Aalto-yliopisto
Aalto-universitetet
Aalto University

Master's Programme in Finance

Enhancing Factor Timing with Machine Learning: A Decision Tree Approach

Oscar Lindholm & Elmeri Perkiö

Master's thesis
2025

Aalto-yliopisto
Aalto-universitetet
Aalto University

Author Oscar Lindholm & Elmeri Perkiö

Title of thesis Enhancing Factor Timing with Machine Learning: A Decision Tree Approach

Programme Master of Science in Economics and Business Administration

Major Finance

Thesis supervisor Markku Kaustia

Date 26.5.2025 **Number of pages** 62+3 **Language** English

Abstract

In this Master's thesis, we investigate the potential to enhance investment returns by dynamically timing equity factor premia in the U.S. market. Addressing a critical gap in existing factor investing literature, this study examines whether it is possible to successfully time equity factor premia using machine learning approaches. Specifically, we assess the effectiveness of Random Forest and Gradient Boosting models in forecasting the relative performance of equity style factors including size, value, profitability, and investment. Our approach employs a monthly forecasting framework that combines key macroeconomic indicators such as inflation, yield spreads, economic activity index, and market volatility with factor momentum signals. The predictive models are trained using a rolling window of the most recent five years of data, allowing the relationships between predictors and factor returns to adapt dynamically to changing market conditions.

Empirical analysis over the out-of-sample period from 2000 to 2024 demonstrates that these machine learning-driven strategies improve predictive accuracy compared to random factor selection and achieve excess returns relative to static allocation strategy. The best-performing Random Forest model delivers annualized returns and Sharpe ratios substantially higher than equal-weighted factor strategy, while Gradient Boosting models also show strong results, broadening the evidence for the potential of predictive modeling in factor investing.

In conclusion, while accurately timing factor returns remains challenging, our research demonstrates that data-driven factor timing models can achieve excess returns over static strategies. This study not only advances the literature on dynamic factor investing, but also offers practical tools for investors seeking to enhance portfolio performance with machine learning-based allocation strategies.

Keywords Factor Investing, Factor Timing, Dynamic Allocation, Machine Learning, Macroeconomic Conditions, Predictive Modeling

Aalto-yliopisto
Aalto-universitetet
Aalto University

Tekijä Oscar Lindholm & Elmeri Perkiö

Työn nimi Faktoreiden Ajoituksen Tehostaminen Koneoppimisen Avulla: Päätöspuulähestymistapa

Koulutusohjelma Kauppatieteiden Maisteri

Pääaine Rahoitus

Vastuuopettaja/valvoja Markku Kaustia

Päivämäärä 26.5.2025 **Sivumäärä** 62+3 **Kieli** Englanti

Tiivistelmä

Tässä tutkielmassa pyritään parantamaan sijoitustuottoja ajoittamalla osakefaktoripreemioita dynaamisesti Yhdysvaltain markkinoilla. Tutkimus tarkastelee verrattain vähän tutkittua aihetta faktoristrategioiden kirjallisuudessa selvittämällä, onko osakefaktoripreemioiden ajoittaminen mahdollista koneoppimismenetelmien avulla. Erityisesti arvioimme Random Forest- ja Gradient Boosting -mallien tehokkuutta ennustaa faktorien, kuten koon, arvon, kannattavuuden ja investointien, suhteellista kehitystä. Lähestymistapamme perustuu kuukausittaiseen ennustejärjestelmään, jossa yhdistyvät keskeiset makrotaloudelliset indikaattorit, kuten inflaatio, korkoerot, talousaktiivisuusindeksi ja markkinavolatiliteetti, sekä faktoreiden momenttisignaalit. Ennustemallit koulutetaan liikkuvalla viiden viimeisimmän vuoden tietojoukolla, mikä mahdollistaa sen, että selittäjien ja faktoreiden tuottojen väliset suhteet mukautuvat dynaamisesti muuttuviin markkinaolosuhteisiin.

Empiirinen analyysi vuosilta 2000–2024 osoittaa, että nämä koneoppimiseen perustuvat strategiat parantavat ennustetarkkuutta verrattuna satunnaiseen faktorivalintaan ja saavuttavat ylituottoa staattisiin allokaatioihin verrattuna. Parhaiten suoriutunut Random Forest -malli tuottaa vuosituottoja ja Sharpen lukuja, jotka ovat selvästi korkeampia kuin tasapainotetussa faktoristrategiassa. Myös Gradient Boosting -mallit osoittavat vahvaa suorituskykyä, mikä laajentaa empiiristä näyttöä ennustemallien potentiaalista faktoripohjaisessa sijoittamisessa.

Vaikka faktorituottojen ajoittaminen on vaikeaa, tutkimuksemme osoittaa, että datalähtöisillä faktorien ajoitusmalleilla voidaan saavuttaa ylituottoa staattisiin strategioihin verrattuna. Tämä tutkielma syventää dynaamisten faktoristrategioiden kirjallisuutta ja tarjoaa käytännön työkaluja sijoittajille, jotka pyrkivät parantamaan salkkujensa tuottoja koneoppimiseen perustuvien allokaatioratkaisujen avulla.

Avainsanat Faktorisijoittaminen, Faktorien Ajoittaminen, Dynaaminen Allokaatio, Koneoppiminen, Makrotaloudelliset Olosuhteet, Ennustava Mallinnus

Table of Contents

1. Introduction

Over the past few decades, factor investing has emerged as a prominent approach in both financial research and investment practice. This approach systematically targets exposure to quantifiable investment attributes, commonly known as factors (such as value, size, and quality), which have been empirically linked to superior risk-adjusted returns. Foundational studies demonstrated that portfolios tilted toward these factors can outperform broad market benchmarks, solidifying the academic case for factor-based strategies (Fama & French, 1993; Carhart, 1997; Asness et al., 2013). In tandem with this growing body of evidence, factor investing has gained widespread adoption: asset managers now routinely implement factor-focused strategies to enhance portfolio diversification, manage risk, and seek incremental returns. Indeed, factor investing (sometimes branded as "smart beta") is increasingly viewed as a cornerstone of modern asset management, blending elements of passive indexation with active management techniques in the pursuit of improved long-term performance.

However, the popularity of factor investing has also led to an overwhelming surge in proposed factors, a phenomenon commonly referred to as the "factor zoo." This proliferation has sparked considerable debate over the relevance and reliability of many newly proposed factors. To address this complexity, this thesis deliberately focuses only on the most widely recognized and economically grounded factors.

While the successes of static factor strategies are widely recognized, a critical question persists: Can investment performance be further improved by dynamically adjusting factor exposures in anticipation of changing market conditions? Both academics and practitioners continue to debate the merits of attempting to time factor returns, overweighting or underweighting certain factors based on economic indicators or market signals to enhance returns.

Perspectives on the viability of factor timing remain polarized. Skeptics argue that factor timing is inherently challenging, emphasizing that straightforward diversification across multiple factors typically outperforms timing attempts. They highlight that attempts to tactically overweight or underweight specific factors inherently reduce diversification, leading to concentrated exposures and higher vulnerability to forecast errors, in addition to increased transaction costs from frequent portfolio adjustments.

Conversely, optimists acknowledge these challenges but argue that investors who understand the economic and market drivers of factor performance can still achieve superior results through informed, selective factor timing.

Responding to this gap in the literature, this thesis empirically investigates whether a dynamic factor-timing approach can meaningfully enhance equity portfolio performance. The research employs two machine learning techniques, Gradient Boosting and Random Forest algorithms, to predict factor returns. Consistent with methodologies employed by Ilmanen et al. (2014), we condition these factor returns on selected macroeconomic indicators known to influence risk premia and expected returns. Specifically, we incorporate three macroeconomic variables: the Consumer Price Index percentage change (CPI%), reflecting inflationary pressures; the spread between the 10-Year Treasury yield and the 3-Month Treasury yield (T10Y3M), reflecting yield curve slope and economic expectations; and the Chicago Fed National Activity Index (CFNAI), summarizing economic activity relative to historical trends.

To further enhance predictive accuracy, market volatility is incorporated through a GARCH (1,1) model, providing an effective measure of volatility clustering commonly observed in financial markets. Recognizing the importance of momentum, we incorporate factor-specific momentum features following Moskowitz et al. (2012), calculating moving average returns from months t-12 to t-2 for each factor to capture persistent momentum trends and avoid short-term reversal effects. Consistent with Ehsani and Linnainmaa (2022), momentum is employed exclusively as a timing mechanism aligned with historical trends and macroeconomic conditions, rather than as a standalone factor.

To empirically evaluate the dynamic factor timing approach, this study employs three distinct datasets: classical Fama-French long-short portfolios and constructed long-only factor portfolios (both beginning in April 1967), as well as investable MSCI factor indices (starting July 1995). Predictive models are trained using a rolling-window approach, applying a threshold of at least five years of historical data. Machine learning hyperparameters are optimized exclusively on data available before January 2000. Consequently, all reported performance metrics from January 2000 to November 2024 represent genuine out-of-sample predictive accuracy across various market conditions and portfolio types.

The results demonstrate that these dynamic factor allocation strategies generally outperform equal-weight equity factor benchmarks from January 2000 to November

2024, with notable Sharpe ratio improvements in the MSCI dataset (Random Forest: 0.71, Gradient Boosting: 0.64 vs. Equal Weight: 0.62) and modest gains in long-short (Random Forest: 0.64, Gradient Boosting: 0.60 vs. Equal Weight: 0.54) and long-only datasets (Random Forest: 0.66, Gradient Boosting: 0.64 vs. Equal Weight: 0.63). Investors could therefore consider exploring these types of strategies within asset allocation decisions, adjusting exposures based on macroeconomic and market indicators.

The remainder of this thesis is structured as follows. Section 2 provides a thorough review of theoretical foundations and empirical literature on factor investing and market timing, pinpointing the research gap. Section 3 describes the data and methodology employed, including predictive variables and modeling techniques. Section 4 presents empirical results from dynamic factor allocation experiments. Finally, section 5 concludes with key insights, implications for investment practice, and recommendations for future research.

2. Literature Review

2.1. Evolution of Factor Premia

The foundations of modern asset pricing theory were established by contributions linking risk and return through optimal portfolio selection. Markowitz (1952) developed mean–variance theory, formalizing that investors seek portfolios maximizing expected returns for a given risk level through diversification. Building upon Markowitz, Sharpe (1964) and Lintner (1965) introduced the Capital Asset Pricing Model (CAPM), a single-factor model connecting an asset's expected return directly to market risk. Despite its conceptual elegance, CAPM drew criticism due to restrictive assumptions and inability to explain observed return anomalies.

In response to these limitations, Ross (1976) introduced the Arbitrage Pricing Theory (APT), allowing multiple sources of systematic risk beyond market risk alone. According to APT, if an asset's price deviates from implied multi-factor values, investors exploit mispricing through zero-net-investment portfolios, pushing prices toward equilibrium. This insight expanded the theoretical framework, motivating empirical exploration of additional risk premia.

Empirical research in the 1980s identified firm characteristics associated with abnormal returns. Stattman (1980) and Rosenberg et al. (1985) found high book-to-market ("value") stocks earned superior returns, while Banz (1981) observed smaller firms outperform larger ones. Bhandari (1988) documented a positive relationship between leverage and returns. These findings laid the foundation for Fama and French's (1992, 1993) three-factor model, incorporating size and value factors. Variables like earnings-to-price and leverage appeared significant in earlier studies, but effects diminished once size and value were included.

Yet even the Fama–French three-factor model left certain patterns unexplained. A prominent anomaly is momentum: stocks that performed well in the recent past tend to continue outperforming (Jegadeesh & Titman, 1993). Carhart (1997) showed adding momentum significantly improved explanatory power. This four-factor model demonstrated momentum represents either an extra risk factor or persistent behavioral biases.

Researchers have also emphasized liquidity as critical. Amihud (2002) demonstrated less liquid stocks, characterized by higher price impacts per unit traded, offer higher expected returns as compensation for trading frictions. Pastor and Stambaugh (2003) introduced a market-wide liquidity factor, demonstrating stocks sensitive to aggregate liquidity shocks earn additional returns as illiquidity compensation.

Fama and French later expanded their model differently. Their five-factor model (Fama & French, 2015) augmented market, size, and value with profitability (Novy-Marx, 2013) and investment intensity. This extension improved explanatory power but raised questions about redundancy—such as whether new factors subsume the value effect—and omission of momentum, which Carhart's work showed to be important. In response, Fama and French (2018) introduced a six-factor model explicitly including momentum alongside market, size, value, profitability, and investment. Challenges defining and measuring factors initially kept momentum and liquidity out of their models despite empirical support.

Still, there is no consensus regarding which factors represent fundamental risk versus artifacts of data mining. The abundance of proposed factors—the "factor zoo"—implies many empirical patterns may fail out-of-sample or lack economic substance. Alternative frameworks challenge the Fama and French paradigm; Hou, Xue, and Zhang's q-factor model (2019), using investment and profitability factors, explains anomalies without explicitly including value. This ongoing debate highlights the dynamic evolution of asset pricing theory as scholars seek economically meaningful risk premia rather than spurious correlations. It also poses a practical challenge: given many potential factors, how should investors select and effectively combine them?

2.2. Time Varying Factor Premia

Investors, faced with numerous sources of risk premia, have increasingly turned to multifactor strategies that simultaneously harvest multiple equity factors. A key advantage of combining several factors is diversification, as long–short factor portfolios often have low correlations with each other due to differences in underlying economic fundamentals or behavioral effects. By allocating across several uncorrelated factor premia, portfolios can achieve more robust and stable performance compared to traditional asset-class diversification strategies. Empirical evidence supports this insight: Ilmanen and Kizer (2012) showed that market-neutral portfolios diversified across multiple style factors attain substantially higher Sharpe ratios and lower volatility relative to conventional equity-bond portfolios. Similarly, Ilmanen et al. (2014) found that well-established equity factors—such as market, size, value, and low volatility—historically exhibit near-zero average correlations, underscoring their independence. Consequently, portfolios blending these factors have consistently delivered positive risk-adjusted returns across diverse market environments and macroeconomic conditions, highlighting their resilience and effectiveness in mitigating volatility and drawdowns.

Importantly, the benefits of multifactor strategies largely stem from the dynamic rather than static nature of risk premia. Classic studies established that expected returns systematically vary with changing economic and market conditions. Campbell and Shiller (1988) demonstrated that valuation indicators, such as earnings yields, have predictive power for future stock returns, indicating that expected equity returns fluctuate alongside fundamental economic shifts. Similarly, Fama and French (1989) documented that variables like dividend yields and default credit spreads forecast market returns, typically rising during periods of economic weakness or financial strain and declining in stronger economic environments. These insights highlight the importance of recognizing that equity premia adjust over time, rather than remaining constant.

Extending these findings to factor premia specifically, a growing body of research emphasizes that equity factors themselves exhibit distinct cyclical performance patterns closely linked to macroeconomic conditions (Guidolin & Timmermann, 2008; Ang & Timmermann, 2012; Ilmanen et al., 2014). Studies demonstrate that distinct macroeconomic regimes—characterized by specific growth, volatility, and correlation profiles—systematically influence the behavior of size, value, momentum, quality, and other factors. Each factor responds uniquely to macroeconomic drivers such as economic growth, inflation, interest rates, market volatility, and liquidity conditions. This heterogeneous factor sensitivity means that at any given moment, certain factors may experience premium expansion while others contract, allowing a multifactor portfolio to better withstand individual macroeconomic shocks. Consequently, diversification across multiple factors inherently provides smoother returns and enhanced stability, particularly during turbulent economic periods.

Investor behavior and factor popularity represent another key dimension influencing time-varying factor returns. Evidence from Lou et al. (2019) indicates that factor premia inversely relate to investor attention and fund flows into those strategies. Factor portfolios typically deliver higher returns when undervalued and overlooked, and lower returns when they become popular and expensive. While this contrarian approach aligns with the principle of "buying low and selling high," practical implementation remains challenging.

Collectively, these insights highlight the nuanced nature of multifactor investing. Thoughtfully combining factors yields clear diversification benefits, but the existence of cyclicality in factor premia suggests that investors might achieve even better results by adjusting factor exposures in line with prevailing conditions. Rather than adhering strictly to fixed factor weights, there may be significant value in tilting toward factors expected to outperform under specific macroeconomic or market environments, motivating the exploration of factor timing strategies.

2.3. Timing the Factor Premia

Given the cyclical behavior of factors, a natural question is whether investors can improve performance by tactically rotating among factor exposures, overweighting factors before they outperform and underweighting them before they underperform. However, attempting to time markets or factors remains one of the most challenging and debated topics in finance. While the potential to outperform static buy-and-hold strategies is appealing, successfully forecasting shifts in factor returns is notoriously difficult due to the complexity and dynamic nature of underlying market forces. Empirical evidence shows that predictors which seem reliable in one sample often fail in another as market regimes change or as investors arbitrage away past patterns. Indeed, many variables that appear statistically significant in-sample lose their predictive power out-of-sample or in later periods, underscoring the instability of return predictability (Goyal & Welch, 2008; Goyal et al., 2022). In other words, a timing model that works well in one economic context might falter once conditions change.

A seminal contribution highlighting these challenges was made by Goyal and Welch (2008), who conducted a comprehensive evaluation of equity premium predictors proposed in prior literature. They examined 29 different variables (ranging from valuation ratios to macroeconomic indicators), comparing their out-of-sample forecasting performance against a simple historical average benchmark. The sobering result was that, with few exceptions, most predictors failed to consistently outperform the naive benchmark once tested out-of-sample. Even variables with strong in-sample correlations and plausible economic stories lost potency over extended periods or different market regimes. An updated analysis by Goyal et al. (2022), incorporating data through 2021, reinforced these conclusions: most proposed predictive variables continue to exhibit considerable instability, and only a small subset (such as certain yield measures or trend signals) show any persistent forecasting ability. Together, these studies drive home the difficulty of accurately predicting even the aggregate stock market premium. They suggest that any successful timing approach must be adaptive and robust to regime changes, rather than reliant on a fixed historical relationship.

Notably, while extensive research has examined market timing for the equity premium, comparatively less attention has been devoted specifically to timing other equity factor premia. Although the regime-dependence and macroeconomic sensitivities of factor returns are well-established, systematic approaches for dynamically rotating among factors remain relatively uncommon in the academic literature. One reason for this gap may be uncertainty about which predictors are most relevant for different factors. Unlike broad equity market timing—where a relatively common set of indicators (e.g. the dividend yield, yield curve slope, default spread) has been studied—factor timing involves a larger number of potential signals and no universally agreed framework. Additionally, as Ilmanen (2014) points out, tactical factor allocation poses a dual forecasting challenge: investors must anticipate the direction of macroeconomic conditions and understand how individual factors are likely to respond. Accurately interpreting both the economic environment and factor sensitivities makes effective factor timing especially complex.

