


Figure 19: Actual vs Predicted 1-Month Returns for All ETFs (Random Forest)



References

Agapova, A. (2011). Conventional mutual index funds versus exchange-traded funds. Journal of Financial Markets, 14(2), 323–343. https://doi.org/10.1016/j.finmar.2010.10.005

Bryzgalova, S., Pelger, M., & Zhu, J. (2022). Forest through the Trees: Building Cross-Sections of Stock Returns. Journal of Finance, 77(4), 2093–2136. https://doi.org/10.1111/jofi.13129

Charupat, N., & Miu, P. (2013). The Pricing and Performance of Leveraged Exchange-Traded Funds. Journal of Banking & Finance, 37(12), 4763–4767. https://doi.org/10.1016/j.jbankfin.2013.08.014

Chong, J., & Phillips, G. (2015). Sector rotation with macroeconomic factors. Journal of Asset Management, 16(1), 14–26. https://doi.org/10.1057/jam.2014.16

Cohen, R. B., & Polk, C. (1998). The Impact of Industry Factors in Asset-Pricing Tests. Working Paper, Harvard Business School.

Gastineau, G. L. (2004). The benchmark index ETF performance problem. Journal of Portfolio Management, 30(2), 96–103. https://doi.org/10.3905/jpm.2004.319939

Gu, S., Kelly, B. T., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. Review of Financial Studies, 33(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Hegde, S. P., & McDermott, J. B. (2004). The liquidity of alternative market centers: Evidence from ETFs. Journal of Financial Markets, 7(3), 263–288. https://doi.org/10.1016/j.finmar.2004.01.002

Phua, P. K., Chung, Y. P., & Tjia, H. (2010). Sector rotation strategy based on macroeconomic indicators. Working Paper, Singapore Management University.

Stangl, J. T., Jacobsen, B., & Visaltanachoti, N. (2009). Sector Rotation over the Business Cycle. Working Paper, Massey University. Available at SSRN: https://ssrn.com/abstract=1396346

Wang, Y., & Xu, T. (2020). Does Sector Rotation Enhance Portfolio Performance? Journal of Portfolio Management, 46(5), 56–70. https://doi.org/10.3905/jpm.2020.1.188

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. Journal of Finance, 61(4), 1645–1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

Bali, T. G., Engle, R. F., & Murray, S. (2020). Factor models, machine learning, and asset pricing. Journal of Econometrics, 217(2), 273–297. https://doi.org/10.1016/j.jeconom.2020.05.008

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? Review of Financial Studies, 21(4), 1509–1531. https://doi.org/10.1093/rfs/hhm055

Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. Journal of Financial Economics, 25(1), 23–49. https://doi.org/10.1016/0304-405X(89)90095-0

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

Goyal, A., & Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies, 21(4), 1455–1508. https://doi.org/10.1093/rfs/hhn005

Kelly, B., & Pruitt, S. (2013). Market timing around the world. Journal of Financial Economics, 111(1), 137–153. https://doi.org/10.1016/j.jfineco.2013.01.006

Kothari, S. P., Shanken, J., & Sloan, R. G. (1995). Another look at the cross-section of expected stock returns. Journal of Finance, 50(1), 185–224. https://doi.org/10.1111/j.1540-6261.1995.tb05166.x

Moskowitz, T. J., & Grinblatt, M. (1999). Do industries explain momentum? Journal of Finance, 54(4), 1249–1290. https://doi.org/10.1111/0022-1082.00146

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. Management Science, 60(7), 1772–1791. https://doi.org/10.1287/mnsc.2013.1828

Nguyen, T., Phan, H., & Wang, L. (2020). Machine learning for active portfolio management. Quantitative Finance, 20(12), 2089–2104. https://doi.org/10.1080/14697688.2020.1731827

Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. Journal of Financial Economics, 104(2), 288–302. https://doi.org/10.1016/j.jfineco.2011.12.004

Zhang, L. (2005). The value premium. Journal of Finance, 60(1), 67–103. https://doi.org/10.1111/j.1540-6261.2005.00726.x

Gastineau, G. L. (2004). The benchmark index ETF performance problem. Journal of Portfolio Management, 30(2), 96–103. https://doi.org/10.3905/jpm.2004.319939

Hegde, S. P., & McDermott, J. B. (2004). The liquidity of alternative market centers: Evidence from ETFs. Journal of Financial Markets, 7(3), 263–288. https://doi.org/10.1016/j.finmar.2004.01.002

Phua, P. K., Chung, Y. P., & Tjia, H. (2010). Sector rotation strategy based on macroeconomic indicators. Working Paper, Singapore Management University.

Stangl, J. T., Jacobsen, B., & Visaltanachoti, N. (2009). Sector rotation over the business cycle. Working Paper, Massey University. Available at SSRN: https://ssrn.com/abstract=1396346

Wang, Y., & Xu, T. (2020). Does sector rotation enhance portfolio performance? Journal of Portfolio Management, 46(5), 56–70. https://doi.org/10.3905/jpm.2020.1.188