Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Machine Learning for Stock Selection and Portfolio Construction

A Systematic Investing Approach for European Equities

By

Michaïl Eric Maréchal

Supervised by

Dr. Nicholas Danks

MSc Business Analytics 2024/25

Student Declaration

I have read the University's code of practice on plagiarism. I hereby certify this material, which I now submit for assessment on the programme of study leading to the award of MSc Business Analytics is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited within the text of my work.

Student Name: Michaïl Maréchal

Student ID Number: 24340220

Student Signature:

Michaïl Maréchal

Date: 25/07/2025

AI Declaration

I declare that this work has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar.

Generative AI Declaration
Please choose A or B with regards to your use of ChatGPT & other generative AI tools in this project: ☐ **Nothing to declare.** I did not use ChatGPT or any other generative AI software. (see note) ✓ **B. I used ChatGPT** or other generative AI software (see note) NOTE: • If you answer A and the corrector/supervisor, finds evidence that you have indeed used ChatGPT, this behaviour will be considered as unethical and you will be penalized accordingly with reference to the TCD policy on plagiarism. • If you answer B, please clearly explain for which chapters or parts of your dissertation you used ChatGPT and how it helped you to improve your learning process within ethical guidelines. You may include your answer – 300 to 600 words approx.- in the appendix A.

Signed: Michaël Maréchal ID No. 24340220

Date: 25/07/2025

Abstract

The present work assesses the potential benefits in terms of risk adjusted returns of imposing constraints on the mean variance portfolio weights. In this context the expected returns are proxied using the predicted probabilities of outperforming the median weekly return of the considered stocks in the subsequent week. The securities considered are the historical members of the Stoxx 50 and the test window spans with weekly observations from January 2019 to March 2025. Those probabilities are obtained by training five machine learning models: Lasso, Ridge, Elastic Net, Random Forest and Light GBM. Their predictions are then combined into two Ensemble models.

Then different portfolios are formed using equal or value weighted approaches and mean variance based on the predictions from the models. Their results are compared to two benchmarks which are the Stoxx 50 and the equal weighted portfolio of all the considered companies. It is found that imposing constraints on the weights of the mean variance portfolios leads to better risk adjusted results, the optimal range is found to be around 20-28%. The equal weighted benchmark performs the best during the expansion phase and portfolios based on machine learning signal perform best during times of slowdown or recession. Particularly, the 30% constrained mean variance portfolio performs the best, almost two times better than the benchmark on risk adjusted metrics and with lower maximum drawdowns, confirming its utility as a hedge during market stress. Finally, macroeconomic indicators are found to be relevant for all models except for Random Forest.

Table of Contents

List of Tables

List of Figures

I. Introduction

Asset pricing is the field of finance that examines the factors that determine the prices and returns of financial assets. Some of the major contributions include the Capital Asset Pricing Model (CAPM) and the Fama French Five Factor model. These linear frameworks explain expected returns through exposure to different risk factors.

However, Gu et al. (2020) point out an important limitation of linear models is their impossibility of capturing non-linear relationships. The further development of computational capabilities, coupled with more comprehensive data availability enabled the use of more complex statistical models, known as machine learning (ML). Those models and algorithms allow to handle more data, with more flexibility and most importantly to capture non-linear relationships between variables. The use of regression trees and neural networks significantly contributed to increase the predictivity of stock returns and the identification of new hidden factors.

Early studies found exceptionally good results, showcasing high risk adjusted returns and alpha. However, Avramov et al. (2023) observes that most of the abnormally high returns of ML strategies are linked to the absence of economic restrictions. Most of the companies considered by such strategies face limits to arbitrage making those theoretical gains impossible to materialise. Additionally, their high turnover makes them unprofitable when reasonable trading fees are considered. Nonetheless, ML strategies show resilience during period of market stress, as suggested by their lower maximum drawdown and their better relative performance.

A major shortcoming of portfolios assembled using signals captured by machine learning is their tendency to invest in companies characterised by a lower market capitalisation, facing limits to arbitrage and display extreme returns behaviour. To overcome this problem, Breitung (2023) suggests instead of predicting the expected return to predict the probability of outperformance. This approach is shown to avoid investing in companies' previously favoured by ML strategies and prefer more robust stocks.

Wolff & Echterling (2024) following the same approach tested several models as well a combined forecast, surprisingly finding that linear models such as Lasso, Ridge or Elastic Net perform similarly to more advanced deep learning models for such a task. Both studies found strong risk adjusted results and the presence of abnormal returns measured by alpha. Neither explored the potential of including macroeconomic indicators, although previous research has emphasised their role at expressing market expectations about future economic states.

In their study Breitung (2023) used the predicted probabilities of outperformance as a proxy for expected returns to construct mean variance portfolios. However, as already pointed out by Green & Hollifield (1992) mean variance optimisation problem can recommend taking extreme positions into securities. This is due to an erroneous estimate of either the expected returns or the covariance matrix. A potential solution to limit the impact of wrong estimates would be to impose constraints on the weights the different assets can take. This leads to a variance-bias trade-off, where tougher constraints contribute to decrease the variance.

Building on the results of previous studies highlighting the increased performance offered by machine learning and particularly when applied for an outperformance classification task, this paper aims to investigate the hypothesised benefits of imposing constraints on the mean variance portfolio weights in the context of using the predicted probability of outperformance for a given stock as a proxy for expected returns. The primary objective will be to identify if those constraints can lead to better risk adjusted returns. Additionally, it will also investigate the optimal level of constraint, how do the different portfolios perform in different economic cycles and to what risk factors do they load. Finally, the relevance of including macroeconomic indicators for the task at hand will be assessed.