Despite these challenges, the literature suggests that incorporating economic fundamentals and regime information can improve timing performance, as opposed to relying purely on extrapolation of past returns. Rather than chasing short-lived statistical patterns, grounding timing decisions in broad economic indicators or theoretical relationships may yield more stable and economically intuitive strategies. In practice, this means using variables that reflect the state of the economy and financial markets—such as growth trends, inflation regimes, interest rate term structures, or measures of market stress—as inputs to any factor timing model. Below, we review several key macroeconomic and market indicators that have been explored as signals for timing factor premia.

Among macroeconomic indicators, **economic growth** is widely recognized as a fundamental driver shaping the performance of equity factor strategies. Both theoretical and empirical evidence suggests that factor returns systematically vary with changes in the underlying economic conditions. Intuitively, during periods of slowing economic growth or rising uncertainty, investors become more risk-averse, demanding higher returns as compensation for bearing increased risks (Fama & French, 1989). Economic slowdowns or recessions significantly influence asset prices by altering investors' perceptions of risk and required returns (Guidolin & Timmermann, 2007). Defensive factors, such as quality and profitability, tend to outperform in these low-growth or recessionary periods due to their stability and lower sensitivity to macroeconomic deterioration. Conversely, procyclical factors, such as size and value, typically demonstrate superior performance during robust economic growth phases, as these periods favor firms more sensitive to economic expansion (Liew & Vassalou, 2000; Vassalou, 2003).

Empirical studies provide consistent evidence supporting this relationship between economic growth and factor returns. Early research by Ferson and Harvey (1991) highlighted significant differences in factor performance across economic expansions and recessions. Hodges et al. (2017) further documented that economically sensitive factors, such as size and value, frequently underperformed during U.S. recessions due to deteriorating fundamentals and limited financial flexibility. Conversely, defensive strategies, notably quality and low-volatility, offered meaningful outperformance during downturns, reflecting their effectiveness as safer investment options. During economic recoveries and expansions, procyclical factors such as momentum and value typically rebounded strongly, benefiting from improving corporate earnings expectations and rising investor risk appetite (Hodges et al., 2017).

Several empirical studies suggest that economic growth indicators can reliably predict the relative attractiveness of different factor exposures. Liew and Vassalou (2000) and Vassalou (2003) demonstrated a robust connection between macroeconomic growth prospects and subsequent returns of size (SMB) and value (HML) factors. Their analyses indicated that these factors systematically delivered higher returns when future economic growth was expected to be strong, implying that investors might effectively anticipate shifts in factor premia by closely monitoring economic conditions. Practically, indicators such as GDP growth rates, industrial production indices, employment trends, or leading economic indicators can provide actionable signals for dynamic factor allocation. For example, deteriorating economic conditions could trigger a tactical shift toward defensive factors, whereas improving economic indicators might justify an increased allocation to procyclical factors such as value, size, or momentum Polk et al. (2020)

Recent research has provided concrete examples of frameworks leveraging economic growth signals for factor timing. Polk et al. (2020) developed a dynamic strategy that identified business cycle regimes in real-time using composite indicators, such as manufacturing surveys and new orders data. Their strategy systematically overweighted growth-sensitive factors during economic expansions and defensively positioned factors ahead of downturns, generating substantially higher risk-adjusted returns compared to a static factor allocation. Similarly, Hodges et al. (2017) incorporated explicit economic regime classifications—expansion, slowdown, recession, recovery—into a factor timing framework, finding this regime-based approach yielded superior Sharpe ratios relative to other predictive indicators studied.

Closely intertwined with economic growth, **inflation** represents another critical macroeconomic variable influencing factor returns. Inflation impacts equity returns primarily by eroding real cash flows and prompting adjustments in interest rates, thus altering equity valuations. Higher inflation frequently triggers tighter monetary policy, leading to increased discount rates, which in turn depress equity valuations and slow real earnings growth. For instance, Fama and Schwert (1977) identified a consistent negative relationship between U.S. stock returns and both expected and unexpected inflation, indicating that equities typically underperform when inflation accelerates. This phenomenon may partially arise from investor misperceptions, known as inflation illusion, whereby investors initially underestimate the negative impact of inflation on real returns. Such misperceptions can lead to stocks becoming temporarily overvalued in real terms, followed by subsequent corrections once investors recognize the true impact of rising prices (Campbell & Vuolteenaho, 2004).

Within factor investing, inflation's predictive influence typically operates indirectly through its connections with the business cycle and interest rates, but its impact remains significant. Historically, spikes in inflation often coincide with late-stage economic overheating or unexpected supply shocks, subsequently prompting economic slowdowns or recessions as central banks aggressively raise interest rates to curb inflationary pressures. Chen (2009) provided empirical evidence that inflation indicators effectively forecast bear markets, demonstrating that substantial increases in both expected and realized inflation significantly raise the probability of future equity market downturns, even after accounting for other macroeconomic factors.

For investors employing factor timing strategies, monitoring inflation conditions can be particularly valuable. Periods of accelerating or unexpectedly high inflation generally correspond to lower equity returns, negatively affecting procyclical factors such as value and size, which tend to underperform along with the broader market during inflationary episodes. Investors witnessing sustained increases in inflation may thus consider shifting toward assets more resilient to inflation, such as commodities or inflation-linked securities, or toward defensive equity factors that typically withstand rising interest rates better. Conversely, environments of stable or declining inflation generally favor risk assets. As inflation pressures diminish, interest rates often decrease, and investor confidence typically improves, providing support to growth-oriented or cyclical factor strategies.

Practically, distinguishing broadly between low, moderate, and high inflation regimes can meaningfully enhance factor timing approaches. Investors can proactively adjust their factor exposures based on inflationary environments, helping to preserve real returns. For instance, if inflation enters an upward trend, a factor-timing strategy might overweight sectors or factors characterized by shorter-duration cash flows—less vulnerable to rising interest rates—and reduce exposure to growth-focused factors, which often carry longer-duration cash flows and are typically more adversely affected by inflation increases. While inflation alone does not wholly determine factor premia, it acts as a valuable supporting signal, highlighting broader macroeconomic shifts that influence factor performance across various market conditions.

Yield spreads, defined as differences between interest rates of varying maturities or credit qualities, aggregate market expectations about economic growth, inflation, and risk. This makes them effective forward-looking signals for factor returns. Two types of yield spreads are particularly informative: the term spread, measured as the difference between long-term and short-term government bond yields, and the credit spread, defined as the difference between yields on corporate bonds and comparable Treasury securities. These measures have a long-established role in forecasting market returns, and recent evidence suggests they also capture risks associated with certain factor premia. For example, Hahn and Lee (2006) found that yield spreads effectively capture variations in returns traditionally attributed to size (SMB) and value (HML) factors, suggesting that part of these factor premia compensates investors for exposure to yield-curve and credit risks. Intuitively, a flattening or inverted yield curve signals expectations of economic slowdowns or recessions. Historically, value and small-cap factors have tended to underperform in these periods, reflecting reduced investor appetite for risk. Similarly, widening credit spreads indicate deteriorating financial conditions and higher default risks, disproportionately harming economically sensitive companies linked to the size factor.

Because yield spreads encapsulate collective market expectations about growth and risk often well before traditional economic data signals emerge, they have proven particularly useful in factor timing models. Chen (2009) highlighted the predictive power of yield spreads, identifying the slope of the yield curve as one of the most reliable leading indicators of bear markets. Specifically, an inverted term spread is a well-known early indicator of recessions and subsequent market downturns. Investors can leverage these signals to adjust factor exposures tactically; when the yield curve inverts, a prudent response would involve reducing exposure to procyclical factors such as value, size, and momentum, and increasing exposure to defensive factors like quality and low volatility, which tend to perform relatively better during downturns. Conversely, a steepening yield curve typically signals improving economic prospects, a scenario historically beneficial for procyclical factors. Similarly, narrowing credit spreads indicate improving sentiment, potentially justifying increased allocations toward higher-risk, cyclical factors. Thus, incorporating yield spreads into factor timing strategies can improve forecasts of factor returns by identifying economic regime shifts that purely valuation-based indicators might miss. Given their documented relationship with the business cycle, yield spreads can serve as useful indicators within factor timing approaches, potentially providing early signals of shifts in factor performance.

Market volatility is an essential indicator of investor sentiment and market stability, carrying significant implications for factor returns. Periods of low volatility typically reflect calm market conditions and greater investor confidence, whereas high volatility signals distress, heightened uncertainty, and increased investor risk aversion. Early work by Turner et al. (1989) established that equity markets fluctuate distinctly between these volatility regimes, with subsequent studies, such as Ang and Timmermann (2012), emphasizing that understanding volatility-driven regime shifts is crucial for explaining patterns in factor performance.

Factor premia often behave very differently depending on the prevailing volatility environment. In periods of heightened volatility, typically occurring during market downturns or crises, defensive factors such as quality and low-volatility strategies generally outperform more aggressive factors like momentum or high-beta portfolios. Conversely, during stable, lower-volatility market conditions, pro-cyclical and higher-risk factors usually deliver superior returns. These observed patterns imply that volatility is a valuable timing signal for dynamically adjusting factor exposures according to the current market environment.

Empirical evidence consistently supports volatility-based factor timing as an effective risk management strategy. Fleming et al. (2001) demonstrated significant improvements in risk-adjusted returns from portfolios that reduced exposure during volatile periods and increased exposure during calm conditions. Extending this concept to factors, Moreira and Muir (2017) showed that strategies scaling exposure inversely with recent realized volatility significantly boosted the risk-adjusted performance of common factors like value, momentum, and profitability. In practice, investors can utilize volatility indicators such as the VIX index or recent realized volatility of factor returns to systematically modulate factor exposures, reducing drawdowns in turbulent markets and enhancing overall portfolio stability.

Complementing macroeconomic and volatility-based signals, **momentum** serves as a significant technical indicator for factor timing. Momentum refers to the tendency of assets with recent strong performance to continue outperforming in the near term, a phenomenon well documented not only in individual stocks but also at the factor level. Lewellen (2002) provided early evidence that major equity factors exhibit momentum. Specifically, portfolios sorted by characteristics such as size or value not only possess long-term return premia but also demonstrate short-term momentum, meaning factors that performed well in recent periods tend to continue outperforming in subsequent months. This suggests momentum arises partly from common economic shocks or investor underreaction affecting groups of stocks simultaneously rather than solely from stock-specific news.

Asness et al. (2013) showed that value and momentum strategies are pervasive across asset classes globally. Notably, they identified a consistent negative correlation between the returns of value and momentum strategies across markets. This inverse relationship complicates multifactor investing because periods of strong momentum performance often coincide with weaker value returns, and vice versa. Recent studies differentiate between time-series momentum, where a factor's past returns predict its own future performance, and cross-sectional momentum, reflecting relative performance across multiple factors. Ehsani and Linnainmaa (2022) provided robust evidence indicating that much of what appears as momentum is attributable to the time-series momentum of individual factors. Practically, this implies that investors could dynamically overweight factors demonstrating recent strong performance and underweight those with weaker recent trends. While momentum-based factor rotation has been shown to enhance performance, careful implementation is required due to risks associated with trend reversals and the costs of higher portfolio turnover.

However, momentum strategies have important limitations. Bender et al. (2018) and Hodges et al. (2017) cautioned that momentum signals are highly sensitive to regime changes and the length of estimation periods. The strength of factor autocorrelations diminishes after approximately one year, underscoring the need for precise timing and the integration of complementary predictive signals. Hodges et al. (2017) suggested enhancing momentum strategies by incorporating additional indicators, such as valuation spreads, macroeconomic outlooks, and factor volatility. Similarly, Barroso and Santa-Clara (2015) proposed a risk-managed momentum approach, highlighting that dynamically scaling positions based on recent volatility significantly improves risk-adjusted returns by mitigating severe drawdowns typical in momentum investing.

In summary, momentum can effectively enhance factor timing by exploiting persistence in recent factor returns. Nevertheless, due to the inherent risks of trend reversals and sensitivity to changing market conditions, momentum strategies are typically integrated with broader frameworks that incorporate macroeconomic indicators and volatility signals. Approaches that combine multiple predictive signals have gained attention, as they can mitigate individual signal limitations and enhance the robustness of factor-timing models. To effectively handle the complexity of combining these signals, advanced analytical methods capable of detecting nonlinear relationships and subtle interactions among predictors are increasingly explored, highlighting the potential role of machine learning techniques discussed in the following section.

2.4. Machine Learning Methods for Factor Timing

Recent advances in machine learning offer promising tools for enhancing factor timing by addressing the inherent nonlinearities and complex interactions involved in forecasting asset returns. Traditional linear econometric models often rely on restrictive assumptions and can therefore overlook critical predictive relationships within financial data. By contrast, machine learning methods, particularly Random Forests and Gradient Boosting, effectively capture subtle interactions and nonlinear patterns among various economic indicators, enabling more robust predictive performance. Empirical research supports these claims: Gu et al. (2020) demonstrate that tree-based methods substantially outperform linear regressions in out-of-sample return predictions due to their ability to automatically incorporate complex interactions without extensive manual intervention.

A significant advantage of Random Forests and Gradient Boosting is their flexibility in simultaneously handling a wide array of predictive indicators, including macroeconomic variables, valuation metrics, and momentum signals. This capability allows investors to systematically extract actionable insights from diverse market information, facilitating the construction of more informed and responsive factor allocation strategies. Further evidence from Cakici et al. (2023) indicates that these methods consistently improve forecasting accuracy across varied market conditions, effectively isolating valuable predictive signals from noisy environments. Similarly, Kelly et al. (2021) find that machine learning-based asset pricing models deliver economically meaningful enhancements to investment performance, underscoring their practical applicability in quantitative asset management.

While the advantages of machine learning models are evident, their effective deployment in factor timing also requires careful implementation to mitigate certain well-known challenges. Overfitting remains a primary concern, especially in financial contexts characterized by limited historical data and evolving market dynamics. Arnott et al. (2019) highlight that such challenges are longstanding in quantitative finance, emphasizing the importance of rigorous validation strategies, including out-of-sample testing, cross-validation, and feature selection, to enhance predictive reliability.

Interpretability of model predictions also remains crucial, particularly for ensuring investor confidence and transparency. Although machine learning methods have a reputation for opacity, tree-based models such as Random Forests and Gradient Boosting offer a notable advantage in interpretability through mechanisms like variable importance metrics and partial dependence analyses. These tools provide investors with valuable insights into which specific signals influence model decisions, aligning quantitative predictions with economic rationale Freyberger et al. (2020)

Financial markets evolve over time, altering the relationships between indicators and returns. Campbell and Thompson (2008) and Rapach et al. (2010) stress that forecasting models must adapt accordingly. Rolling-window retraining and adaptive updating allow machine learning models to preserve accuracy and interpretability under changing conditions. Empirical evidence in market timing, risk modelling, and portfolio construction shows that these techniques capture nonlinear signal interactions and offer improvements relative to conventional methods.

3. Data and Forecasting Strategies

This chapter lays the essential groundwork for our dynamic factor-allocation analysis in Chapter 4 by first describing the construction of our return and predictor datasets, then documenting the empirical patterns that motivate conditional timing, and finally presenting the modelling framework that converts these patterns into actionable signals.

In Section 3.1, we detail our data sources and constructions. On the return side, we employ the traditional long-short Fama-French five factors, as well as constructed long-only factor portfolios and investable MSCI proxies, to capture size, value, profitability, and investment premia. On the predictor side, we compile a set of macroeconomic and technical features—year-over-year CPI% changes, the 10-Year minus 3-Month Treasury yield spread, the Chicago Fed National Activity Index, conditional volatility from a GARCH (1,1) model, and 12-month factor momentum—each carefully lagged to reflect real-time data availability.

Next, in Section 3.2, we replicate and extend the empirical approach of Ilmanen (2014) to illustrate the characteristics of our data and clearly demonstrate the time-varying nature of factor premia, thus laying essential groundwork for our main empirical analysis. Specifically, we confirm that factor premia systematically vary across inflationary, yield curve, economic growth, and volatility regimes. These findings reinforce Ilmanen's insight that no single factor consistently outperforms, highlighting the potential to exploit regime-dependent factor behavior through dynamic allocation strategies.

Finally, Section 3.3 introduces our predictive modelling framework, describing a systematic yet adaptive approach designed to capture time-varying factor premia through dynamic allocations. We detail our methodology of fitting decision-tree ensemble models, specifically Random Forest and Gradient Boosting, utilizing a rolling five-year training window to flexibly model nonlinear interactions without restrictive distributional assumptions. These models identify predictive relationships between historical feature values and subsequent factor outperformance. The resulting predicted probabilities form the basis for dynamic factor allocations, the performance and implications of which are analyzed and presented in Chapter 4.

3.1. Data Selection

3.1.1. Return Data

To effectively examine how factor returns vary across different macroeconomic environments, we primarily utilize the Fama-French 5-factor model due to its extensive empirical support in capturing cross-sectional variations in stock returns. Fama and French (2015) introduced this model to extend earlier factor models by incorporating profitability (RMW) and investment (CMA) factors alongside the traditional market, size (SMB), and value (HML) factors. This extension allows a deeper understanding of the underlying economic drivers affecting asset returns beyond the explanations provided by simpler factor structures. In our empirical analysis in Chapter 4 we utilize three distinct datasets, starting with the academically established long-short Fama–French five factors, then moving to custom long-only Fama–French portfolios that address practical leverage and implementation constraints, and finally incorporating investable long-only MSCI factor indices for real-world application. This staged approach provides a comprehensive view across different factor constructions and settings, and all three datasets are employed in our dynamic allocation framework.

First, we employ the Fama-French long-short factor methodology, constructing factors by taking long positions in portfolios of stocks with desirable financial characteristics and short positions in portfolios with less favorable characteristics within the U.S. equity market universe. Specifically, the size factor (SMB) captures differences between small-cap and large-cap stocks based on market capitalization, while the value factor (HML) contrasts high-value (high book-to-market) stocks against low-value counterparts. Additionally, the profitability factor (RMW) distinguishes firms with robust versus weak profitability, and the investment factor (CMA) differentiates between firms with conservative and aggressive investment policies. These factors are zero-cost portfolios designed to isolate the respective premia, with monthly data sourced from the Kenneth R. French Data Library.

Acknowledging practical constraints that some investors face, such as leverage restrictions and limitations on short-selling activities (Frazzini & Pedersen, 2014), we also constructed long-only versions of these Fama-French factors. These long-only factors were created using three sets of Fama-French-style portfolios: (1) six portfolios formed on Size and Book-to-Market (2×3), (2) six portfolios formed on Size and Operating Profitability (2×3), and (3) six portfolios formed on Size and Investment (2×3).

Specifically, for the Value (HML) factor, we formed portfolios consisting exclusively of small-value and big-value stocks (top 30% highest Book-to-Market). Similarly, for the Profitability (RMW) factor, we invested solely in small and big stocks with the highest operating profitability (top 30%), explicitly omitting low-profitability stocks. For the Investment (CMA) factor, we selected small and big stocks characterized by the lowest investment growth (most conservative firms), disregarding high-investment growth stocks. By carefully selecting these size-characteristic subportfolios, our long-only approach effectively isolates each factor's premium without the need for short positions, offering practical and clear insights into the performance and robustness of long-only investment strategies.