This will be achieved by training five machine learning models with weekly observation for a binary classification task. Their outputs will be combined into two Ensembles as those were found to be better performing than individual models (Krauss et al., 2017; Wolff & Echterling, 2024). Each stock at each time step is either labelled 1 or 0 depending on if it outperformed the median returns of the considered stocks the following week. The considered companies are the historical constituents of the Stoxx 50 for the period of analysis. The training and validation data spans from October 2004 to December 2018 and the test set is from January 2019 until March 2025. Then portfolios will be assembled using simple techniques (equal and value weighted) and mean variance using the predicted probability of outperformance subtracted by 0,5 as a proxy for expected returns, then compared to two benchmarks. Those will be the Stoxx 50 and the equal weighted portfolio of all the considered stocks, as it was found to be a simple benchmark difficult to beat in terms of risk adjusted metrics (DeMiguel et al., 2009). Next different levels of constraints will be imposed on the mean variance portfolios, and it will be analysed how they behave during different economic cycles as defined by the European Business Cycle Clock (Eurostat, 2025). Additional, exposure to risk factors will be measured using the Fama French Five Factor model. Lastly, feature importance will be scrutinised based on the Shapley values, with particular attention given to macroeconomic factors.

This paper aims to contribute to the literature in three ways. Firstly, it is the first to explore imposing constraints to the weights of mean variance portfolios in this context. Secondly, it is the first to include macroeconomic indicators as suggested by Breitung (2023). As mentioned, those offer valuable information about the current market state and investors' expectations. Lastly, it is the first to focus solely on European Equities through the historical members of the Stoxx 50. Previous studies have focused on either the United States or the broader stock market. Already in Breitung (2023) European stocks showed slightly different patterns such as the absence of sensitivity to volume-based indicators compared to other regions.

II. Literature Review

1. Asset Pricing and Portfolio Construction

1.1 CAPM

Asset pricing is the area of financial research concerned with exploring the factors that influence the prices and returns of financial assets. One of its foundational concepts is the Capital Asset Pricing Model (CAPM) which is the result of the combined works of Treynor (1961), Sharpe (1964), Lintner (1965) and Mossin (1966). It is directly derived from the mean variance portfolio introduced by Markowitz (1952). It transforms the algebraic statement into a testable prediction about the relation between risk and expected return. According to it idiosyncratic risk, which is the risk associated to a single security, should not be rewarded. This is because by investing into multiple, preferably uncorrelated stocks, it can be diversified away and thus markets should not assign any premiums for bearing idiosyncratic risk. Consequently, the only factor influencing equity returns is exposure to market risk, as it is the only risk that cannot be diversified away. The CAPM equation is formulated as follows:

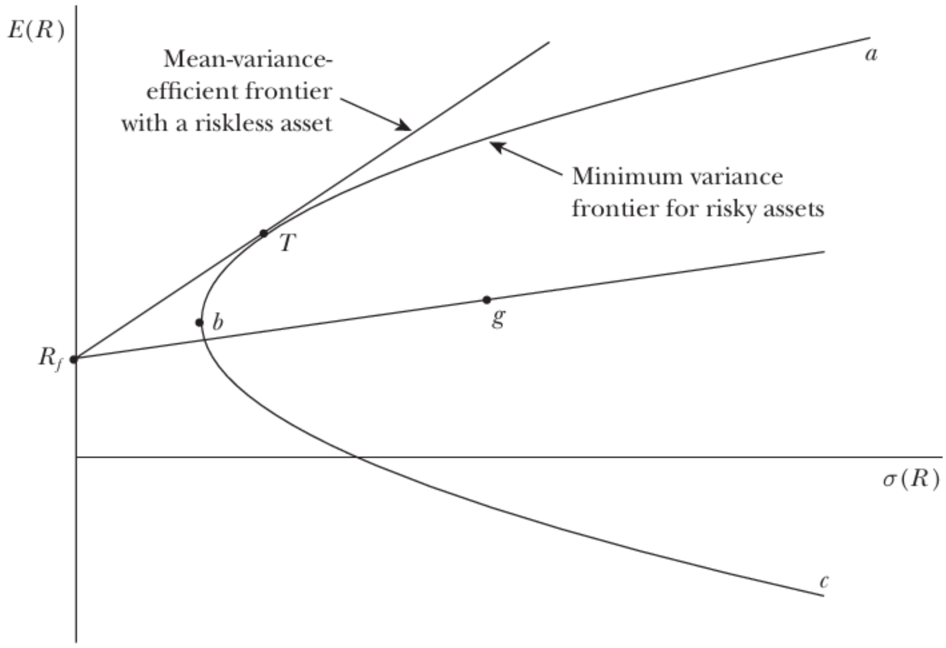
$$ER_i = R_f + \beta_i(ER_m - R_f)$$

Where the exposure to the market is captured in the term β, ER_i is the expected return from the investment, R_f is the risk-free rate and (ER_m – R_f) the market risk premium, with ER_m the expected market return. Beta merely represents the correlation between the asset and the market, essentially representing the sensitivity of the asset's return to variation in the market's return. It is calculated as follows:

$$\beta = \frac{Covariance(R_e, R_m)}{Variance(R_m)}$$

With R_e the return of an individual equity and R_m the return of the overall market. Additionally, Sharpe and Lintner added two assumptions: 1. Complete agreement, where investors agree on the joint distribution of assets from t-1 to t. The distribution should be true and is the one from which returns to test the model are drawn from. 2. Borrowing and lending at the risk-free rate are available for all investors and no matter the amount.