Constructing and regularly rebalancing long-only or long-short factor portfolios can entail significant operational complexity and trading costs, especially for individual investors and smaller institutions. To offer a more investable perspective, our third dataset comprises long-only MSCI factor indices, serving as readily accessible proxies for each Fama-French factor alongside our two Fama-French datasets in Chapter 4.

Specifically, we use the MSCI USA Value Index for HML, capturing large- and mid-cap securities with high book-to-price, forward earnings-to-price, and dividend yields;

the MSCI USA Quality Index for RMW, targeting firms with strong ROE, stable earnings growth, and low leverage; the MSCI USA Low Size Index for SMB, weighting stocks inversely by market cap; and the MSCI USA Minimum Volatility Index for CMA, emphasizing low-volatility names as a stand-in for conservative investment. While the latter is not a perfect match for CMA's traditional definition, it embodies a cautious investment tilt.

These MSCI indices are well-established, investable vehicles that reduce the need for bespoke portfolio construction and rebalancing, making factor strategies more practical and potentially lowering transaction burdens, while offering efficient access to factor exposures.

3.1.2. Feature Data

Consistent with methodologies employed by Ilmanen et al. (2014), we condition these factor returns on selected macroeconomic indicators that have been shown to influence risk premia and expected returns. Specifically, we selected three macroeconomic variables: the Consumer Price Index percentage change (CPI%), capturing inflation dynamics and purchasing power erosion; the spread between the 10-Year Treasury yield and the 3-Month Treasury yield (T10Y3M), which provides insights into the yield curve slope and future economic expectations; and the Chicago Fed National Activity Index (CFNAI), a composite indicator summarizing economic activity and serving as a proxy for economic growth by indicating expansions or recessions relative to historical trends. These indicators were chosen due to their demonstrated predictive power and relevance in explaining variation in factor premia across differing economic states, as evidenced by prior literature including Ilmanen et al. (2014).

Since some macroeconomic data are subject to release delays, we implemented appropriate lags to prevent look-ahead bias. Specifically, CPI data are typically released between the 10th and 15th day of the subsequent month, requiring a one-month lag. Similarly, the CFNAI data, released around the 22nd to 24th day of the following month, were also lagged by one month. Incorporating these lags ensures that our forecasts reflect only information that was realistically available at the time predictions were made, thus maintaining the integrity of the predictive framework.

Recognizing the importance of capturing momentum in factor returns, we incorporate a factor-specific momentum feature based on the methodology of Moskowitz et al. (2012), who demonstrate the economic significance of time series momentum across various asset classes. Specifically, we calculate factor-specific momentum for each factor (SMB, HML, RMW, CMA) by employing a moving average of returns from month t-12 to t-2, explicitly excluding the most recent month to mitigate short-term reversal effects. By utilizing this moving average approach based on fully lagged historical returns, we effectively capture persistent momentum trends, enhancing predictability and refining factor allocation decisions. Consistent with Ehsani and Linnainmaa (2022), we do not include momentum as a standalone factor; instead, we leverage momentum signals exclusively as a timing mechanism aligned with historical trends and current macroeconomic conditions.

To incorporate market volatility into our analysis, we employed a GARCH (1,1) model applied to daily returns of the market excess return (MKT-RF), sourced from the Kenneth R. French Data Library. The GARCH (1,1) model, introduced by Bollerslev (1986), effectively captures volatility clustering—a prominent characteristic in financial market data, where periods of high volatility tend to be followed by similarly volatile periods.

Formally, the GARCH (1,1) model specifies conditional variance as:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

where σ_t^2 represents the predicted variance at time t, ω denotes the long-term average variance, α captures the sensitivity to recent squared returns (shocks), and (β) measures volatility persistence. Parameters were estimated via maximum likelihood estimation using Python's `arch` library.

After fitting the GARCH model to daily market returns for each month, we extracted the conditional volatility estimated at the month's end, providing a forward-looking volatility prediction for the subsequent month. This approach allows us to evaluate how anticipated fluctuations in market uncertainty influence the behavior and pricing of the factors under consideration.

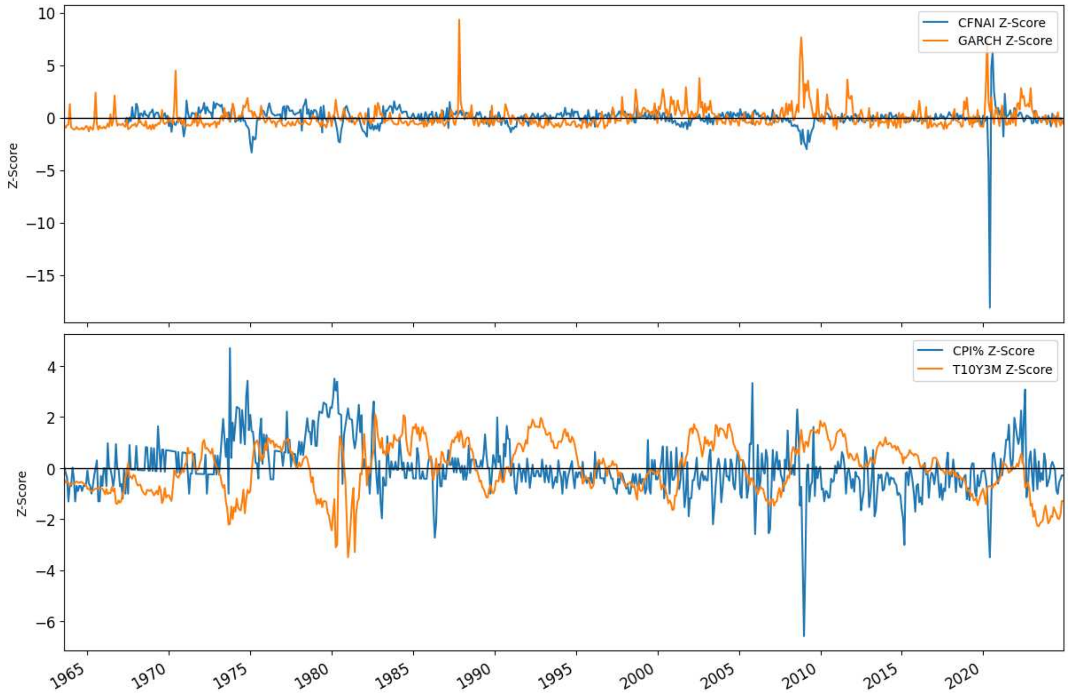
Macroeconomic indicators—CPI%, T10Y3M, and CFNAI—are sourced from the Federal Reserve Economic Data (FRED) database, maintained by the Federal Reserve Bank of St. Louis. Utilizing data from FRED ensures the consistency, credibility, and comparability of our results with existing literature and established research practices.

3.2. Time-Varying Style Premia

To conclusively demonstrate the time-varying nature of risk premia in our dataset, we replicate and extend the conditional factor analysis of Ilmanen et al.,(2014) using the classical long-short Fama-French 5-Factors. We start by visualizing key macroeconomic indicators known to influence factor performance: year-over-year CPI changes (CPI%), the 10-Year minus 3-Month Treasury yield spread (T10Y3M), the Chicago Fed National Activity Index (CFNAI), and market volatility estimated by a GARCH (1,1) model. Each indicator is standardized into monthly Z-scores, enabling clear identification of periods marked by macroeconomic stress or stability. Figure 1 illustrates these standardized indicators. Prominent episodes such as the 2008 financial crisis and the 2020 COVID-19 downturn are distinctly visible in the economic activity and volatility indicators, while significant inflationary episodes, notably in the late 1970s, early 1980s, and 2022, are captured by the CPI%. The yield curve spread (T10Y3M) clearly highlights shifts corresponding to major economic turning points, particularly around the 2008 crisis.

Figure 1 – Standardized Z-Scores of Macroeconomic Indicators

In figure 1, the upper panel depicts standardized monthly Z-scores of economic activity (CFNAI) and market volatility (GARCH (1,1)). The lower panel displays standardized Z-scores for year-over-year inflation (CPI%) and the yield-curve spread (T10Y3M). Extreme positive or negative values indicate heightened macroeconomic stress or relative stability.



By explicitly conditioning factor returns on distinct macroeconomic conditions—captured by CPI%, T10Y3M, CFNAI, and GARCH (1,1)—we can systematically identify and quantify these regime-specific behaviors. This structured approach illustrates how certain factors tend to deliver stronger performance in particular macroeconomic contexts, supporting the economic rationale for considering dynamic, regime-aware factor allocation strategies.

Table 1 – Factor Sharpe Ratios Across Macroeconomic Regimes

Table 1 reports Sharpe ratios for the Fama-French factors—Size (SMB), Value (HML), Investment (CMA), and Profitability (RMW)—across economic regimes defined by standardized macroeconomic indicators. Regimes labeled as greater (>) or less (<) than zero indicate conditions above or below historical averages.

	CPI% > 0	CPI% < 0	T10Y3M > 0	T10Y3M < 0	CFNAI > 0	CFNAI < 0	GARCH > 0	GARCH < 0
SMB	0.007	0.396	0.454	0.050	0.115	0.351	-0.257	0.638
HML	0.583	0.128	0.552	0.117	0.468	0.177	0.217	0.435
CMA	0.698	0.225	0.617	0.284	0.537	0.330	0.812	0.152
RMW	0.717	0.271	0.759	0.189	0.456	0.427	0.628	0.297

The results summarized in Table 1 reinforce the hypothesis of regime-dependent factor performance. Specifically, during high inflationary periods (CPI% > average), the profitability (RMW, Sharpe ratio approximately 0.72) and investment (CMA, Sharpe ratio approximately 0.70) factors exhibited relatively high Sharpe ratios, indicating their stronger performance under elevated inflation conditions. Conversely, the size factor (SMB) displayed notably weaker performance (Sharpe ratio approximately 0.01) in these same environments.

Analyzing factor returns conditional on the yield spread (T10Y3M), we observe superior performance for the profitability (RMW, Sharpe ratio approximately 0.76) and investment (CMA, Sharpe ratio approximately 0.62) factors during periods characterized by steeper yield curves (T10Y3M > average). Additionally, the size (SMB, Sharpe ratio approximately 0.45) and value (HML, Sharpe ratio approximately 0.55) factors also delivered positive returns in these environments, reflecting generally favorable conditions for riskier investments. Conversely, flatter yield-curve periods (T10Y3M ≤ average) were associated with weaker factor performances, as indicated by notably lower Sharpe ratios across all factors.

Considering economic activity as measured by CFNAI, periods of positive economic growth (CFNAI > average) resulted in relatively strong performances for the investment (CMA, Sharpe ratio approximately 0.54) and value (HML, Sharpe ratio approximately 0.47) factors. In contrast, during periods of negative economic growth (CFNAI ≤ average), Sharpe ratios declined to approximately 0.30 for CMA and 0.18 for HML.

However, the size factor (SMB) performed relatively better (Sharpe ratio approximately 0.28) during these downturns, underscoring the diverse sensitivity of factors to broader economic conditions.

Extending our analysis, we investigate combined macroeconomic regimes (see Appendix Figure A1). When simultaneously evaluating economic growth (CFNAI) and inflation (CPI%), distinct patterns emerge. The investment factor (CMA) displayed notably strong performance during periods of concurrent economic expansion and rising inflation (Growth Up & Inflation Up, Sharpe ratio approximately 1.03). Conversely, the size factor (SMB) exhibited stronger returns in environments marked by economic slowdowns coupled with low inflation (Growth Down & Inflation Down, Sharpe ratio approximately 0.46).

Similarly, the analysis of combined yield spreads (T10Y3M) and volatility regimes (GARCH (1,1)) reveals meaningful insights (see Appendix Figure A2). In periods characterized by steeper yield curves and elevated volatility (Term Spread Up & Volatility Up), profitability (RMW, Sharpe ratio approximately 1.04) and investment (CMA, Sharpe ratio approximately 0.85) factors exhibited notably high Sharpe ratios, highlighting their resilience in uncertain market conditions. Meanwhile, the SMB factor performed robustly during periods of steep yield curves accompanied by low volatility (Term Spread Up & Volatility Down, Sharpe ratio approximately 0.65), emphasizing opportunities for capturing premia associated with riskier small-cap stocks during stable, growth-oriented periods.

3.3. Forecasting Strategies

3.3.1. Predictive Models

In this section, we introduce a predictive modeling framework designed to forecast the likelihood of each factor outperforming in subsequent periods. Our approach incorporates previously discussed macroeconomic indicators—including inflation, yield spreads, economic activity, and market volatility—as well as factor-specific momentum features.

Traditional linear regression models have long served as foundational tools in financial forecasting; however, they depend heavily on restrictive assumptions such as linearity and independence among predictors. These assumptions frequently limit their capacity to capture the complex, dynamic, and nonlinear relationships inherent in financial markets. Additionally, linear models often struggle to accurately interpret macroeconomic signals due to inherent noise, structural shifts, and evolving relationships between predictors and factor returns (Hamilton, 1989).

To address these limitations, we employ two decision tree-based models capable of effectively capturing nonlinearities and intricate interactions among predictors. Compared to traditional linear regression approaches, decision tree-based methods offer greater flexibility, dynamically adapting to evolving market conditions without restrictive assumptions about data distributions. Consequently, these models serve as robust tools for reliably forecasting factor performance.

Beyond predictive accuracy, decision tree-based models also achieve a valuable balance between model complexity and interpretability. Compared to less interpretable machine learning techniques such as neural networks, commonly described as black box models, decision trees explicitly illustrate the decision-making pathways underlying their forecasts. This inherent transparency provides clear economic insights, enhancing both credibility and practical applicability in financial decision-making contexts.

During the training phase, our models systematically learn relationships between predictive indicators and historical outcomes. By explicitly including historical instances identifying the top-performing factor each month, the models can detect intricate interactions and nonlinear dependencies between macroeconomic indicators and factor returns. This learned relationship enables reliable forecasts of factor outperformance probabilities in subsequent out-of-sample periods. These probabilistic forecasts directly translate into dynamic factor exposure weights, facilitating portfolio adjustments aimed at achieving superior returns compared to static, equally weighted strategies.

To enhance predictive accuracy and ensure robust generalization, model hyperparameters were optimized exclusively using historical data from April 1972 through December 1999, based solely on the Fama-French five-factor long-short dataset. These hyperparameters were then held constant and consistently applied out-of-sample from January 2000 onward for the long-short factors. Furthermore, applying these same hyperparameters and predictor features directly to the two additional datasets—the Fama-French long-only portfolios and investable MSCI factor indices—without further adjustments, provides a rigorous test of the models' ability to generalize across diverse factor constructions. This design minimizes concerns related to overfitting and offers a clear framework to evaluate the true predictive capability of our modeling approach.

To maintain predictive relevance over time, our forecasting approach employs a rolling five-year training window (60 months). This dynamic training window ensures that models continuously incorporate recent market developments, thereby enhancing the robustness and timeliness of forecasts. The comprehensive evaluation of model performance, presented in Chapter 5, considers predictive accuracy, realized portfolio returns, and stability to assess the practical applicability of the approach.

3.3.2. Random Forest

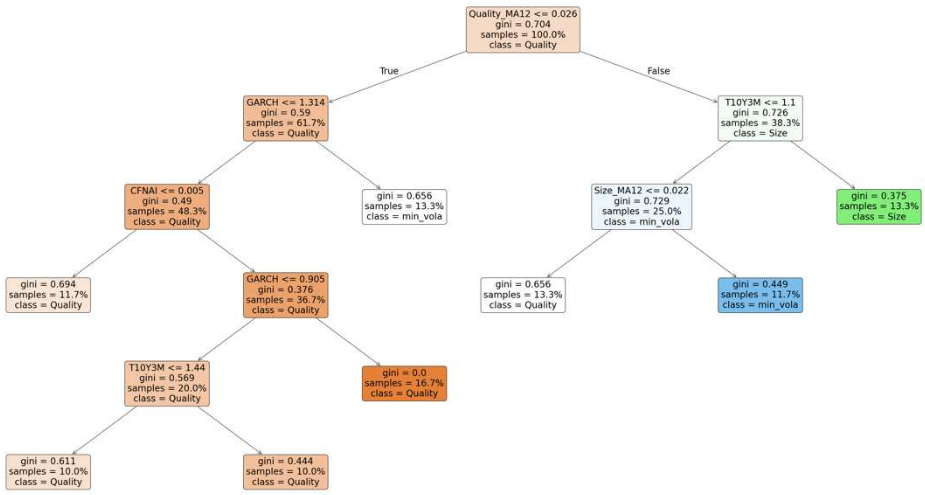
Random Forest models offer notable advantages in financial forecasting because they naturally identify complex, nonlinear relationships without extensive feature engineering. This capability is well documented by Gu et al. (2020), who demonstrated Random Forests' superior performance relative to linear models, especially in predicting stock returns. Building upon these advantages, we adopt the Random Forest methodology introduced by Breiman (2001), an ensemble approach that aggregates predictions from multiple independently trained decision trees.

In our implementation, each tree is trained using historical data from a rolling five-year window. The algorithm partitions this historical dataset by iteratively selecting optimal feature thresholds to maximize differentiation among historically outperforming factors. These splits, typically chosen based on the Gini impurity criterion, progressively segment the data into increasingly homogeneous subsets, enhancing each tree's predictive accuracy.

Figure 2 illustrates a trained decision tree as a flowchart, systematically directing each month's latest feature data through a sequence of binary splits defined by learned thresholds. At each node, the tree evaluates current macroeconomic and momentum indicators relative to historical patterns, sequentially narrowing conditions until reaching a terminal node (leaf). Each leaf identifies the historically top-performing factor for that particular combination of conditions. If predictions relied solely on one decision tree, the monthly allocation would reflect the historical frequency of each factor being the top performer under similar conditions.

Figure 2 – Example Decision Tree

Figure 2 presents a single decision tree from our Random Forest model.



In practice, the Random Forest method aggregates predictions from multiple decision trees. Each tree independently predicts the outperforming factor based on historical training relationships, contributing a single "vote." Aggregating these votes across many trees yields probabilistic forecasts of factor performance, significantly enhancing prediction robustness. These probabilities directly inform dynamic factor weights for portfolio allocation.

To enhance predictive accuracy and maintain robustness, we optimized the Random Forest hyperparameters using historical data from the Fama-French five-factor long-short dataset covering April 1972 to December 1999.

Specifically, we set the number of trees (n_estimators) to 100, balancing predictive accuracy with computational efficiency. Tree depth (max_depth) was unrestricted, allowing the model to fully capture complex interactions. Additionally, we considered all available features at each split (max_features=None) to maximize interaction detection. To enable subtle differentiation among data points, we specified a minimum of two samples required to split a node (min_samples_split=2) and set a minimum of five samples per leaf (min_samples_leaf=5) to ensure statistical robustness. Finally, following recommendations by Hastie et al. (2009), we selected the Gini impurity criterion to minimize misclassification at each node, further improving the model's predictive reliability.