Figure 1: *CAPM Investment Opportunities*



Note. From Fama & French (2004) page 27.

Figure 1 presents the CAPM investment opportunities. The X axis represents risk measured by the standard deviation of portfolio returns and the Y axis the expected returns. The curve abc is referred to as the minimum variance frontier, which encompasses the combinations of expected return and risk for portfolios of risky assets that are minimised for return variance (risk) at different levels of expected return. Those portfolios do not include risk-free borrowing and lending. It visually depicts the trade-off investors face between risk and expected returns. Without the presence of risk-free borrowing or lending only portfolio above b along abc are mean variance efficient. Adding risk-free borrowing turns the efficient set into a straight line as any combinations between risk-free lending and positive investment in g will be along the line passing through R_f and g. Any points to the right of g would imply borrowing at the risk-free rate and investing the proceeds to increase exposure to g. The mean variance efficient portfolio, the one offering the highest Sharpe Ratio[1], can be found by searching for the tangent to the efficient frontier and passing through R_f. The line passing through the two points is called the Capital Market Line (CML). The tangency portfolio T becomes theoretically the only risky

[1] It is a risk adjusted return measure. It is computed by subtracting the risk-free rate from the return of the asset and dividing the result by its standard deviation. $SR = (R_i - R_f)/\sigma_i$

portfolio an investor needs as they could combine it with the risk-free asset to achieve any desired risk-return level. Often the market portfolio measured as a broad equity index is used as a proxy for the tangency portfolio in CAPM.

1.2 Evolution of Asset Pricing

Despite its strong academic anchorage and its reference status the CAPM always has performed empirically poorly. This is the result of theoretical failings and simplistic assumptions (Fama & French, 2004). Additionally, some aspects cannot be directly tested, as for instance stocks have to be compared against a "market portfolio", which in principle should also include not just traded financial assets (consumer durables, real estate or human capital) which are hard to measure. Black (1972) further contributed to soften the assumptions by proposing that assets with $\beta = 0$ do not necessarily have to be the risk-free interest rate as previously claimed by Sharpe and Lintner.

Early research supported that the market proxies used were on the minimum variance frontier and that the beta premium is positive. However, two problems became apparent. First, estimates of beta for individual assets were imprecise, this was resolved forming ranked portfolios for which estimates could be more reliable and leading to the procedure to become standard in empirical testing (Jenson et al., 1972). Second, the regression residuals have common sources of variation, such as industry. Additionally, research found that Sharpe-Lintner CAPM was "flatter" than expected, meaning that its intercept is greater than the average risk-free rate and the coefficient of beta is less than expected (Fama & French, 1992). The returns on the low beta portfolios are too high and the opposite is true for high beta portfolios.

In the 1970s and 1980 more evidence appeared that the CAPM was missing factors. A common theme in research was that ratios incorporating stocks prices have information about expected returns missed by beta. This is logic as a stock price depends on the expected cash flow and the expected returns to which this cash flow is discounted to the present. Fama & French (1992) synthesise the empirical evidence and confirm with a cross-section regression approach that size, earnings-price, debt-equity and book to market ratios hold explanatory power.

A major extension to the CAPM is Merton's (1973) Intertemporal Capital Asset Pricing Model (ICAPM). CAPM's assumption is that investors are only concerned with the wealth their portfolio produces at the end of the current period, ICAPM also considers the opportunities they will have to consume or invest the payoff. When forming their portfolio at time t investors consider how their wealth will vary with future state variables (labour income, prices of consumption goods and the nature of portfolio opportunities). Similarly, ICAPM investor prefer high expected returns and low return variance but are also concerned with the covariance of portfolio returns with state variables.

In pursuit of specifying those state variables Fama & French (1993) take an indirect approach and highlight the higher average return of small stocks and high book-to-market stocks suggesting unidentified state variables which are priced separately from the market beta. The main flaw of their three-factor model from a theoretical perspective is that the factors added were not motivated as relevant state variables of concern for investors, but rather they are the product of brute force constructs meant to capture previously uncovered patterns.

However, some academicians disagree that the new factors represent a valid extension to the CAPM. Their claim is that the market tries to set prices according to the CAPM, but violations happen due to investors overaction. This is because the biggest contributor is the book to market factor (B/M) and therefore, the "behaviouralist" claim that this factor exposes investors extrapolation of past performance, resulting in stock prices too high for growth (low B/M) and on the opposite for firms fallen in disfavour (high B/M). Unfortunately, it is not possible to test their theory if prices are rational or not, as it would require taking a stance on what the market is trying to achieve and what does risk mean. The factor approach does not depend on taking such as stance and remains relevant for asset pricing.

The research for new risk factors is a continuous process, leading current academics to argue for the potential inclusion of ESG as a new factor. This would suggest that companies with better ESG scores would have lower expected returns than their counterparts as the market would be offering a premium for firms carrying more climatic risk (Cornell, 2021).

1.3 Fama French 5 Factor Model and Abnormal Return Measurement

The three-factor model incorporates the market, small and book to market factors. The small minus big factor (SMB) reflects that small stocks are found to have higher average returns than their counterparts. Additionally, their returns covary more with one another than with returns of large stocks. The high minus low (HML) represent the similar observation that stocks with a high book to market ratio outperform those with a low one. Those additions significantly contributed to explain a part of the errors observed in the CAPM. The model was later augmented by two new factors robust minus weak (RMW) and conservative minus aggressive (CMA)(Fama & French, 2015). Companies with a robust operating profitability are found to outperform those with a weaker profitability. Moreover, on average companies with a more conservative investment approach outperform those with a more aggressive strategy. Resulting in the following equation:

$$ER_i = R_f + \beta_{i,MKT}(ER_m - R_f) + \beta_{i,SMB}SMB + \beta_{i,HML}HML + \beta_{i,RMW}RMW + \beta_{i,CMA}CMA$$

Jensen (1968) introduce an approach to test asset pricing models and use them to measure abnormal portfolio performance. Theoretically the regression model should explain all the equity returns and thus lead to an error term, referred to as *alpha*, equal to 0. However, it is empirically found that they are not able to capture the entirety of the variation in equity returns. Consequently, a talented fund manager could potentially generate a positive alpha, suggesting that their fund obtains higher returns than expected for a given level of risk. This measure can be captured precisely by subtracting the risk-free rate from both sides and thus essentially making the alpha the intercept. Using an Ordinary Least Squared (OLS) regression, regressing the risk factors on the excess returns allows to calculate the coefficient and significance of this alpha term. This serves as an important measure to evaluate abnormal returns referred to as Jensen's alpha.

$$ER_i - R_f = \alpha + \beta_{i,MKT}(ER_m - R_f) + \beta_{i,SMB}SMB + \beta_{i,HML}HML + \beta_{i,RMW}RMW + \beta_{i,CMA}CMA$$

1.4 Mean Variance

The mean variance portfolio was introduced by Markowitz (1952) and as mentioned is the principle building block for most models in asset pricing and the foundation of Modern Portfolio Theory (MPT). It is an optimisation problem to find a trade-off between expected return maximisation and volatility (risk) minimisation. This is achieved by computing the efficient frontier which is a curve representing the maximised expected returns for any predetermined level of risk. Returning the optimal investment strategy. Both opposite objectives can be combined in the following mathematical expression:

$$\min_{w} w^T \Sigma w \; - \theta * w^T \mu$$

Subject to:

$$w^T 1 = 1 \; or \; 0$$

$$\theta \geq 0$$

Optionally for long only:

$$w \geq 0$$

Where w is the portfolio weights, Σ the covariance matrix of asset returns and μ the vector of expected asset returns. Θ represents the risk aversion parameter and determines if more emphasis should be put on minimising risk or maximising expected returns. When set to 0 it equates to solving for the minimum variance portfolio. Usually in most settings its value is set to 1, as it is the case in the other papers referenced. For long only portfolios weights have to sum to 1 to form the unit cost portfolio. On the contrary for long short portfolios weights have to sum to 0 thus forming the dollar neutral portfolio. The two main inputs for this optimisation problem are Σ and μ, despite them representing a view on the future movements of asset returns they are usually proxied with historical values. Now with the emergence of powerful prediction models those are used to forecast values for μ.

Although the mean-variance framework theoretically provides an optimal solution, it is frequently observed that some portfolio weights take on extreme values. Those are usually attributed to estimation errors of the expected returns or the covariance matrix. Green & Hollifield (1992) offered a first view on this claiming that those extreme positions are not necessarily due to estimation error, but rather under certain population structures this can be the optimal solution. Additionally, this can be the result of a dominant factor effect like a strong correlation between returns leading the model to take extreme positions in assets with different betas to eliminate both the factor and residual risk. Jagannathan & Ma (2003) argue for the use of constraints, even wrong ones, to improve out of sample performance. The number of parameters to be estimated and potentially erroneous grows exponentially with the number of assets considered. This can lead to overfitting and unstable weights leading to poor out of sample performance. Constraints can be thought of as a regularisation tool, reducing variance in the bias-variance trade-off. Those constraints do not necessitate to be theoretically "correct" as they simply reduce exposure to noisy and wrong estimates. Resulting in more robust portfolio weights, lower out of sample variance and better risk adjusted returns. This paper building on the results obtained from Jagannathan & Ma (2003) explores the potential benefits of imposing constraints on the portfolio weight using probabilities of outperformance as a proxy for μ.

1.5 Considerations

Other portfolio construction techniques were considered, namely Omega Factor (Keating & Shadwick, 2002) and Black-Litterman asset allocation model (Black & Litterman, 1990). Omega Factor optimises the portfolio for the Omega Ratio which accounts for the entire distribution of returns (incorporating asymmetries and fat tails, which MV does not) and tries to optimise the relative probability of portfolio return or loss exceeding a critical value. It was left out given that Ma et al. (2021) finds that mean variance significantly performs better when combined with machine learning forecasts. The Black Litterman portfolio directly builds on mean variance. It allows to incorporate views one could have about expected returns and associate with them a level of confidence. Although it was not included as it is more complex and also not the usual procedure followed in the literature, it would still nonetheless be an interesting direction to explore for a future work.

2. Machine Learning in Asset Pricing

2.1 Foundational Papers

Despite their wide acceptance the asset pricing models presented suffer from multiple limitations. Consequently, Gu et al. (2020) explored if machine learning (ML) methods can improve empirical asset pricing models by extracting signals from a large set of characteristics. Researchers have increasingly adopted these methods in financial research due to a further development of computational capabilities, coupled with more comprehensive data availability which has enabled to leverage the potential of machine learning approaches. The motivations to implement such techniques stem from the several challenges faced by those so-called linear 'traditional' models. Firstly, they tend to perform poorly in high dimensionality, as they fail to accommodate the many features. Secondly, financial data is characterised by a low signal to noise ratio, easily leading to overfitting and increased difficulty to capture meaningful results. This translates into satisfactory in sample results, but the performance significantly degrades for out of sample predictions. Lastly, their linear nature makes them unable to capture non-linear patterns and identify significant interactions between features. Machine learning offers the possibility to alleviate these issues faced by linear models and potentially uncover new "hidden" significant parameters for asset pricing.