3.3.3. Gradient Boosting

Gradient Boosting, proposed by Friedman (2001), offers a distinctive advantage for financial forecasting due to its iterative, sequential approach to error correction. Unlike methods such as Random Forests, which combine independent trees, Gradient Boosting sequentially constructs each tree to specifically address errors identified in previous iterations. This targeted error-correcting mechanism allows Gradient Boosting to precisely capture subtle predictive signals and nuanced interactions in financial market data, resulting in consistently robust and accurate forecasts. Gu et al. (2020) further validated Gradient Boosting's predictive advantage, demonstrating its effectiveness relative to traditional models, particularly when modeling complex financial relationships.

In our implementation, the Gradient Boosting model's hyperparameters were optimized using the same historical dataset and training period as previously described for the Random Forest model (April 1972 to December 1999, using the Fama-French five-factor long-short dataset). Specifically, we selected 200 trees (n_estimators) to achieve a balance between predictive accuracy and computational efficiency, and limited tree depth (max_depth) to 15 to control complexity and reduce overfitting. A learning rate (learning_rate) of 0.17 was chosen to facilitate efficient convergence while preserving sufficient flexibility to model underlying relationships. To further enhance generalization and minimize correlation among individual trees, each tree was trained on randomly selected subsets comprising 75% of monthly observations (subsample) and 50% of available predictor features (colsample_bytree). Additionally, we set a minimum child weight (min_child_weight) of 7 to prevent excessively granular data partitions, thereby stabilizing predictions and further mitigating the risk of overfitting. For multiclass classification, the model employed the mlogloss evaluation metric, which quantifies the accuracy of predicted class probabilities by penalizing incorrect classifications, thus providing clear probabilistic interpretations of factor forecasts.

Through these carefully tuned regularization techniques, Gradient Boosting effectively manages overfitting and simultaneously offers valuable interpretability via feature importance metrics, enhancing both the economic intuition and transparency of the predictive outcomes.

4. Empirical results

This section empirically assesses the predictive capability of macroeconomic and factor momentum indicators on factor returns. We systematically test whether decision-tree-based machine learning methods can enhance factor allocation and improve return predictions relative to traditional methods.

Our empirical analysis is divided into three subsections. Subsection 4.1 employs traditional long-short Fama-French 5-factor data, serving as a baseline scenario. Subsection 4.2 replicates this analysis with long-only factor portfolios to accommodate practical investment constraints. Lastly, Subsection 4.3 evaluates the performance using investable MSCI factor indices to approximate realistic investment conditions for investors.

4.1. Fama-French 5-Factors

4.1.1. Strategy Allocations

We begin our empirical analysis using the traditional Fama–French long-short factors: SMB (size), HML (value), CMA (investment), and RMW (profitability). These factors are inherently market-neutral and typically exhibit modest absolute returns. The market factor is intentionally excluded from our analysis due to its fundamental difference from these characteristic-driven factors—it captures broad market risk rather than distinct factor-specific premia. Its inclusion would compromise direct comparability, given its long-only nature. By excluding the market factor, we clearly isolate the specific conditions under which characteristic-based factors thrive or underperform.

Table 2 – Factor performance statistics April 1967-Dec 2024 (Long-Short)

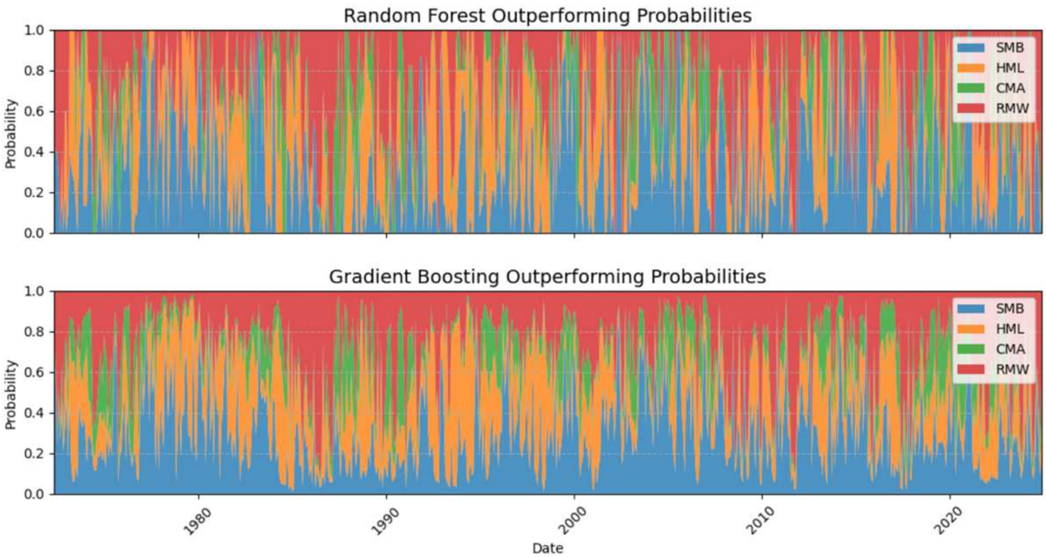
Table 2 reports historical performance metrics for the Fama–French long-short factors. The "win rate" indicates how often each factor achieved the highest monthly return relative to the other factors.

Factor	Annualised Return	Annualised Volatility	Cumulative Return	Win Rate
SMB	1.9%	10.6%	216.7%	30.8%
HML	2.9%	10.4%	473.2%	23.7%
CMA	2.9%	7.2%	472.9%	16.0%
RMW	3.1%	7.7%	568.9%	29.4%

Our dataset spans from April 1967 to December 2024, yielding 736 monthly observations based on the earliest availability of all required predictive features. The factor performance statistics summarized in Table 2 highlight clear distinctions across the factors. SMB delivered episodic periods of strong outperformance, achieving the highest win rate, yet its returns were the lowest overall, emphasizing significant cyclicality and relatively modest long-term performance. In contrast, RMW consistently provided strong cumulative and annualized returns with moderate volatility and a comparatively high win rate, underscoring its stable and persistent performance. HML and CMA generated similar annualized returns but with notably different volatility and win-rate profiles: HML experienced considerably higher volatility and a moderate frequency of outperformance, whereas CMA combined lower volatility with the least frequent monthly outperformance. These patterns align closely with prior empirical findings that investment- and profitability-based factors (CMA, RMW) generally exhibit greater stability compared to the more cyclically sensitive size and value factors (SMB, HML) (Fama & French, 2015; Hou et al. (2015).

Figure 3 – Factor allocations of machine learning models April 1972 - December 2024 (Long-Short)

Figure 3 illustrates the dynamic probabilities of factor outperformance as predicted by the machine learning models. Each factor's assigned probability, indicated by the area size, represents the models' confidence in that factor's likelihood of outperforming in the subsequent month. These probabilities directly translate into portfolio weights for dynamic factor allocation strategies.



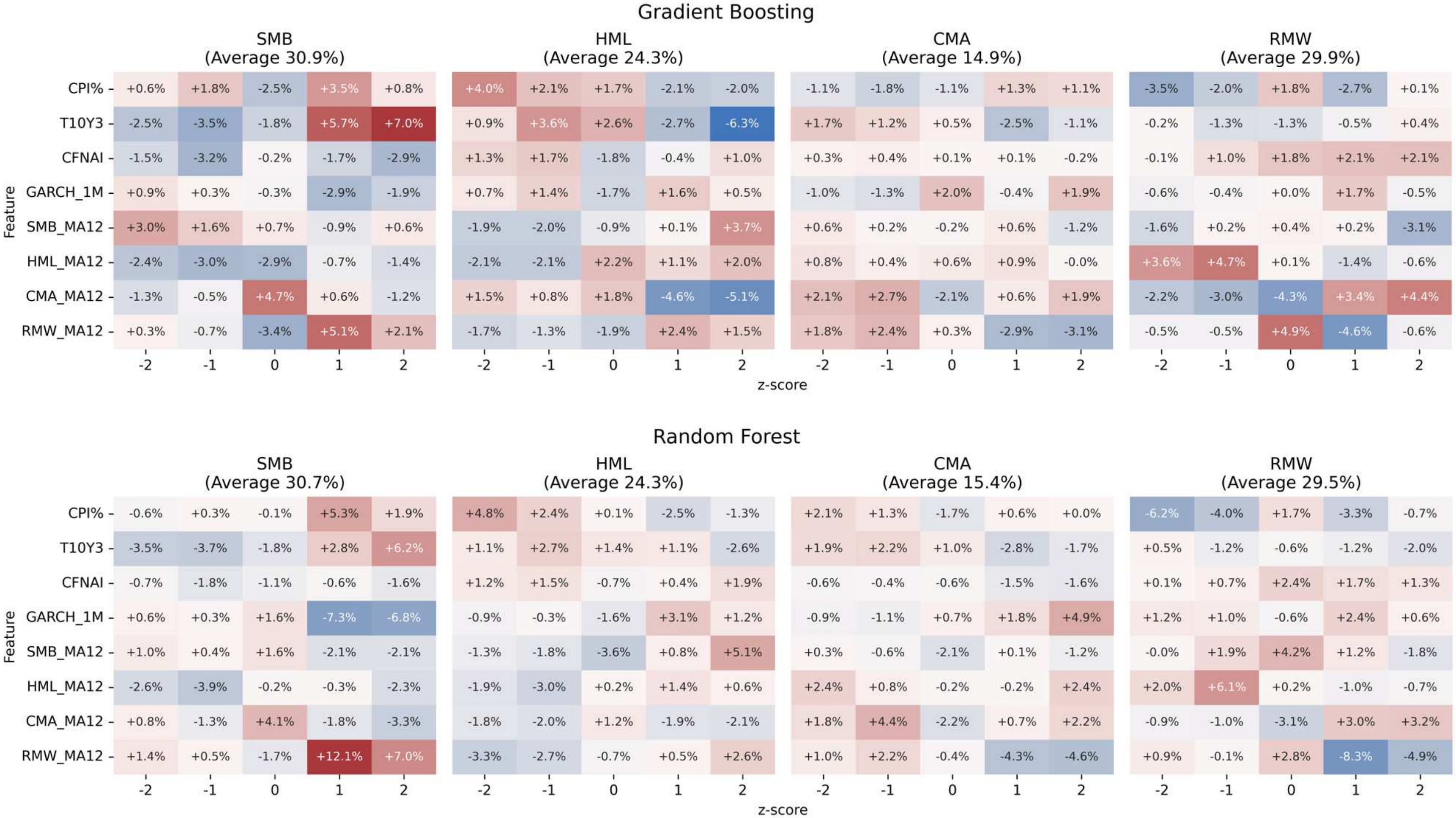
We employ a rolling-window training approach with a 60-month (5-year) interval. Specifically, the initial five-year period (April 1967–March 1972) serves exclusively as training data to generate the first forecast for April 1972. Subsequently, the window rolls forward by one month, continuously updating the training data and generating new forecasts month by month. Figure 3 illustrates the dynamic factor allocations determined by our two machine-learning models: Random Forest and Gradient Boosting. Each month, the models assign portfolio weights based on predicted probabilities of factor outperformance for the subsequent month. Compared to an equal-weighted benchmark (1/N, or 25% per factor), both models typically produce more aggressive, concentrated allocations.

The models' forecasts reflect relationships learned between historical factor performance and predictive feature values. Following each training cycle, the optimized decision trees utilize the most recent month-end data to generate these forecasts. A higher assigned probability indicates stronger model confidence in a factor's likely outperformance, driven by predictive signals from the input features. The degree of allocation concentration thus mirrors the clarity and consistency olof these predictive signals—robust, aligned signals lead to higher factor concentrations, while conflicting signals yield more balanced distributions. Notably, this level of concentration is also influenced by the models' hyperparameters chosen during optimization.

Overall, the Random Forest model demonstrates greater conviction, allocating an average of 67.8% monthly to its highest-ranked factor. Gradient Boosting, by comparison, adopts a more diversified stance, averaging 53.5% for its top factor. Visually, the allocations exhibit pronounced cyclicality, capturing periods dominated by specific factors as well as frequent reallocations in response to changing market and economic conditions. This dynamic behavior highlights the models' potential responsiveness to shifting environments. The predictive accuracy of these allocations and their impact on actual factor portfolio returns are comprehensively evaluated in Sections 5.2.2 and 5.2.3.

Figure 4 – Partial dependence heatmaps for Random Forest and Gradient Boosting (Long-Short)

Figure 4 illustrates how factor allocations systematically deviate from their long-term average under different macroeconomic and momentum conditions (expressed as z-scores). Each heatmap shows the average deviation of factor allocations relative to their overall historical average, highlighting how the model adjusts factor weights based on varying conditions.



To examine how factor allocations vary systematically under different macroeconomic and momentum conditions, Figure 4 illustrates the dynamic responses of our models. Overlaid on these heatmaps are the average factor weights allocated by our models from April 1972 to December 2024. Throughout this period, the models have favored SMB (30.7%) and RMW (29.5%) on average, allocated moderately to HML (24.3%), and notably underweighted CMA (15.4%). These average weights reflect factor predictions based solely on signals derived from macroeconomic and factor momentum features.

To further clarify these systematic variations, the heatmaps display deviations in factor allocations from each factor's long-term average weight, calculated across the entire historical period. These partial dependency analyses illustrate how the models dynamically adjust allocations under changing macroeconomic and momentum conditions. Specifically, each heatmap quantifies average factor allocations under distinct feature states, where these states are defined by standardizing feature values relative to their historical means calculated across the entire sample period, rather than for each individual window. Feature conditions are categorized into z-score bins ranging from -2 to +2; for instance, the '+1' z-score bin includes observations between 0.5 and 1.5 standard deviations above the long-term mean, whereas the '-1' bin covers observations between -1.5 and -0.5 standard deviations below the mean.

Both models exhibit clear patterns in allocation behavior, with some aligning closely with established economic theory while others remain less interpretable. The size factor (SMB), known for cyclical sensitivity, sees notably higher allocations during periods characterized by a positive term spread (T10Y3M) and substantially reduced allocations when the term spread is lower or below its historical average. This aligns with economic intuition, as an inverted or narrowing term spread typically signals recessionary conditions or economic distress, situations in which small-cap stocks tend to underperform due to their cyclical sensitivity and greater vulnerability to financial distress (Fama & French, 1993; Gertler & Gilchrist, 1994). High market volatility (GARCH (1,1)) also significantly decreases SMB allocations, consistent with the known sensitivity of small-cap stocks to heightened risk environments (Ang et al., 2006; Frazzini & Pedersen, 2014). Interestingly, inflation (CPI%) has an inconsistent effect on SMB, suggesting the models perceive inflation less critically for small-cap performance, diverging somewhat from conventional expectations.

Conversely, both models generally allocate lower weights to the value factor (HML) during periods of elevated inflation (positive CPI% z-bins). This aligns with existing literature indicating that value stocks, typically characterized by higher leverage and cyclical sensitivity, often experience reduced margins and profitability during inflationary periods, negatively impacting their relative performance (Fama & French, 1996; Asness et al. (2013). Similarly, HML allocations notably decline when the term spread (T10Y3M) is high, especially at extreme positive levels (+2 standard deviations) in the Gradient Boosting model. Such reductions likely reflect anticipation that steeper yield curves, indicative of potential monetary tightening, may disproportionately disadvantage leveraged, cyclical value stocks (Gertler & Gilchrist, 1994). However, allocations to HML do not consistently decrease during periods of elevated market volatility. Notably, the Random Forest model slightly increases HML allocations at high volatility bins (+1 and +2), possibly suggesting that value stocks may become relatively resilient or undervalued amid turbulent market conditions.

Although both models generally underweight the investment factor (CMA), their responses to macroeconomic indicators reveal nuanced patterns. The Random Forest model moderately increases CMA allocations during low-inflation scenarios, whereas the Gradient Boosting model demonstrates the opposite tendency, favoring CMA during high-inflation periods. Both models consistently reduce CMA allocations during periods of positive term spreads, suggesting diminished attractiveness in anticipated economic expansions. Nevertheless, CMA exhibits relative resilience during conditions of elevated market volatility, particularly within the Random Forest model, aligning with literature emphasizing its defensive characteristics in uncertain markets (Fama & French, 2015; Hou, Xue & Zhang, 2015; Fama & French, 2017). Overall, these nuanced allocation patterns may reflect CMA's smoother return profile and lower volatility, potentially challenging the models' ability to capture clear, consistent macroeconomic drivers.

Both models allocate comparatively less to the profitability factor (RMW) during periods characterized by notably low inflation (CPI z-bins -2 and -1), suggesting the models perceive profitability as less advantageous in deflationary or subdued inflation environments, although this relationship remains somewhat ambiguous due to inconsistent allocations across inflation bins. Interestingly, RMW uniquely exhibits identifiable responses to changes in the CFNAI economic indicator, receiving slightly above-average allocations when economic activity, as measured by CFNAI, is neutral or positive (bins 0, +1, +2). This positive allocation aligns well with RMW's defensive characteristics and stable performance throughout various economic conditions (Novy-Marx, 2013). Despite this defensive nature, neither yield spread (T10Y3M) nor market volatility (GARCH (1,1)) demonstrates clear, systematic allocation adjustments for RMW. This observation is somewhat counterintuitive, given prior literature consistently identifies profitability as a robust defensive factor typically preferred during economic downturns and heightened market uncertainty (Fama & French, 2015; Novy-Marx, 2013).

The momentum signals (MA12) reveal nuanced and mixed influences on factor allocations, and interpretations should remain cautious given the complexity of these interactions. Both Random Forest and Gradient Boosting demonstrate a modest tendency to reduce SMB allocations following periods of strong SMB momentum, possibly indicating a contrarian stance or caution regarding potential reversals after extended rallies. Notably, periods of positive RMW momentum substantially increase SMB allocations—particularly pronounced in the Random Forest model (+12.1%, +7.0%)—suggesting that after sustained performance from the defensive profitability factor (RMW), the models might cautiously anticipate a rotation toward more procyclical factors like SMB as market conditions stabilize or improve following challenging periods. However, clearly interpreting these effects remains challenging because factors may cross-sectionally utilize each other's momentum signals, potentially capturing overlapping predictive information.

Momentum effects for HML allocations remain subtle and mixed, with Random Forest moderately increasing allocations in response to positive SMB and RMW momentum, while Gradient Boosting exhibits less consistent responses. CMA allocations show clearer directional effects: notably decreasing during strong positive RMW momentum, potentially reflecting a perceived shift in relative attractiveness. Conversely, CMA allocations increase meaningfully when CMA or RMW momentum is negative, potentially suggesting a mild contrarian stance or defensive positioning. RMW allocations exhibit clear contrarian tendencies, notably reducing allocations following periods of strong positive RMW momentum. Overall, traditional momentum strategies do not clearly dominate, and both contrarian and momentum-driven adjustments appear context-dependent, underscoring the complexity involved in dynamically incorporating momentum signals into factor allocation decisions.

However, it is essential to recognize that these partial dependencies, represented as z-scores, simplify inherently complex interactions by presenting averaged relationships calculated over an extensive historical dataset spanning more than 50 years. Individual predictions for any given month can, and often do, significantly diverge from these averages due to dynamic interactions among multiple features and the specific macroeconomic and market conditions prevailing at particular points in time. Therefore, partial dependencies should be interpreted cautiously—they provide generalized insights rather than precise representations of the models' actual decision-making process, which simultaneously evaluates multiple interacting features rather than assessing them independently.

4.1.2. Predictive Performance

To assess the predictive capabilities of our two machine learning classifiers, Random Forest (RF) and Gradient Boosting (GB), we utilize several standard classification metrics:

- **Accuracy**: the proportion of all predictions that the model gets right, indicating its overall correctness.
- **Precision**: the ratio of true positive predictions to all positive predictions made, showing how often the model's positive calls are correct.
- **Recall**: the ratio of true positive predictions to all actual positive cases, reflecting the model's ability to identify every positive instance.
- **F1-score:** the harmonic mean of precision and recall, offering a single metric that balances accuracy of positive predictions with completeness.

Table 3 – Aggregate classification metrics for model predictions (Long-Short)

Table 3 presents aggregate-level classification performance metrics across all factors for Random Forest and Gradient Boosting models.

Model	Accuracy	Precision	Recall	F1-score	Samples
Random Forest	29.1%	29.0%	29.1%	29.1%	632
Gradient Boosting	32.4%	31.5%	32.4%	31.8%	632

Model performance results summarized in Table 3 indicate that both models surpass the baseline accuracy of 25%, which would be expected from random selection. The Gradient Boosting (GB) model consistently outperforms the Random Forest (RF) model, achieving higher accuracy (32.4% vs. 29.1%) along with incremental gains in precision, recall, and F1-score.

Despite modest accuracy levels, even small improvements in predictive accuracy can be highly meaningful in the context of finance. Given the inherently noisy and unpredictable nature of short-term factor performance forecasting, slight informational advantages or incremental predictive enhancements can translate into economically significant outcomes, especially when leveraged repeatedly over numerous investment decisions.

Table 4 – Factor level classification metrics for model predictions (Long-Short)

Table 4 illustrates how predictive performance varies across factors and models by presenting factor-specific classification metrics.

Model	**Random Forest**			**Gradient Boosting**			
Factor	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**	**Samples**
SMB	33.3%	32.1%	32.7%	34.8%	37.9%	36.3%	190
HML	29.9%	30.8%	30.3%	34.4%	34.6%	34.5%	159
CMA	13.8%	13.0%	13.4%	17.9%	10.9%	13.5%	92
RMW	31.3%	32.5%	31.9%	32.5%	35.6%	34.0%	191

Table 4 provides a detailed per-factor breakdown of precision, recall, and F1-scores for both models, highlighting substantial variation in predictive performance across factors. Both models demonstrate their strongest predictive ability for the Size (SMB) factor, with Gradient Boosting achieving a recall of nearly 38%, correctly identifying approximately four out of every ten SMB-leading months. Similarly, Gradient Boosting exhibits balanced precision and recall around 34% for the Value (HML) and Profitability (RMW) factors, indicating effective utilization of underlying macroeconomic and momentum signals. In contrast, predictive accuracy notably deteriorates for the Investment (CMA) factor, with recall declining sharply to approximately 13% for Random Forest and 11% for Gradient Boosting. This weaker performance likely stems from an inherent limitation in our dataset's structure: since our models are explicitly trained to identify the single winning factor each month, the infrequent occurrence of CMA as the top-performing factor results in fewer observations from which the models can effectively learn and recognize the specific economic environments or market conditions under which CMA thrives. Consequently, accurately forecasting CMA performance becomes inherently more challenging.

Table 5 – Confusion matrices of model predictions (Long-Short)

Table 5 presents confusion matrices comparing predicted versus true labels for each factor. Rows represent the true factor, and columns represent predicted factors, with each row summing to 100%. Higher diagonal percentages indicate better classification accuracy.

	Random Forest				Gradient Boosting			
	Predicted winner							
True winner	SMB	HML	CMA	RMW	SMB	HML	CMA	RMW
SMB	32.1%	25.3%	12.6%	30.0%	37.9%	22.1%	10.0%	30.0%
HML	26.4%	30.8%	13.2%	29.6%	30.8%	34.6%	5.7%	28.9%
CMA	28.3%	23.9%	13.0%	34.8%	28.3%	19.6%	10.9%	41.3%
RMW	28.3%	23.6%	15.7%	32.5%	31.4%	23.6%	9.4%	35.6%

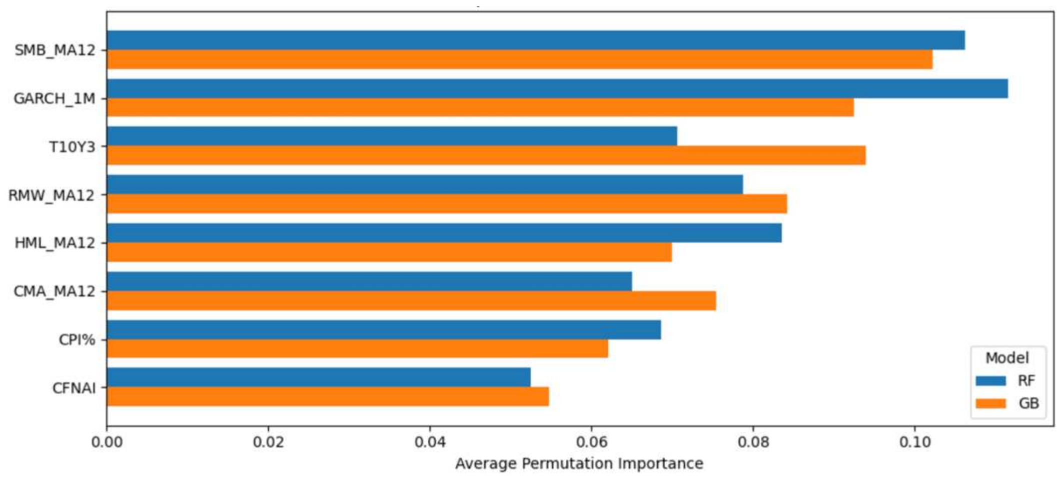
The confusion matrices in Table 5 further illuminate these discrepancies by providing a detailed visualization of classification errors. Misclassifications often occur between factors exhibiting similar economic or market characteristics; notably, CMA is frequently misclassified as RMW, likely reflecting shared defensive or profitability-driven attributes that complicate accurate factor-specific classification. This pattern underscores the complexity of isolating factor-specific signals from broader economic trends. It is important to emphasize that our evaluation focuses solely on correctly identifying the single top-performing factor each month, thus presenting a somewhat binary perspective of prediction accuracy.

In practice, however, a model might still allocate significant portfolio weight to the actual winning factor even when it is not explicitly predicted as the top factor, potentially mitigating the negative impact of misclassification. Conversely, even when a misclassified factor performs relatively well, incorrect predictions can still lead to suboptimal portfolio allocations and subsequent drawdowns. From a practical investment perspective, recall is particularly critical, as failing to identify the true top-performing factor can significantly reduce portfolio returns. Here, GB's superior recall performance across SMB, HML, and RMW is especially valuable, whereas both models' limited ability to detect CMA highlights an area for future methodological improvement.

Another approach to assessing how our models utilize predictive features is through permutation importance analysis, illustrated in Figure 5. Permutation importance quantifies the decrease in a model's accuracy when values of a particular feature are randomly shuffled, effectively replacing meaningful predictive information with random noise. Intuitively, if a feature is critical for prediction, shuffling its values will considerably reduce model accuracy, whereas less important features have minimal impact. Given our models' accuracies of 29.1% (RF) and 32.4% (GB), shuffling most analyzed features significantly reduces accuracy, often bringing it closer to or even below the random baseline accuracy of 25%.

Figure 5 – Average permutation importance of predictive features (Long-Short)

Figure 5 shows the average permutation importance across all rolling-window training periods for Random Forest (RF) and Gradient Boosting (GB) models. Higher permutation importance indicates a feature's greater predictive contribution to the models' forecasting accuracy.



Short-term volatility (GARCH (1,1)) and momentum in size factor returns (SMB_MA12) emerge as the most influential features for both models, highlighting the critical role of volatility and recent performance trends in predicting factor performance. The term spread (T10Y3) also holds notable predictive importance, particularly for Gradient Boosting, suggesting that GB may more effectively exploit signals related to economic expectations and monetary policy conditions. Additionally, momentum in profitability (RMW_MA12) and value (HML_MA12) factors demonstrates consistent predictive value across both algorithms.

Broader macroeconomic indicators, specifically inflation (CPI%) and overall economic activity (CFNAI), exhibit comparatively limited predictive importance. Although these features have clear theoretical linkages to factor returns, their incremental predictive value within our models appears modest relative to short-term volatility (GARCH (1,1)), factor momentum, and the term spread (T10Y3).

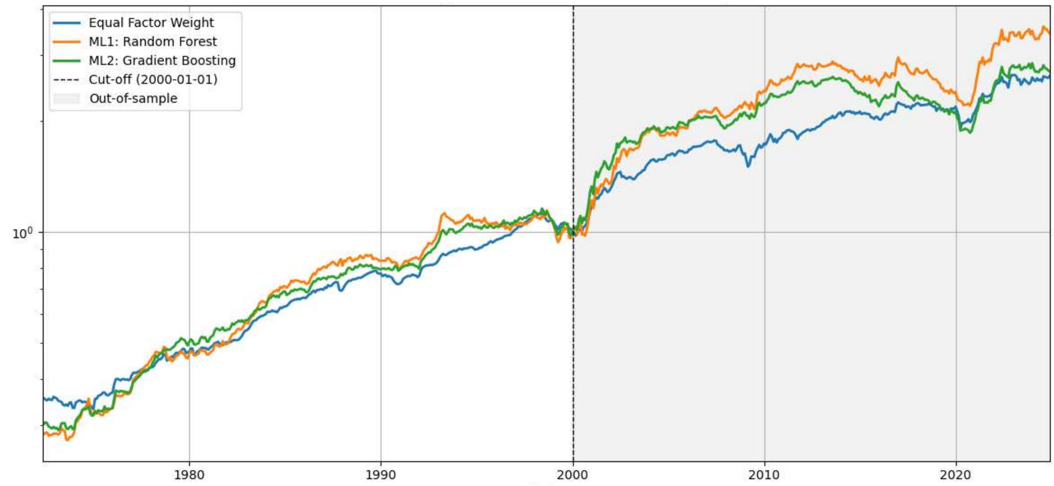
Overall, permutation importance analysis confirms that both machine learning models predominantly rely on recent volatility conditions and specific momentum signals from factor returns. The comparative analysis further highlights subtle differences between the two algorithms in leveraging macroeconomic and momentum indicators, reinforcing the potential advantage of Gradient Boosting in extracting nuanced predictive relationships. These findings underline the complexity involved in feature selection and signal interpretation, crucial for understanding the performance and limitations of these predictive approaches.

4.1.3. Strategy Returns

We now examine the return performance of our Fama-French 5-factor long-short Machine Learning (ML) strategies—specifically, the Random Forest and Gradient Boosting models—in comparison to the Equal Weighted (EW) factor strategy. Figure 6 presents the cumulative returns of these strategies across two distinct periods: the in-sample hyperparameter optimization period (1972-04-30 – 1999-12-30) and the out-of-sample evaluation period (2000-01-30 – 2024-11-30). Both ML strategies clearly outperform the EW strategy during the out-of-sample period, as illustrated by the cumulative returns plotted on a logarithmic scale and rebased on January 1, 2000.

Figure 6 – Cumulative Strategy Growth (Long-Short)

Figure 6 shows the cumulative returns (logarithmic scale) of Random Forest, Gradient Boosting, Equal Weighted, and Market benchmarks from 1972 to 2024. All curves are rebased on January 1, 2000, clearly distinguishing the in-sample optimization period (1972–1999) from the out-of-sample evaluation period (2000–2024).



During the in-sample hyperparameter optimization period, both ML models demonstrated strong performance, as expected, given their optimized hyperparameters tailored specifically to this dataset. Specifically, as shown in Table 6, the Random Forest strategy yielded an annualized return of approximately 4.72% with a Sharpe ratio of 0.80, while the Gradient Boosting strategy achieved an annualized return of approximately 4.37% but with a notably higher Sharpe ratio of 0.90, reflecting its more efficient risk-adjusted returns. Both clearly surpassed the Equal Weighted strategy, which produced an annualized return of 2.82% and a Sharpe ratio of 0.69 during this same period. Moreover, Gradient Boosting exhibited the lowest maximum drawdown (-15.44%), followed closely by Random Forest (-17.94%), both superior to the Equal Weighted strategy (-20.19%), highlighting effective downside risk management by the ML approaches during this period.

Table 6 – Annualized Performance Metrics (Long-Short)

Table 6 shows annualized returns, volatility, Sharpe ratios, and maximum drawdowns for Random Forest, Gradient Boosting, Equal Weighted, and Market strategies during the pre-2000 period.

1972-04-30 - 1999-12-30

Strategy	Annualised Return	Annualised Volatility	Sharpe Ratio	Max Drawdown
Equal Weight Return	2.82%	4.06%	0.69	-20.19%
Random Forest	4.72%	5.93%	0.80	-17.94%
Gradient Boosting	4.37%	4.87%	0.90	-15.44%

In out-of-sample evaluation period, both ML strategies maintained solid performance, as presented in Table 7. The Random Forest strategy achieved an annualized return of approximately 5.30% and a Sharpe ratio of 0.64, while the Gradient Boosting strategy delivered an annualized return of approximately 4.26% with a Sharpe ratio of 0.60. Both ML strategies clearly outperformed the Equal Weighted strategy, which yielded a lower annualized return of 3.26% with a Sharpe ratio of 0.54. Interestingly, despite Gradient Boosting's higher accuracy in selecting the correct factors, its returns were comparatively lower than those of the Random Forest strategy. This outcome likely results from Gradient Boosting's more conservative weighting approach, with an average maximum factor weight of around 54% compared to Random Forest's more aggressive average maximum factor weight of 68%.

Analyzing maximum drawdowns further highlights the risk characteristics of these strategies. During the out-of-sample period, both ML strategies encountered slightly higher drawdowns, with Random Forest at -26.02% and Gradient Boosting at -29.43%, exceeding the Equal Weighted strategy's drawdown of -23.93%. These increased drawdowns indicate that, while ML strategies can yield superior returns, their more aggressive factor allocations might amplify risk in adverse market conditions.

Table 7 – Annualized Performance Metrics (Long-Short)

Table 7 shows annualized returns, volatility, Sharpe ratios, and maximum drawdowns for Random Forest, Gradient Boosting, Equal Weighted, and Market strategies during the out-of-sample period.

2000-01-30 - 2024-11-30

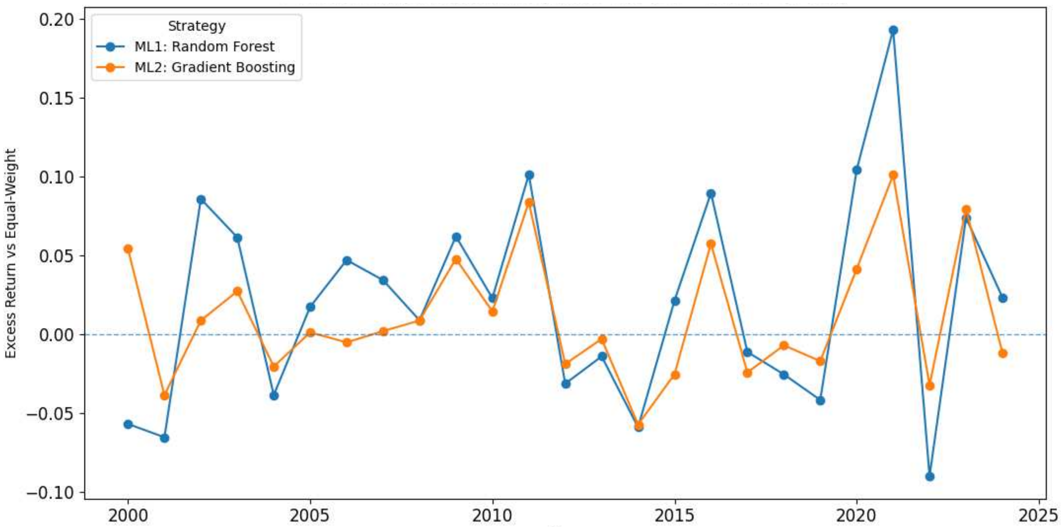
Strategy	Annualised Return	Annualised Volatility	Sharpe Ratio	Max Drawdown
Equal Weight	3.26%	6.02%	0.54	-23.93%
Random Forest	5.30%	8.28%	0.64	-26.02%
Gradient Boosting	4.26%	7.11%	0.60	-29.43%

Further analysis of yearly performance from 2000 onwards, shown in Figure 7, reveals that the Random Forest strategy achieved positive excess returns relative to the Equal Weighted strategy in 16 out of 25 years, highlighting its consistency in generating market outperformance. In comparison, the Gradient Boosting strategy recorded positive excess returns in 14 out of 25 years, reflecting its more cautious approach that often leads to moderate, but steadier, outperformance.

An examination of annual excess returns underscores considerable year-to-year variability for both machine learning strategies. The Random Forest strategy demonstrated pronounced peaks and troughs in performance, notably achieving exceptionally strong returns in years such as 2020 and 2010, illustrating its ability to aggressively capitalize on market opportunities through dynamic factor allocation. Conversely, it also experienced significant underperformance in specific years, notably 2001 and 2022. By contrast, the Gradient Boosting strategy exhibited fewer dramatic fluctuations, indicating a steadier performance pattern. This stability reflects Gradient Boosting's conservative and risk-conscious weighting approach, appealing to investors seeking predictable yet modest outperformance relative to the benchmark.

Figure 7 – Annual Excess Returns Analysis (Long-Short)

Figure 7 shows annual excess returns of the Random Forest and Gradient Boosting strategies relative to the Equal Weighted strategy. The analysis highlights year-to-year variability and the frequency of outperformance across the evaluation period (2000–2024).



We further evaluate the strategies' risk-adjusted performance through regression analyses against the Fama-French 5-factor model over the full evaluation period (April 1972 – November 2024), as presented in Table 8. The Random Forest strategy delivered a positive monthly alpha of 0.163% (annualized alpha approximately 1.98%) with a statistically significant alpha t-statistic of 2.674. Similarly, the Gradient Boosting strategy produced a positive monthly alpha of 0.105% (annualized alpha approximately 1.26%), also achieving statistical significance with an alpha t-statistic of 2.590. Both machine learning strategies demonstrated clear alpha generation relative to the Equal-Weighted strategy, highlighting their capability to generate statistically significant outperformance over the extended evaluation period. However, regression results for the post-2000 subperiod (Appendix Table A3) reveal considerably lower and statistically insignificant monthly alphas (Random Forest: 0.107%, t-stat 1.080; Gradient Boosting: 0.035%, t-stat 0.546), likely reflecting both reduced statistical power due to fewer observations and a genuine decline in alpha magnitude in more recent years, potentially due to greater market efficiency or increased competition in factor-based strategies.