Consequently, Gu et al. (2020) observes that ML approaches demonstrate significant performance gains over traditional counterparts, resulting in economic gains for investors. Tree-based models and neural networks outperform linear models such as OLS, Ridge and Lasso, with the top performers being neural networks followed by Random Forest. Those predictive gains can be mainly attributed to ML's ability to capture interactions and nonlinearities. Additionally, the paper highlights the difficulty of predicting individual stocks, but when forming quantile portfolios return differences become large and economically meaningful. Building on this Chen et al. (2023) uses Long Short-Term Memory (LSTM) with an incorporated economic structure for a no-arbitrage condition to directly predict the stochastic discount factor (SDF). This approach significantly outperforms others and suggests the importance of embedding economic structure into deep learning models. Moreover, this paper strongly supports the importance of macroeconomic states when predicting equity returns.

2.2 Extensions

Avramov et al. (2023) further expands on the models proposed by Gu et al. (2020) and Chen et al. (2023). The main hypothesis they try to test is whether the abnormally high returns of ML strategies is due to the absence of economic restrictions. To do so they test the models on different sets, firstly including all companies, second excluding microcaps, third excluding firms which do not have credit ratings and finally, excluding no rated and distressed firms. Companies that would fall within one of those categories are more likely to be mispriced due to limits to arbitrage. However, their hard to arbitrage status makes them hard to trade in real life, consequently leading to higher transaction costs and lower profitability. Their results show a strong gap in performance between models trained with and without economic restrictions. Models trained on an economically restricted set (no microcap, non-rated or distressed firms) perform significantly worse than their counterparts, leading to no abnormal performance as indicated by alpha. The deterioration is mostly substantial for deep learning models, regular models still underperform but their drop in precision is milder. This suggests that most ML signals are founds in stocks which are hard to arbitrage and resultingly mispriced. Additionally, the severe drop in returns for deep learning model suggests that nonlinearities and interactions are mostly relevant for difficult to value and arbitrage stocks. The implications are that although

ML trading strategies can show strong performance in research, their gains are hard to materialise. Moreover, ML trading strategies achieve a high turnover, therefore, in the presence of reasonable trading costs almost all the meaningful risk adjusted returns are eroded with no alpha left on the table. Nonetheless ML strategies show strong resilience during periods of market stress characterised by lower drawdowns and prove to be a good hedge during market crises, despite the extreme positions implied by the mean variance portfolio. Furthermore, the models only load to known risk factors (FF6[2]) suggesting that no new hidden factor has been identified, and it invests in companies which are consistent with anomaly-based trading strategies. This means that despite their opaque nature those models successfully identify mispriced stocks in line with solid economic foundations and empirical findings. Lastly, it is observed for all models that they perform substantially better during periods of stress (high market volatility and low liquidity), except for linear models which are more consistent through all cycles.

Previous studies where only concerned with predicting the returns or the stochastic discount factor (SDF). Although those would still implement those predictions in portfolios this would be done using more naïve approaches, such as an equal weighted portfolio long in the top decile and short in the bottom decile. Ma et al. (2021) explores if return predictions from machine learning models could be used as input for more sophisticated portfolio construction techniques. With that regards they implement portfolios based on mean variance with forecasting (MVF) and Omega Factor (OF) following the approach presented in Yu et al. (2020). Their results demonstrate significantly better results for portfolios constructed using those techniques as compared to the naïve approaches. Moreover, they find that Random Forest (RF) performs best, followed by Long Short-Term Memory (LSTM) and that portfolios built using MVF outperform those constructed using OF. Those results remain valid even after accounting for transaction costs and despite the higher turnover of RF+MVF.

Another key question in asset pricing is which factors are relevant and should be included in prediction models. Given the ability of machine learning models to accommodate a large number features it is easier to include many predictors and potentially uncover surprising interacting variables. With regards to that Zheng & Lucey (2025) forecast the returns of the S&P Global Clean Energy Index including media-related sentiment factors, sourced from the GDELT database. They test seven prediction models and two models to simulate future feature

[2] Fama French Six Factors models. It additionally includes the momentum factor on top of the other five standard risk factors.

values. For the latter task Random Forest substantially outperforms LSTM consistently. For short term forecasting (one week) ARIMA and Prophet are the most reliable, but media-related factors do not play a crucial role in the prediction. Nonetheless, on a medium (one month) and long (six months) term media related factors are found to be important. For the medium term it was observed that using ARIMA in a first place to establish a baseline for the index and then apply Prophet to refine the prediction range yields the best results. Finally, for the long-term VAR offers the best accuracy. Overall, Prophet demonstrated the strongest ability to predict trends.

Finally, Maréchal (2024) explored the predictive power of political factors for ASEAN Markets using LSTM. Political Stability, Control of Corruption and Government Effectiveness were found to have a statistically significant contribution using Permutation Feature Importance. Interestingly, Regulatory Quality did not show significance at an aggregate level, but was significant on a per country basis, highlighting the difficulty to capture all the connections between variables and the dilution that can happen on an aggregate level. Finally, those factors demonstrated a stronger relevance for countries that experienced recent serious political turmoil such as Thailand and Indonesia. Overall, this study stressed the important effects of political stability and good governance on a country's economy and stock market.