Table 8 – Fama-French 5-Factor Regression Analysis (Long-Short)

Table 8 shows regression results showing monthly alphas, factor loadings (betas), and corresponding t-statistics for Random Forest, Gradient Boosting, and Equal-Weighted strategies over the full evaluation period (April 1972 – November 2024).

Model	α (%)	α t-stat	Mkt β	SMB β	HML β	RMW β	CMA β	Adj. R^2	Ann. α (%)
Random Forest	0.163***	2.674	-0.018	0.277***	0.154***	0.279***	0.348***	0.489	1.978
Gradient Boosting	0.105***	2.590	-0.009	0.292***	0.184***	0.280***	0.295***	0.688	1.263
Equal-Weight	0.000***	19.58	0.000	0.250***	0.250***	0.250***	0.250***	1.000	0.000

Note: *p<0.1; **p<0.05; ***p<0.01

4.2. Fama-French 5-Factors Long-Only

4.2.1. Strategy Allocations

Our second dataset replicates the analysis from Section 4.1.1, utilizing the same predictive features and covering the identical sample period (April 1967 to December 2024). However, this time we employ long-only factor portfolios, reflecting practical investment constraints commonly faced by investors, such as limitations on leverage and short-selling. Rather than constructing market-neutral factors, we analyze portfolios consisting solely of long positions in the top 30% of each characteristic—SMB, HML, CMA, and RMW—thus omitting the short leg. This adjustment significantly alters factor characteristics by increasing factor correlations and volatility due to inherent market exposure. The primary objective of this analysis is to evaluate whether these structural changes influence the predictability of factor returns and affect the identification of macroeconomic and momentum conditions that enhance or diminish factor performance.

Table 9 – Factor performance statistics April 1967-Dec 2024 (Long-Only)

Historical performance metrics for the long-only factors. The "win rate" denotes how frequently each factor delivered the highest monthly return.

Factor	Annualised Return	Annualised Volatility	Cumulative Return	Win Rate
SMB	12.6%	20.5%	146 881.0%	27.7%
HML	14.0%	18.2%	307 691.2%	28.1%
CMA	13.7%	17.3%	269 528.4%	18.0%
RMW	13.3%	17.4%	214 197.8%	26.2%

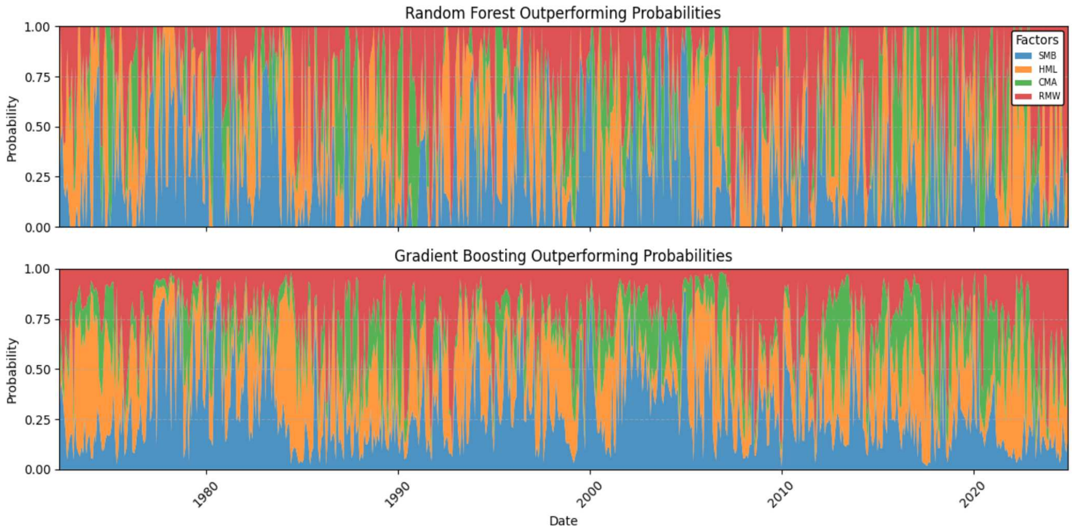
Table 9 reports historical performance metrics for the long-only factors. Compared to the long-short factors analyzed in Section 4.1.1, these portfolios naturally exhibit significantly higher annualized returns, ranging from 12.6% (SMB) to 14.0% (HML), reflecting their inherent positive market exposures. However, this increased return potential is accompanied by substantially higher volatility levels, consistent with the loss of market neutrality. SMB, consistently the most volatile factor across both datasets, again shows the highest annualized volatility (20.5%), while HML, CMA, and RMW demonstrate slightly lower volatility levels.

Similar patterns emerge from the win rates as observed in Section 4.1.1. SMB and HML both exhibit relatively high win rates around 28%, reinforcing their tendency for periodic outperformance despite elevated volatility. RMW maintains a moderately high win rate of 26.2%, indicating stable performance with occasional market-leading returns. CMA, consistent with its long-short counterpart, displays the lowest win rate (18.0%), reflecting steady but comparatively infrequent monthly leadership.

For consistency, we apply the same hyperparameters for both models as used in the long-short dataset. However, the models are naturally retrained using the new data, since the returns and their dependencies on underlying features inherently differ due to the absence of market neutrality in these long-only factors.

Figure 8 – Factor allocations of machine learning models April 1972-December 2024 (Long-Only)

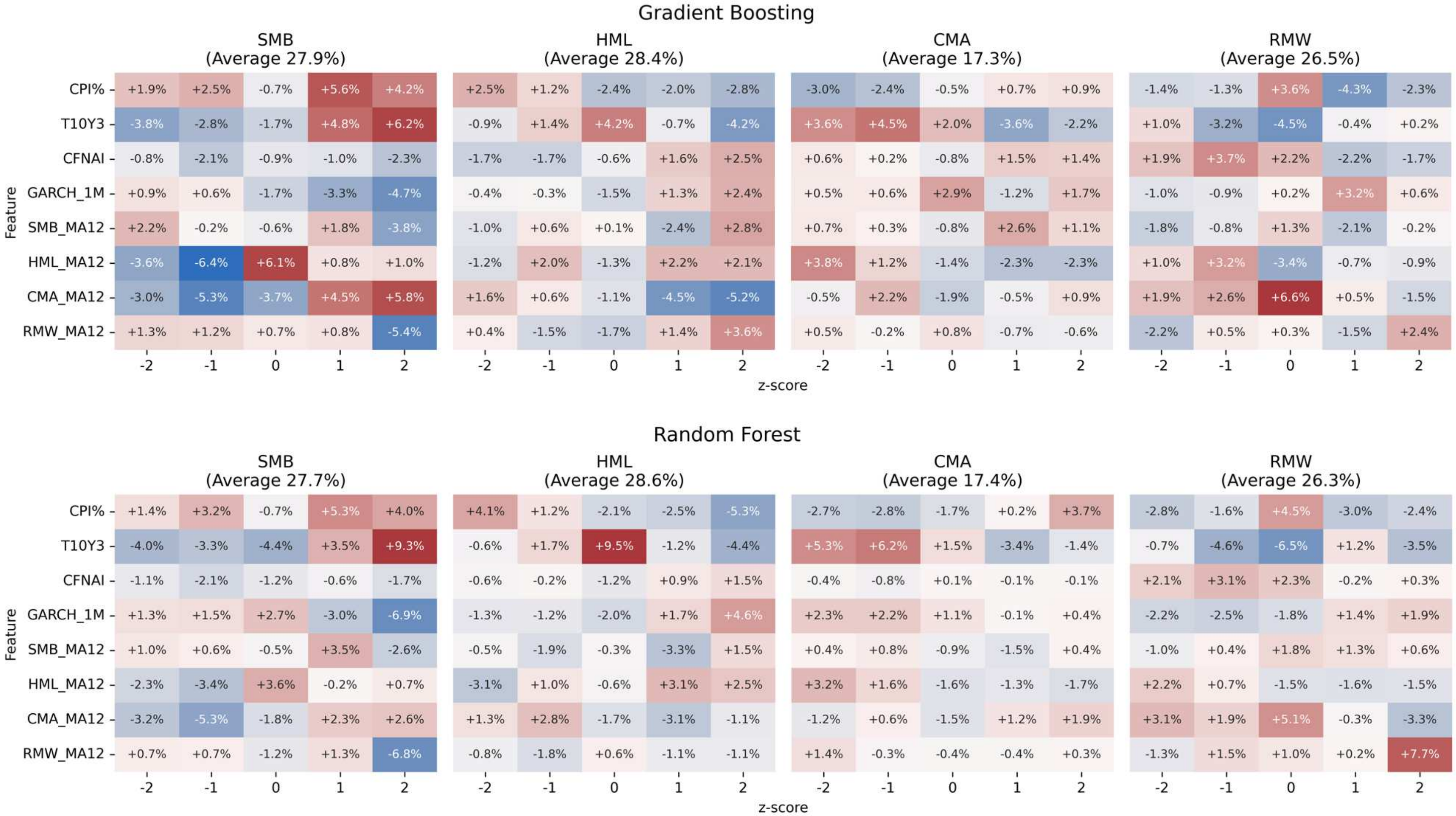
Figure 8 illustrates dynamic probabilities of factor outperformance predicted by the machine learning models, now applied to the long-only factor dataset. Each factor's assigned probability (area size) reflects the models' confidence in its next-month outperformance, directly translating into dynamic portfolio weights.



Similarly to long-short setting, we observe clear differences in allocation behavior between the two models in Figure 8. Random Forest consistently exhibits more aggressive and concentrated factor allocations, rapidly rotating positions in response to shifting market conditions. It assigns an average weight of approximately 66.2% to its top-ranked factor each month. Conversely, Gradient Boosting maintains comparatively more stable and diversified allocations, averaging around 53.2% for its leading factor, suggesting a more cautious stance that relies on steadier positions and less frequent rotations. Notably, these average top-ranked weights closely mirror those observed in the long-short analysis (Random Forest: 67.8%, Gradient Boosting: 53.5%), underscoring how significantly these allocation patterns are influenced by the chosen hyperparameters, which were kept identical across both analyses.

Figure 9 – Partial dependence heatmaps for Random Forest and Gradient Boosting (Long-Only)

Figure 9 illustrates how the models systematically adjust factor allocations in response to varying macroeconomic and momentum conditions (expressed as z-scores). Each heatmap shows the average deviation of allocations from the factor's own long-term average weight, highlighting how factor weights dynamically respond across different market environments.



As we replicate the analysis from Section 4.1.1 using long-only factor portfolios, we observe similar underlying relationships between factor allocations and macroeconomic indicators in Figure 9. However, removing short positions and introducing inherent market exposure results in more pronounced and clearer interactions relative to the market-neutral long-short setting. This amplification aligns with economic intuition, as factor returns now explicitly incorporate broader market dynamics, thus increasing their correlations and sensitivities to macroeconomic conditions.

This heightened sensitivity is particularly evident for the SMB factor. Both models significantly increase SMB allocations during periods of higher inflation, consistent with the notion that elevated inflation typically coincides with stronger economic activity, benefiting smaller, economically sensitive firms (Fama & French, 1993; Gertler & Gilchrist, 1994). Similarly, the relationship between SMB allocations and the yield spread (T10Y3M) becomes clearer; allocations rise markedly when the term spread is positive, highlighting SMB's pronounced cyclical exposure to economic growth expectations. Additionally, SMB allocations consistently decline during episodes of elevated market volatility (GARCH (1,1)), emphasizing its vulnerability in high-risk market environments—a relationship notably clearer than in the long-short context.

The HML factor also exhibits clearer relationships, particularly with inflation. Both models consistently allocate more heavily to HML during low-inflation periods and significantly reduce allocations amid high inflation, aligning with economic theory predicting that value stocks typically face margin compression and lower profitability under inflationary pressures (Fama & French, 1996; Asness et al. (2013). Additionally, HML allocations explicitly decrease when the yield spread (T10Y3M) reaches high levels, reflecting anticipation of tighter monetary conditions adversely affecting leveraged value firms. Notably, HML allocations distinctly increase during periods of rising market volatility, particularly within the Random Forest model, suggesting resilience or opportunistic positioning in turbulent markets.

CMA allocations likewise exhibit clearer cyclical sensitivity. Both models significantly increase allocations during periods of negative term spreads, indicating robust defensive positioning in anticipation of economic downturns or uncertainty—a relationship substantially clearer than previously observed. CMA also displays distinct allocation reductions during low-inflation scenarios, reversing the ambiguous pattern noted in the long-short analysis. Furthermore, allocations to the profitability factor (RMW) similarly show clearer directional responses. Both models notably increase

RMW allocations during neutral to positive economic activity (CFNAI), underscoring its stability across varied economic conditions. RMW allocations distinctly rise during elevated market volatility (GARCH (1,1)), reinforcing its defensive role. CMA allocations, while generally defensive during uncertain economic conditions (particularly negative term spreads), show less explicit responsiveness to volatility than RMW or HML, suggesting nuanced differences in the models' interpretations of defensive characteristics. Conversely, RMW's relationships with inflation and the yield spread remain nuanced, lacking consistent directional patterns.

Transitioning to long-only factor portfolios notably strengthens and clarifies momentum-driven allocation patterns compared to the long-short analysis, although some relationships remain complex. Contrarian behavior remains evident, particularly for SMB and CMA momentum. Both models consistently reduce SMB allocations following strong recent SMB performance, suggesting deliberate rotations away from extended cyclical rallies. Similarly, strong positive momentum in CMA leads to increased SMB allocations, indicating cyclical repositioning after defensive periods.

Conversely, clearer evidence of traditional trend-following emerges for the HML and RMW factors, as both models explicitly increase allocations following periods of strong recent outperformance. Additionally, strong positive RMW momentum notably reduces SMB allocations, highlighting defensive shifts away from cyclical exposures during profitability-driven rallies. These intensified yet nuanced momentum relationships underscore how removing short positions and introducing market exposure amplify previously subtle momentum-based predictive behaviors

4.2.2. Predictive Performance

We now assess the predictive performance of the Random Forest (RF) and Gradient Boosting (GB) models in the long-only investment setting, highlighting differences relative to the previously analyzed long-short scenario. Transitioning from long-short to long-only strategies affect factor correlations and modifies sensitivities to macroeconomic and momentum indicators, potentially impacting model accuracy.

Table 10 summarizes aggregate-level predictive accuracy for both models. Overall, Random Forest exhibits a slight improvement in accuracy, while Gradient Boosting experiences a modest decline; however, both models notably exceed the 25% random selection baseline. These variations likely stem from structural differences: Random Forest's ensemble approach robustly handles increased factor correlations, whereas

Gradient Boosting is more susceptible to noise and minor fluctuations due to its sequential modeling strategy.

Table 10 – Aggregate classification metrics for model predictions (Long-only)

Table 10 presents aggregate-level classification performance metrics across all factors for Random Forest and Gradient Boosting models.

Model	Accuracy	Precision	Recall	F1 -score	Samples
Random Forest	29.6%	29.7%	29.6%	29.6%	632
Gradient Boosting	30.2%	29.4%	30.2%	29.5%	632

Factor-specific results in Table 11 further illustrate predictive performance differences between the two models. Random Forest demonstrates improvements in precision and recall metrics across most factors, notably exhibiting substantial gains in CMA predictions. This aligns with earlier partial dependence analyses suggesting clearer predictive signals from indicators such as the term spread (T10Y3) and CPI% in the long-only setting. However, Random Forest experienced a slight decline in SMB predictions, likely due to heightened volatility associated with SMB, which inherently complicates accurate forecasting.

Table 11 – Factor level classification metrics for model predictions (Long-Only)

Table 11 illustrates how predictive performance varies across factors and models by presenting factor-specific classification metrics.

	Random Forest			Gradient Boosting			
Factor	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**	**Samples**
SMB	29.2%	28.7%	28.9%	30.8%	35.1%	32.8%	171
HML	33.9%	31.5%	32.7%	33.8%	39.2%	36.3%	181
CMA	21.2%	22.3%	21.7%	21.7%	13.4%	16.6%	112
RMW	31.5%	33.3%	32.4%	28.5%	26.8%	27.6%	168

Gradient Boosting's factor-level performance shows a nuanced pattern. It significantly improves predictions for HML and CMA, as indicated by higher F1-scores, suggesting potentially clearer signals within the long-only framework. Specifically, for HML, increased recall paired with lower precision indicates heightened sensitivity at the cost of more false positives. However, Gradient Boosting's predictive effectiveness notably declined for SMB and RMW, likely due to increased volatility and noise from removing short-selling constraints.

Despite these improvements, CMA remains particularly challenging to predict for both models. Its persistently lowest "win rate" among factors, although slightly improved in the long-only scenario, inherently limits the models' ability to establish robust predictive patterns. Moreover, the economic signals influencing CMA performance are potentially subtler or less consistently responsive to macroeconomic and momentum indicators, adding further predictive complexity.

Table 12 – Confusion matrices of model predictions (Long-Only)

Table 12 presents confusion matrices comparing predicted versus true labels for each factor. Rows represent the true factor, and columns represent predicted factors, with each row summing to 100%. Higher diagonal percentages indicate better classification accuracy.

	Random Forest				**Gradient Boosting**			
	Predicted winner							
True winner	**SMB**	**HML**	**CMA**	**RMW**	**SMB**	**HML**	**CMA**	**RMW**
SMB	28.7%	25.7%	18.1%	27.5%	35.1%	29.2%	9.9%	25.7%
HML	23.2%	31.5%	18.8%	26.5%	27.1%	39.2%	11.6%	22.1%
CMA	31.2%	22.3%	22.3%	24.1%	31.2%	29.5%	13.4%	25.9%
RMW	25.0%	25.0%	16.7%	33.3%	30.4%	33.3%	9.5%	26.8%

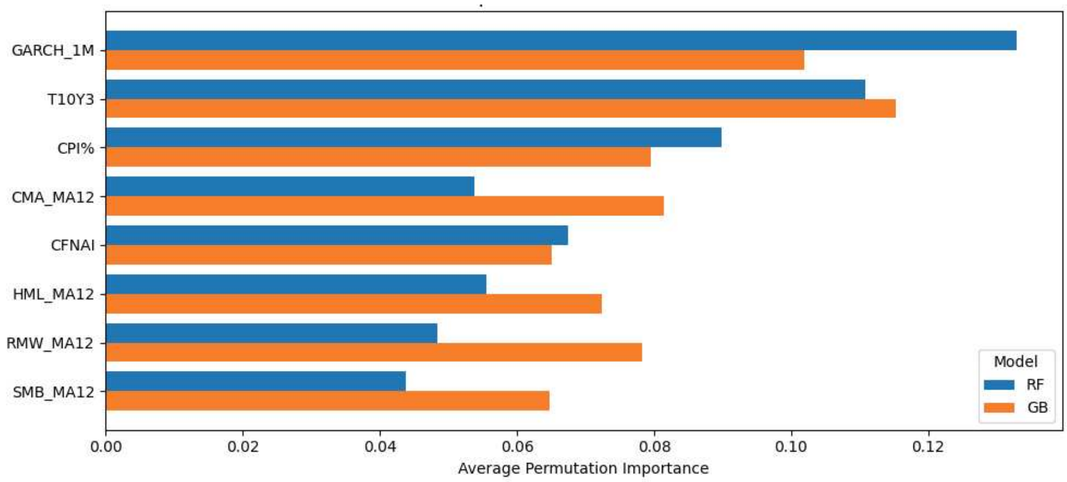
The confusion matrices in Table 12 illustrate nuanced differences relative to the earlier long-short analysis. Gradient Boosting shows clearer classification boundaries for SMB and HML compared to other factors, though notable misclassifications persist, especially between CMA and RMW. This overlap suggests shared defensive or profitability-driven characteristics complicating accurate factor-specific classification.