2.3 Forecast Probabilities of Outperformance

Machine learning approaches have contributed to substantially improve prediction accuracy for equity returns compared to traditional model, nevertheless, this improved precision is still hard to materialise. It is argued that this is linked to the fact that ML trading strategies tend to invest more into smaller, less liquid, volatile stocks or firms under financial distress (Avramov et al., 2023). Consequently, ranking based portfolios are associated with higher volatility, larger maximum drawdowns and higher kurtosis of returns. To alleviate this problem (Breitung, 2023) suggests turning the regression into a classification problem. This is achieved by labelling each stock at each time step as an outperformer (label 1) if its next month's stock return is higher than the median next month's stock return for all stocks in the dataset and otherwise an underperformer (label 0). To confirm this statement, they construct two random forest models, one regression predicting absolute stock returns and another classifying stocks at each period into one of the two categories. The dataset is composed of a wide selection of liquid international stocks for which only the technical indicators are considered. Those are retrieved using the Technical Analysis (TA) Python Package and can be divide into five broad groups: momentum, volume, volatility, trend and others. As anticipated, the classification model was less biased towards smaller stocks. Additionally, it generated results with less volatility, maximum drawdown and kurtosis. Despite its lower riskiness it also outperformed the regression model on total returns, translating into a Sharpe Ratio of 1,95. Subsequently, they evaluate if refitting the model using a rolling window leads to better results than a static model. They find that a static model leads to a higher accuracy in parallel of keeping the ability to assess whether captured patterns remain valid over time. They also investigate which features are driving the model's decisions. To do so they implement Shapley Values (Shapley, 1953; Lundberg & Lee, 2017). The paper finds that trend-based features are the most important and interestingly volume-based features are less important for European equities compared to other regions. Additionally, companies expressing a lower average daily return, short term and to a lesser extend long volatility in the last fifty days have positive Shapley values. A factor that was found to very important is the difference between short- and long-term volatility, stocks relatively more volatile in the short term compared to their long-term trend are more likely to be classified as an outperformer. Lastly, the trend indicator Moving Average Convergence Divergence (MACD) has a strong impact, where stocks with relative stronger short-term trend compared to their long-term average are more likely to underperform. For this approach to translate into profitable investments it does not specifically require high accuracy, but rather the ability to significantly differentiate top and bottom performers. To verify this, they compute the decile portfolios and observe a significant return spread between the equal weighted top and bottom decile portfolios suggesting that this approach is appropriate for portfolio rankings. Furthermore, they find that the top decile outperforms the equal weighted reference benchmark. Leading to returns two to three times less volatile and lower maximum drawdowns, suggesting that his portfolio is more resilient in times of crises, as supported by a Sharpe Ratio (SR) three times higher (SR=1,81). However, the returns do not follow a normal distribution, as indicated by the negative skewness and high kurtosis, leading to tail risks which are not captured by SR. Therefore, the author follows the approach presented in Pezier (2004) to compute the Adjusted Sharpe Ratio (ASR) which accounts for those aspects. After this manipulation the top portfolios do not show any outperformance, suggesting that investors charge a premium for exposure to tail risk. Additionally, an equal weighted portfolio with a hundred long and short positions yields an annualised return of 40,88% and volatility of 15,97%, resulting in a SR of 2,49. This outperformance is not due to tail risk exposure as skewness is positive and kurtosis is very low. Consequently, leading to a high ASR of 4,1. Similarly value weighted portfolios also show strong results and ensure that the findings persist for larger stocks. Interestingly, those do not display any tail risk, concluding that tail risk exposure is mainly driven by smaller stocks rather than the investment strategy. To ensure robustness against any biases, the portfolios are compared to randomly constructed bootstrap portfolios and controlled for potential country overweighting, where no anomalies are observed. Lastly, when analysing exposure to the Fama French Six Factors model, it is observed that the portfolios load positively to the Market and Small Minus Big (SMB) factors, indicating that part of the high return are driven by small stocks. In parallel, negative exposure to Conservative Minus Aggressive (CMA) and Momentum can be seen. Nonetheless, the abnormal returns remain dominant with a significant alpha. Furthermore, they are the first to suggest using mean variance for portfolio construction using a classification model. They achieve this by subtracting 0,5 to the predicted probability of being part of class 1 and using this new value as a substitute for expected returns (μ) in the mean variance optimisation problem. They compare it to the traditional approach of using past returns as a proxy for future expected returns. An alternative approach which is argued in favour in the literature consists of investing in the minimum variance portfolio as this would not require any estimate for μ, however, doing so would lead to a trade-off between lower volatility and lower returns. The long only and long-short portfolio using outperformance probabilities show stronger results than the portfolios using past returns, characterised by higher returns, SR, ASR, and lower volatility and maximum drawdown. Mean variance portfolios built based on the classification model were less likely to take extreme positions, with a median weight for the largest component being 53,74% versus 80,42%. This is where this work aims to bring the biggest contribution by exploring if imposing constraints on the portfolio weights would lead to better risk adjusted returns. The model performs better during periods of stress (high volatility and bid ask spread). During periods of low volatility and bid ask spread investing in lower-ranked stocks does not generate any alpha. Additionally, lower alpha is observed in recent periods, but it cannot be concluded that this the result of markets getting more efficient but rather linked to the absence of major crisis recently, given that the performance is mostly driven by the short leg of the portfolio. Those results are robust to transaction costs, as even after accounting for them using the bid ask spread as a proxy, abnormal returns in the form of alpha are still persistent.