Random Forest's misclassification errors appear relatively evenly distributed across factors, indicating less distinct factor-specific decision boundaries. Particularly, CMA continues to be commonly confused with RMW, reinforcing their overlapping economic signals. This pattern underscores the ongoing challenge of differentiating factors with correlated macroeconomic sensitivities, a difficulty slightly amplified in the long-only context

To further understand predictive challenges and classification difficulties, permutation importance analysis in Figure 10 reveals how predictor relevance has shifted in the long-only setting. Most notably, macroeconomic indicators—particularly short-term market volatility (GARCH (1,1)) and term spread (T10Y3M)—emerge as significantly more influential for both Random Forest and Gradient Boosting models compared to the long-short scenario. Inflation (CPI%) and broader economic activity (CFNAI) also increase in importance, reflecting enhanced sensitivity of long-only factors to broader economic conditions.

Figure 10 – Average permutation importance of predictive features (Long-Only)

Figure 10 shows the average permutation importance across all rolling-window training periods for Random Forest (RF) and Gradient Boosting (GB) models. Higher permutation importance indicates a feature's greater predictive contribution to the models' forecasting accuracy.



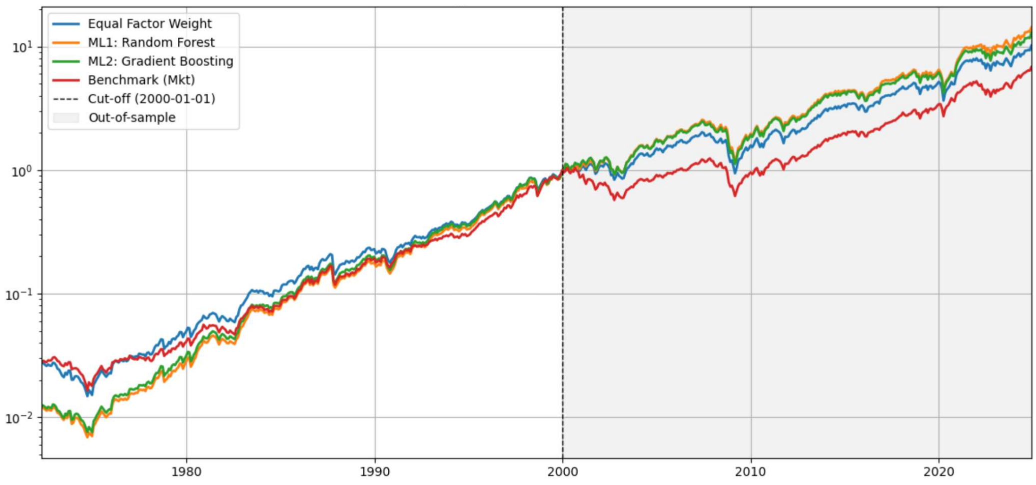
Interestingly, SMB momentum (SMB_MA12), previously among the most critical predictors, notably diminishes in predictive power, becoming the least influential for Random Forest and among the weakest for Gradient Boosting. However, momentum signals for profitability (RMW_MA12), investment (CMA_MA12), and value (HML_MA12) factors maintain moderate to strong predictive roles—especially CMA momentum within Gradient Boosting. This shift suggests that while macroeconomic conditions dominate predictive dynamics in the less market-neutral long-only setting, selected factor momentum signals remain valuable, except notably SMB momentum, which loses clarity and robustness as a predictor.

4.2.3. Strategy Returns

In this section, we evaluate the return performance of our Fama-French long-only factor strategies—specifically, the Random Forest and Gradient Boosting models. These strategies are compared against an Equal Weighted (EW) factor benchmark and a broader Market (Mkt) benchmark sourced from the Kenneth French data library. Figure 11 illustrates the cumulative returns (logarithmic scale) of these strategies, rebased on January 2000. The figure clearly distinguishes two evaluation periods: the in-sample period used for hyperparameter optimization (1972–1999) and the out-of-sample evaluation period (2000–2024). Although hyperparameter optimization was specifically performed for the long-short strategy, the substantial overlap between the long legs of the long-short and long-only strategies justifies evaluating performance over the same out-of-sample period. As depicted in Figure 11, both ML strategies demonstrate notable performance relative to the EW and Market benchmarks, highlighting their relative performance in the long-only context.

Figure 11 – Cumulative Strategy Growth (Long-only)

Figure 11 shows cumulative returns (logarithmic scale) of Random Forest, Gradient Boosting, Equal Weighted, and Market benchmarks over the evaluation period (1972–2024). All curves are rebased on January, 2000, distinguishing the pre-2000 period from the out-of-sample evaluation period (2000–2024).



In the pre-2000 period, both ML strategies demonstrated robust returns, though differences were modest compared to the Equal Weighted (EW) benchmark, as illustrated in Table 13. The Random Forest strategy achieved an annualized return of approximately 15.19% with a Sharpe ratio of 0.84, while Gradient Boosting delivered slightly lower annualized returns of approximately 14.94% with a Sharpe ratio of 0.82.

Both ML strategies outperformed the Equal Weighted strategy, which produced an annualized return of 14.20% and a Sharpe ratio of 0.78. All strategies substantially exceeded the market benchmark, which yielded a lower annualized return of 11.74% and Sharpe ratio of 0.74. An examination of maximum drawdowns for this period indicates similar risk profiles, with Random Forest experiencing a drawdown of -54.27%, Gradient Boosting at -54.81%, and the Equal Weighted strategy at -54.10%, highlighting comparable downside risk characteristics across all factor-based strategies. These drawdowns significantly exceeded the market benchmark's drawdown of -50.39%, underscoring the elevated downside risks inherent in long-only factor-based strategies.

Table 13 – Annualized Performance Metrics (Long-only)

Table 13: Annualized returns, volatility, Sharpe ratios, and maximum drawdowns for Random Forest, Gradient Boosting, Equal Weighted, and Market strategies during the pre-2000 period.

1972-04-30 - 1999-12-30

Strategy	Annualised Return	Annualised Volatility	Sharpe Ratio	Max Drawdown
Equal Weight	16.24%	17.51%	0.93	-45.23%
Random Forest	17.58%	17.42%	1.01	-43.26%
Gradient Boosting	17.47%	17.47%	1.00	-40.34%
Mkt	14.22%	15.84%	0.90	-46.42%

During the out-of-sample evaluation period, ML strategies continued to maintain positive performance relative to benchmarks as seen in Table 14. The Random Forest strategy yielded an annualized return of approximately 12.52% with a Sharpe ratio of 0.66, and the Gradient Boosting strategy achieved an annualized return of approximately 12.12% with a Sharpe ratio of 0.64. These performances surpassed both the Equal Weighted strategy (11.91% return, Sharpe ratio 0.63) and the market benchmark (8.98% return, Sharpe ratio 0.57). Despite Gradient Boosting's higher accuracy in factor selection, its returns were slightly lower, consistent with its conservative weighting approach as previously observed in the long-short setting. Regarding maximum drawdowns, the Random Forest (-54.27%), Gradient Boosting (-54.81%), and Equal Weighted (-54.10%) strategies again showed closely aligned downside risks, suggesting that all factor-based approaches maintained comparable risk management. Nevertheless, all factor strategies exhibited notably higher maximum drawdowns compared to the market benchmark's drawdown of -50.39%, reaffirming the increased downside risk associated with long-only factor strategies

Table 14 – Annualized Performance Metrics (Long-only)

Table 14 shows annualized returns, volatility, Sharpe ratios, and maximum drawdowns for Random Forest, Gradient Boosting, Equal Weighted, and Market strategies during the out-of-sample period.

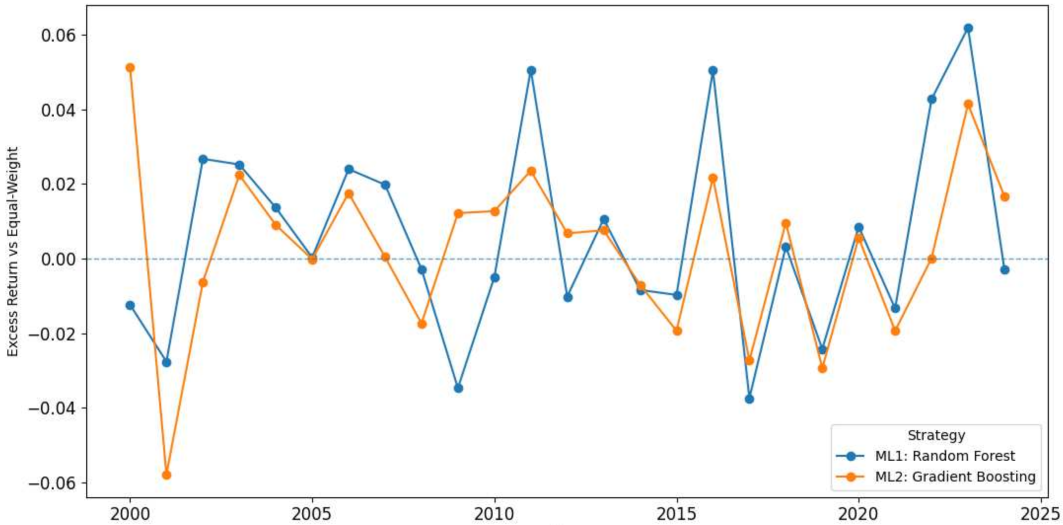
2000-01-30 - 2024-11-30

Strategy	Annualised Return	Annualised Volatility	Sharpe Ratio	Max Drawdown
Equal Weight	11.91%	18.83%	0.63	-54.10%
Random Forest	12.52%	18.94%	0.66	-54.27%
Gradient Boosting	12.12%	19.08%	0.64	-54.81%
Mkt	8.98%	15.84%	0.57	-50.39%

An examination of annual excess returns versus the Equal Weighted strategy from 2000 through 2024 underscores considerable year-to-year variability for both ML strategies which is seen in Figure 12. The Random Forest strategy exhibited significant peaks and troughs, with notably strong outperformance in years such as 2010, 2013, and 2020, achieving positive excess returns in 15 out of 25 years. This highlights its capability to dynamically capitalize on market opportunities through aggressive factor allocations. Conversely, it experienced periods of notable underperformance, particularly in years like 2001 and 2022. In comparison, Gradient Boosting displayed a more stable performance pattern, characterized by fewer dramatic fluctuations and achieving positive excess returns in 13 out of 25 years. This relative consistency underscores Gradient Boosting's conservative, risk-conscious factor-weighting approach, making it particularly suitable for investors prioritizing predictability and steady, moderate outperformance relative to the benchmark

Figure 12 – Annual Excess Returns Analysis (Long-only)

Annual excess returns of the Random Forest and Gradient Boosting strategies relative to the Equal Weighted strategy. The analysis highlights year-to-year variability and the frequency of outperformance across the evaluation period (2000–2024).



Regression analysis against the Fama-French 5-factor model (Table 15) provides additional insights into the relative performance of these strategies. The Random Forest strategy generated a positive monthly alpha of 0.076% (annualized alpha approximately 0.92%), statistically significant at the 5% level with a t-statistic of 2.22. Similarly, the Gradient Boosting model exhibited a positive monthly alpha of 0.060% (annualized alpha approximately 0.73%), also statistically significant at the 5% level (t-statistic of 2.49). Conversely, the Equal Weighted strategy delivered a negative monthly alpha of -0.011% (annualized alpha approximately -0.13%) with a t-statistic of -0.82, indicating no statistically significant deviation from factor-driven returns. Nonetheless, for the post-2000 period (Appendix Table A4), monthly alphas were lower and not statistically significant (Random Forest: 0.062%, t-stat 1.110; Gradient Boosting: 0.030%, t-stat 0.745), again reflecting both fewer observations and smaller effect sizes in the later sample.

Table 15 – Fama-French 5-Factor Regression Analysis (Long-only)

Regression results showing monthly alphas, factor loadings (betas), and corresponding t-statistics for Random Forest, Gradient Boosting, and Equal-Weighted strategies over the full evaluation period (April 1972 – November 2024).

Model	α (%)	α t-stat	Mkt β	SMB β	HML β	RMW β	CMA β	Adj. R^2	Ann. α (%)
Random Forest	0.076**	2.222	1.017***	0.577***	0.202***	0.072***	0.151***	0.975	0.919
Gradient Boosting	0.060**	2.486	1.025***	0.581***	0.200***	0.051***	0.138***	0.988	0.728
Equal-Weight	-0.011	-0.818	1.034***	0.558***	0.214***	0.080***	0.115***	0.996	-0.131

Note: *p<0.1; **p<0.05; ***p<0.01

4.3. MSCI Investable Factors

As outlined earlier, regularly constructing and rebalancing long-only or long-short factor portfolios can pose significant challenges for individual investors and smaller institutions due to substantial time and transaction costs. To provide a more accessible solution, we incorporate investable long-only MSCI factor indices, which offer practical alternatives to directly implementing Fama-French-style factor strategies. These MSCI indices simplify factor investing through easier implementation, reduced transaction costs, and greater accessibility.

We previously selected four MSCI USA indices as approximations for traditional Fama-French factors: MSCI USA Value (for HML), MSCI USA Quality (for RMW), MSCI USA Low Size (for SMB), and MSCI USA Minimum Volatility (representing a conservative proxy for CMA). These indices offer investors cost-effective and investable exposures to desired factor characteristics without short-selling or complex portfolio construction.

We now evaluate the return performance of our MSCI Investable Factor strategies employing the Random Forest and Gradient Boosting models. These results are compared against an Equal Weighted (EW) benchmark consisting of MSCI factors and the MSCI USA index, which serves as the standard US market benchmark. Figure 13 illustrates cumulative returns (logarithmic scale) of these strategies for the period from July 2000 through November 2024, chosen based on the availability of MSCI index price data. This analysis includes cumulative returns, annualized performance metrics, and regression-based statistical evaluations.

Figure 13 – Cumulative Strategy Growth (MSCI)

Figure 13: Cumulative returns of Random Forest, Gradient Boosting, Equal Weighted, and Market benchmarks (2000–2024).

The cumulative returns clearly highlight the performance differences among the strategies. Over the evaluation period, the Random Forest strategy demonstrated substantial cumulative outperformance compared to all other strategies, indicating its effectiveness in dynamically allocating across MSCI factor indices. Conversely, the Gradient Boosting strategy showed more moderate cumulative returns, outperforming the MSCI USA benchmark but closely aligning with the Equal Weighted strategy, suggesting a conservative factor allocation approach.

Examining annualized performance metrics in Table 16 over the evaluation period further highlights these distinctions. The Random Forest strategy generated the highest annualized return (9.82%) accompanied by the strongest Sharpe ratio (0.71), clearly exceeding both the Gradient Boosting strategy and Equal Weighted benchmark, each achieving an annualized return of 8.81% with Sharpe ratios of 0.64 and 0.62, respectively. Notably, all three factor-based strategies significantly outperformed the MSCI USA index, which recorded a considerably lower annualized return of 6.99% and a Sharpe ratio of 0.45.

Table 16 – Annualized Performance Metrics (MSCI)

Table 16 shows annualized returns, volatility, Sharpe ratios, and maximum drawdowns for Random Forest, Gradient Boosting, Equal Weighted, and Market strategies during the out-of-sample period.

2000-07-30 - 2024-11-30

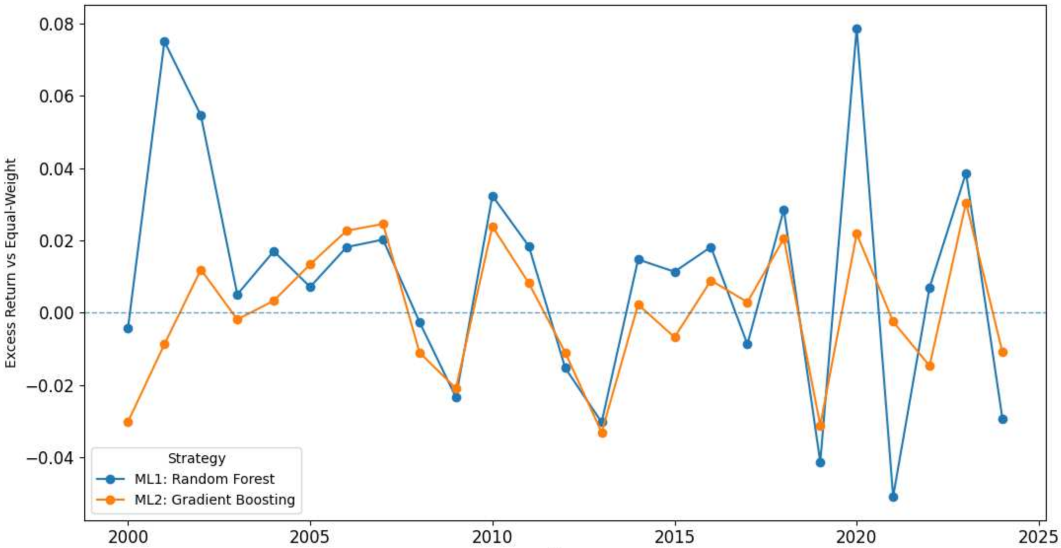
Strategy	Annualised Return	Annualised Volatility	Sharpe Ratio	Max Drawdown
Equal Weight	8.81%	14.18%	0.62	-48.11%
Random Forest	9.82%	13.76%	0.71	-47.02%
Gradient Boosting	8.81%	13.81%	0.64	-47.85%
MSCI USA	6.99%	15.41%	0.45	-52.22%

Analyzing maximum drawdowns further illustrates favorable risk characteristics of the ML strategies. The Random Forest strategy recorded the lowest maximum drawdown of -47.02%, marginally outperforming Gradient Boosting's -47.85%, and significantly better than the Equal Weighted benchmark (-48.11%) and the MSCI USA benchmark (-52.22%). These findings indicate that the ML-driven factor allocations effectively controlled downside risk while enhancing return profiles compared to passive factor approaches and the overall market.