Wolff & Echterling (2024) further contribute by investigating more ML models for similar outperformance probability prediction objectives and include fundamental factors for the constituents of the S&P500. They explore a wide range of models including linear (PCA, Ridge, Lasso and ENet), tree-based (Random Forest and Boosting), deep learning (deep neural networks - DNN and long short-term memory - LSTM) and a combined forecast (simple average of the probabilities predicted by all the models referred to as “ensemble”). Their portfolio building strategy consist of going equally long in the k stocks with highest probability of outperformance. Surprisingly they observe that linear models perform similarly well compared to more complex deep learning models. Comparably to Breitung (2023) it is found that this approach is suitable for portfolio rankings as the 1/5/10% top/bottom have higher accuracy than the rest. ML strategies are found to be more volatile than the equal weighted benchmark, but they still achieve lower maximum drawdown suggesting that they show more resilience in case of extreme market crash. Resultingly, all ML strategies achieved a higher risk adjusted return (Sharpe Ratio) and generate economically significant abnormal returns (alpha). Nonetheless, this comes at the cost of a high turnover in particular for more complex techniques, which is in line with what Avramov et al. (2023) pointed out. The computation of the break-even transaction cost (BTC) confirms that trading costs must be low in order for the strategies to be profitable (0,05% for PCA, 0,21% for ENet and 0,17% for the Ensemble). The optimal portfolio size should be a trade-off between higher returns and volatility due to a smaller size or the opposite if more constituents are included. The ideal portfolio size using the Sharpe Ratio is found to be around 50, which represents being invested in the top 4,3% (50/1164). However, this is only applicable for similar strategies employing an equal weighted portfolio formation approach. Moreover, the ML strategies consistently outperformed the index in all economic cycles and at least performed as well as the benchmark. An important aspect to evaluate the trading strategies is their exposure to risk factors. To test for it the Fama French Six Factors model is used augmented by the betting-against alpha factor (BaB)(Frazzini & Pedersen, 2014). All the models logically load to the market factor, as it is a long only strategy. Additionally, PCA has significant positive exposure to HML and BaB, and negative to RMW. On the other side boosting and neural networks have only momentum as a significant factor. Most importantly after accounting for the different risk factors alpha remains for all models. Furthermore, they perform a 52-week rolling factor exposure analysis of the ensemble to identify the evolution factor exposure over time. High fluctuations suggests that the model does not follow a particular factor strategy but rather performs implicit factor timing. Finally, feature importance is assessed using Shapley values. It is found that technical indicators play a larger role, mostly due to the relatively short time horizon for predictions (1 week). Logistic models found momentum factors to be particularly important, whereas it was Relative Strength Index

(RSI) for neural networks and tree-based models accounted for both technical and fundamental features equally. Interestingly, tree-based models performed better with only fundamental data, contrasting with neural networks which did not seem improvements with its inclusion and rely solely on technical indicators.

3. Machine Learning Models

3.1 Regularisation

Regularisation techniques build upon the simple linear logistic regression for classification tasks. Its main goal is to prevent the risk of overfitting and improve generalisability to new data by curtaining the coefficient estimates within a certain range. In the presence of multicollinearity (correlated features), partly irrelevant parameters or many predictors, regularisation outperforms Ordinary Least Squares (OLS)(Wolff & Echterling, 2024). This is achieved by applying a penalty term on the coefficient estimates. Leading to include the penalty and the residual sum of squares in the minimisation problem. This prevents the model from giving too much importance to one feature and overfit. If predictors are not on the same scale penalisation is not equally applied, therefore, it is necessary to first normalise the data. Hoerl & Kennard (1970) introduced Ridge regression as a regularisation technique which performs linear logistic regression with a penalty factor equal to the square of coefficients (L2). The penalty function is expressed as follows:

$$\min_{\beta} \sum\nolimits_{x,y} log(1 + exp(-\beta' x * y)) + \lambda \beta' \beta$$

Where λ determines the level of shrinkage towards 0 (higher values of λ bring all coefficients closer to 0). However, Ridge coefficients never reach exactly 0. Lambda is the sole hyperparameter that requires tuning. Its optimal value is decided based on cross validation. Another regularisation technique is Lasso (Tibshirani, 1996). The difference is that it uses as a penalty term of the absolute coefficient estimates (L1). This approach leads Lasso to shrink coefficients to 0, essentially performing feature selection and leading to higher model interpretability. However, this introduces its shortcoming, as in case of several correlated variables Lasso will only select one and ignore the others. Both models come with different coefficient estimates, nonetheless, their performance in most cases is extremely similar. Here is the equation of the penalty function for Lasso:

$$\min_{\beta} \sum\nolimits_{x,y} log(1 + exp(-\beta' x * y)) + \lambda|\beta|$$

Finally, Zou & Hastie (2005) suggested combining both models into the "Elastic Net" (ENet). The resulting approach encapsulates both penalty functions (L1 & L2). ENet outperforms Lasso, while also being sparse (shrinking coefficients to 0). Additionally, it introduces a new hyperparameter, as now in addition to selecting the shrinkage coefficient it is also necessary to determine what fraction should be given to each penalty function denoted by α, referred to as the L1 ratio.

$$\min_{\beta} \sum\nolimits_{x,y} log(1 + exp(-\beta' x * y)) + \lambda \left(\alpha|\beta|_1 + \frac{1-\alpha}{2}\beta_2^2\right)$$