Figure 14 illustrates annual excess returns of the Random Forest and Gradient Boosting strategies relative to the Equal Weighted benchmark over the period from 2000 to 2024. It underscores notable year-to-year variability in strategy performance. Random Forest produced positive excess returns in 16 out of 25 years, reflecting its consistent capability to capitalize on market opportunities through dynamic factor allocation. Gradient Boosting generated positive excess returns in 14 out of 25 years, demonstrating a somewhat steadier yet moderately conservative factor allocation approach. The pronounced peaks, especially observed in the Random Forest strategy during periods such as 2001, 2002, and 2020, highlight its tendency towards more aggressive weighting decisions, capturing greater upside potential but also occasionally amplifying periods of underperformance. Conversely, Gradient Boosting exhibited relatively smoother returns, reflecting its risk-conscious weighting methodology.

Figure 14 – Annual Excess Returns Analysis (MSCI)

Annual excess returns of the Random Forest and Gradient Boosting strategies relative to the Equal Weighted strategy. The analysis highlights year-to-year variability and the frequency of out-performance across the evaluation period (2000–2024).



Regression analysis against the Fama-French 5-factor model (Table 15) provides additional insights into the relative performance of these strategies. The Random Forest strategy generated a positive monthly alpha of 0.076% (annualized alpha approximately 0.92%), statistically significant at the 5% level with a t-statistic of 2.22. Similarly, the Gradient Boosting model exhibited a positive monthly alpha of 0.060% (annualized alpha approximately 0.73%), also statistically significant at the 5% level (t-statistic of 2.49). Conversely, the Equal Weighted strategy delivered a negative monthly alpha of -0.011% (annualized alpha approximately -0.13%) with a t-statistic of -0.82, indicating no statistically significant deviation from factor-driven returns. Nonetheless, for the post-2000 period (Figure A4), monthly alphas were lower and not statistically significant (Random Forest: 0.062%, t-stat 1.110; Gradient Boosting: 0.030%, t-stat 0.745), reflecting both fewer observations and smaller effects, which may be attributed to reduced factor return dispersion or broader adoption of these strategies among investors.

Table 17 – Fama-French 5-Factor Regression Analysis (MSCI)

Regression results showing monthly alphas, factor loadings (betas), and corresponding t-statistics for Random Forest, Gradient Boosting, and Equal-Weighted strategies over the full evaluation period (July 2000 – November 2024).

Model	α (%)	α t-stat	Mkt β	SMB β	HML β	RMW β	CMA β	Adj. R^2	Ann. α
Random Forest	0.056	0.873	0.880^{***}	-0.051^{**}	0.004	0.085^{***}	0.171^{***}	0.933	0.670
Gradient Boosting	-0.029	-0.562	0.893^{***}	-0.069^{***}	0.014	0.083^{***}	0.150^{***}	0.955	-0.353
Equal-Weight	-0.058	-1.498	0.924^{***}	-0.069^{***}	0.069^{***}	0.109^{***}	0.092^{***}	0.976	-0.699

Note: *p<0.1; **p<0.05; ***p<0.01

5. Conclusion

Previous academic research has explored the viability and effectiveness of factor timing strategies, but findings on their efficacy remain mixed. This thesis contributes to the ongoing discussion by empirically assessing whether dynamically adjusting factor exposures using machine learning methods, specifically Random Forest and Gradient Boosting algorithms, can enhance portfolio performance. By integrating machine learning with macroeconomic signals, this research provides a novel perspective and illustrates a practical, data-driven example of how dynamic factor timing strategies can be applied and used as a potential framework in factor allocation decisions.

The empirical analysis indicates that machine learning techniques can be effectively applied to capture complex, non-linear relationships among macroeconomic indicators, market volatility, and factor momentum. In particular, incorporating measures of market volatility and the yield spread between the 10-Year Treasury and the 3-Month Treasury (T10Y3M) appeared to significantly enhance the model's ability to predict equity factor returns. Out-of-sample tests from January 2000 to November 2024 indicate that dynamic allocation strategies outperformed equal-weight equity factor benchmarks on a risk-adjusted basis, as measured by monthly Sharpe ratios. The MSCI dataset showed the largest gain (Random Forest 0.71, Gradient Boosting 0.64 vs. Equal Weight 0.62), whereas improvements were narrower in the long-short dataset (0.64 and 0.60 vs. 0.54) and the long-only dataset (0.66 and 0.64 vs. 0.63).

However, the results must be interpreted cautiously, acknowledging the inherent uncertainties and practical challenges associated with predictive modeling and timing factor exposures. While these findings indicate potential benefits, additional research is essential to verify their robustness and generalizability. Future research could explore incorporating transaction costs, liquidity constraints, and evaluating these strategies across other asset classes and geographical regions. Further investigations using alternative modeling approaches, such as deep learning, as well as continuous real-time validation, could provide deeper insights into the practical applicability and limitations of dynamic factor timing.

References

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets, 5(1), 31–56.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. Journal of Finance, 61(1), 259–299.

Ang, A., & Timmermann, A. (2012). Regime changes and financial markets. Annual Review of Financial Economics, 4(1), 313–337.

Arnott, R. D., Harvey, C. R., & Markowitz, H. M. (2019). A backtesting protocol in the era of machine learning. Working paper.

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. Journal of Finance, 68(3), 929–985.

Banz, R. W. (1981). The relationship between return and market value of common stocks. Journal of Financial Economics, 9(1), 3–18.

Barroso, P., & Santa-Clara, P. (2015). Momentum has its moments. Journal of Financial Economics, 116(1), 111–120.

Bender, J., Sun, X., Thomas, R., & Zdorovtsov, V. (2018). The promises and pitfalls of factor timing. Journal of Portfolio Management, 44(4), 79–92.

Bhandari, L. C. (1988). Debt/equity ratio and expected common stock returns. Journal of Finance, 43(2), 507–528.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Cakici, N., Fieberg, C., Metko, D., & Zaremba, A. (2023). Machine learning goes global: Cross-sectional return predictability in international stock markets. Working paper.

Campbell, J. Y., & Shiller, R. J. (1988). Stock prices, earnings, and expected dividends. Journal of Finance, 43(3), 661–676.

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.

Campbell, J. Y., & Vuolteenaho, T. (2004). Inflation illusion and stock prices. American Economic Review, 94(2), 19–23.

Carhart, M. M. (1997). On persistence in mutual fund performance. Journal of Finance, 52(1), 57–82.

Chen, S.-S. (2009). Predicting the bear stock market: Macroeconomic variables as leading indicators. Journal of Banking & Finance, 33(2), 211–223.

Ehsani, S., & Linnainmaa, J. T. (2022). Factor momentum and the momentum factor. Journal of Finance, 77(3), 1877–1919.

Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. Journal of Financial Economics, 25(1), 23–49.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. Journal of Finance, 47(2), 427–465.

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.

Fama, E. F., & French, K. R. (1996). Multifactor explanations of asset-pricing anomalies. Journal of Finance, 51(1), 55–84.

Fama, E. F., & French, K. R. (2015). A five-factor asset-pricing model. Journal of Financial Economics, 116(1), 1–22.

Fama, E. F., & French, K. R. (2017). International tests of a five-factor asset-pricing model. Journal of Financial Economics, 123(3), 441–463.

Fama, E. F., & French, K. R. (2018). Choosing factors. Journal of Financial Economics, 128(2), 234–252.

Fama, E. F., & Schwert, G. W. (1977). Asset returns and inflation. Journal of Financial Economics, 5(2), 115–146.

Ferson, W. E., & Harvey, C. R. (1991). The variation of economic risk premia. Journal of Political Economy, 99(2), 385–415.

Fleming, J., Kirby, C., & Ostdiek, B. (2001). The economic value of volatility timing. Journal of Finance, 56(1), 329–352.

Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. Journal of Financial Economics, 111(1), 1–25.

Freyberger, J., Neuhierl, A., & Weber, M. (2020). Dissecting characteristics nonparametrically. Review of Financial Studies, 33(4), 1974–2025.

Gertler, M., & Gilchrist, S. (1994). Monetary policy, business cycles, and the behavior of small manufacturing firms. Quarterly Journal of Economics, 109(2), 309–340.

Goyal, A., & Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies, 21(4), 1455–1508.

Goyal, A., Welch, I., & Zafirov, A. (2024). A comprehensive 2022 look at the empirical performance of equity premium prediction. The Review of Financial Studies, 37(11), 3490–3557.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. Review of Financial Studies, 33(5), 2223–2273.

Guidolin, M., & Timmermann, A. (2007). Asset allocation under multivariate regime switching. Journal of Economic Dynamics & Control, 31(11), 3503–3544.

Guidolin, M., & Timmermann, A. (2008). International asset allocation under regime switching, skew, and kurtosis preferences. Review of Financial Studies, 21(2), 889–935.

Hahn, J., & Lee, H. (2006). Yield spreads as alternative risk factors for stock returns. Journal of Financial and Quantitative Analysis, 41(2), 245–269.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.

Hodges, P., Peterson, J., & Ang, A. (2017). Factor timing with cross-sectional and time-series predictors. Journal of Portfolio Management, 44(1), 30–43.

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. Review of Financial Studies, 28(3), 650–705.

Ilmanen, A., & Kizer, J. (2012). The death of diversification has been greatly exaggerated. Journal of Portfolio Management, 38(3), 15–27.

Ilmanen, A., Maloney, T., & Ross, A. (2014). Exploring macroeconomic sensitivities: How investments respond to different economic environments. Journal of Portfolio Management, 40(3), 87–99.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock-market efficiency. Journal of Finance, 48(1), 65–91.

Kelly, B., Malamud, S., & Zhou, K. (2021). The virtue of complexity in return prediction. Journal of Finance, 77(4), 1679–1715.

Lewellen, J. (2002). Momentum and autocorrelation in stock returns. Review of Financial Studies, 15(2), 533–564.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics, 47(1), 13–37.

Liew, J., & Vassalou, M. (2000). Can book-to-market, size and momentum be risk factors that predict economic growth? Journal of Financial Economics, 57(2), 221–245.

Lou, D., Polk, C., & Skouras, S. (2019). A tug of war: Overnight versus intraday expected returns. Journal of Financial Economics, 134(1), 192–213.

Markowitz, H. (1952). Portfolio selection. Journal of Finance, 7(1), 77–91.

Moreira, A., & Muir, T. (2017). Volatility-managed portfolios. Journal of Finance, 72(4), 1611–1644.

Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time-series momentum. Journal of Financial Economics, 104(2), 228–250.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. Journal of Financial Economics, 108(1), 1–28.

Pastor, L., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. Journal of Political Economy, 111(3), 642–685.

Polk, C., Haghbin, M., & de Longis, A. (2020). Time-series variation in factor premia: The influence of the business cycle. Journal of Investment Management, 18(1), 17–39.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity-premium prediction: Combination forecasts and links to the real economy. Review of Financial Studies, 23(2), 821–862.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. Journal of Economic Theory, 13(3), 341–360.

Rosenberg, B., Reid, K., & Lanstein, R. (1985). Persuasive evidence of market inefficiency. Journal of Portfolio Management, 11(3), 9–16.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance, 19(3), 425–442.

Stattman, D. (1980). Book values and stock returns. Unpublished manuscript, University of Chicago.

Turner, C. M., Startz, R., & Nelson, C. R. (1989). A Markov model of heteroskedasticity in stock returns. Journal of Business & Economic Statistics, 7(2), 183–189.

Vassalou, M. (2003). News related to future GDP growth and the value premium. Journal of Finance, 58(5), 2219–2246.

Appendix

Figure A1 - Factor Sharpe Ratios under Combined Economic Growth (CFNAI) and Inflation (CPI%) Regimes

Figure A1 illustrates the performance (measured by Sharpe ratios) of the Fama-French factors—Size (SMB), Value (HML), Investment (CMA), and Profitability (RMW)—across combined regimes of economic growth and inflation.

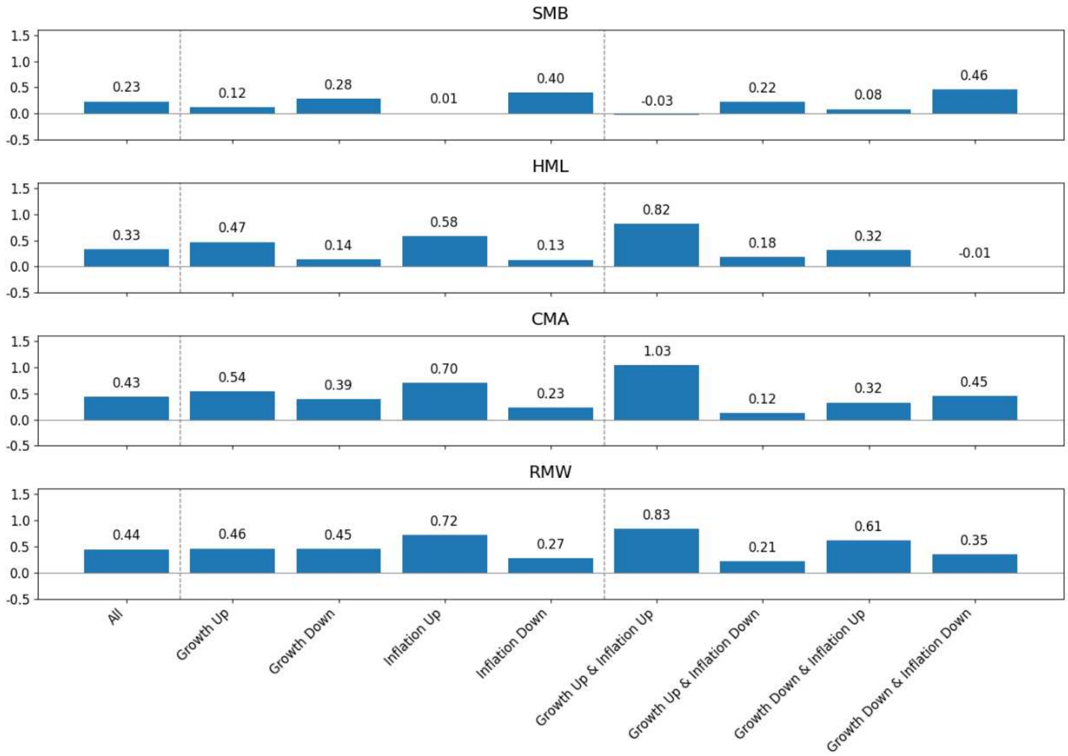


Figure A2 - Factor Sharpe Ratios under Combined Yield Spread (T10Y3M) and Volatility (GARCH (1,1)) Regimes

Figure A2 presents the Sharpe ratios for the Fama-French factors across regimes characterized by different combinations of yield curve slopes and market volatility.

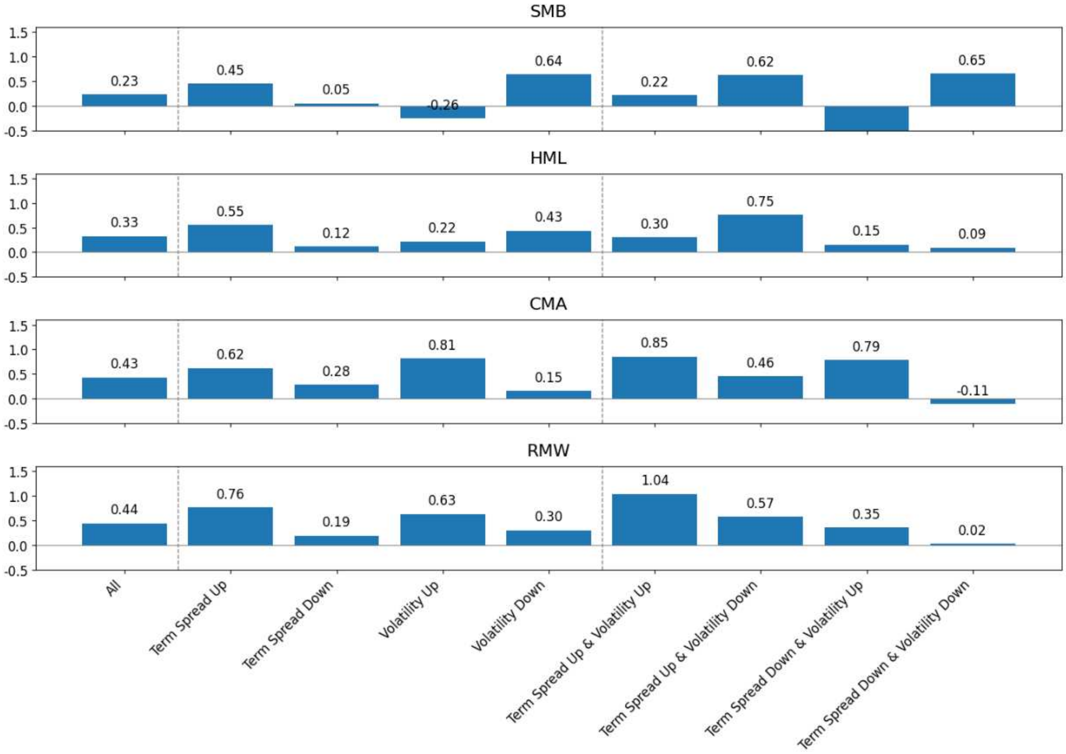


Table A3 – Fama-French 5-Factor Regression Analysis (Long-Short)

Regression results showing monthly alphas, factor loadings (betas), and corresponding t-statistics for Random Forest, Gradient Boosting, and Equal-Weighted strategies over the full evaluation period (January 2000 – November 2024).

Model	α (%)	α t-stat	Mkt β	SMB β	HML β	RMW β	CMA β	Adj. R^2	Ann. α (%)
Random Forest	0.107	1.080	-0.004	0.313***	0.09**	0.335***	0.427***	0.538	1.284
Gradient Boosting	0.035	0.546	-0.003	0.323***	0.153***	0.319***	0.339***	0.740	0.418
Equal-Weight	-0.000	-11.256	0.000	0.250	0.250	0.250	0.250	1.000	-0.000
Note:									*p<0.1; **p<0.05; ***p<0.01

Table A4 – Fama-French 5-Factor Regression Analysis (Long-Only)

Regression results showing monthly alphas, factor loadings (betas), and corresponding t-statistics for Random Forest, Gradient Boosting, and Equal-Weighted strategies over the full evaluation period (January 2000 – November 2024).

Model	α (%)	α t-stat	Mkt β	SMB β	HML β	RMW β	CMA β	Adj. R^2	Ann. α (%)
Random Forest	0.062	1.110	1.012	0.561***	0.247***	0.045*	0.173***	0.972	0.749
Gradient Boosting	0.030	0.745	1.026	0.574***	0.218***	0.030*	0.173***	0.985	0.364
Equal-Weight	0.020	0.918	1.027	0.547	0.247	0.055	0.098	0.996	0.237

Note: *p<0.1; **p<0.05; ***p<0.01