3.2 Random Forest

Tree-based models such as Random Forest (RF) apply a search technique of "divide and conquer learning". Essentially the model divides the data into subgroups for which it tries maximising intra-group homogeneity regarding the target variable (Kubat et al., 1998). Breiman et al. (1984) introduced an algorithm that selects the splits in a top-down fashion. The feature that allows the highest gain in homogeneity is chosen. The tree continues to grow sequentially adding new variables for splitting and groups are becoming more homogenous. The tree stops growing once all the groups are perfectly homogenous or that it has reached one of the imposed limits (maximum depth/nodes or minimum impurity decrease for a split to happen). Those limits are some of the hyperparameters that can be specified for Random Forest. A single tree usually provides a very poor performance as it tends to overfit and only remember observations from the training data, yielding no predictive ability. A solution to overcome this problem was suggested by Breiman (1996), the approach consists of combining multiple decision trees which are each built on a different random subsample of the training data using a bootstrap aggregation method. Finally, the predictions of all the tree are aggregated. This approach alleviates the problem of overfitting. Additionally, Breiman (2001) suggested also applying bagging to the features. This fraction is usually the square root of the number of total features, and this is also the case in this paper. This improved variability of the trees, leading to better results and avoiding relying on a single predictor.

3.3 Light GBM

Light GBM is a tree-based boosting algorithm introduced by Microsoft (Ke et al., 2017). Gradient boosting algorithms try to minimise the prediction error using gradient descent by subsequently adjusting the new tree to correct for the errors made by the previous one. This is achieved by overweighting mislabelled observations. A popular example is XGBoost. However, despite its good performance, efficiency and scalability in high dimension and for large datasets is unsatisfactory. This is because for each split the algorithm scans all the features at every possible split point to estimate the information gain. Light GBM presents novel techniques to speed this process. The first one is to bin continuous features. This significantly reduces memory usage and allows for faster split finding. Secondly, its most unique aspect is that it grows trees leaf-wise instead of level-wise. This leads to deeper and unbalanced trees which fit the data better and obtain higher accuracy. However, this increases significantly the risk of overfitting and therefore, hyperparameter tuning is extremely important for Light GBM. Thirdly, it uses Gradient-Based One-Side Sampling (GOSS). It does not use all the samples but rather keeps only large gradients and randomly samples a share of those with small gradients. This speeds computation without much accuracy loss. Finally, it applies Exclusive Feature Bundling (EFB). This is particularly efficient for sparse (many zero values) high dimensional data. The algorithm will combine features that are considered mutually exclusive (rarely take zero at the same time) essentially making them a single feature and thus reducing the dimensionality.

3.4 Ensemble

An ensemble refers to the combination of the predictions made by base learners. Its most simple approach is to simply average out the forecasts made by the underlying models. This technique although naïve as it does not try to overweight a better performing learner still achieves in a lot of cases the best results (Wolff & Echterling, 2024). Krauss et al. (2017) finds that their ensemble models perform better than individual models (DNN, GBT, RAF). This is because they are less prone to overfitting, more robust and individual models can get stuck in local minima but by combining several predictors it leads to a better approximation of the true unknown function. For an ensemble to produce good results its base learners need to fulfil two conditions. First, they need to be diverse, meaning that their errors should not be strongly correlated. Secondly, they have to be accurate. Bella et al. (2013) implements a weighting scheme induced by the training Area Under the Curve (AUC) to form a performance-based ensemble. It tries to overweight models which are perceived as more performant to generate better overall performance for the ensemble.

3.5 Considerations

Other possible ML models were considered for this work, in particular neural networks given their extensive use in other reference papers and their increased performance. The models considered were a simple feedforward neural network and Long Short-Term Memory (LSTM) for its ability to account for time dependencies. Those were ruled out because as found in Wolff & Echterling (2024) for the task at hand they perform similarly to the models that were included but would have come at a significantly higher computational cost.

4. Macroeconomic Variables

Beyond firm level signals, macroeconomic factors also shape asset returns. This paper follows the recommendation of Breitung (2023) for future research to include those in the context of outperformance prediction. Avramov et al. (2023) in their study already included some, such as: index dividend yield, PE ratio term spread, US T-bill (3 months bond) and default spread.

Flannery & Protopapadakis (2002) investigate the impact of macroeconomic announcements on the US stock returns and volatility. They observe that volume significantly ramps up on the day of announcements, suggesting that investors do monitor closely changes in the economy. In particular, inflation (measured through consumer price index (CPI)), producer price index (PPI), money supply (M1/M2), balance of trade (BOT), employment reports, housing starts, US 3 months and 10 years bond yields were found to be significant. Money supply (M1) affects both returns and volatility. Additionally, when CPI, PPI and M1 announce higher than expected values this leads investors to degrade their prognostics. Only a few macroeconomic factors matter, particularly inflation and M1, but most broad economic indicators have very little effect.

Boons (2016) further research whether macroeconomic state variables can explain cross-sectional variation in individual stocks. For their analysis the index dividend yield, government bonds at different maturities, index volatility, sentiment, term spread, and credit spread are included. Interestingly, index dividend, term spread, and credit spread covary significantly with certain firm characteristics (ex: high book to market or small size) thus explaining a part of the cross-section of individual stock returns. This is especially relevant during economic downturn, as the model performs better.

Celebi & Hönig (2019) observe similar findings when looking at the German stock market, DAX. They analyse the relationship between macroeconomic factors before, during and after crises. They find that before economic downturn macro indictors are less relevant. Long term rates (10Y German Bund), M1, M3 and business sentiment are all found to be significant during and after crises. Short term rate (3M Euribor) is the only exception displaying significance during all economic cycles.

Finally, Ferrer et al. (2016) confirm the strong sensitivity of European equities to interest rates and monetary policies. Interest rates are a key driver of economic activity as they reflect market expectations about the future of the economy and determine the cost of borrowing. In particular, long-term rates have a critical influence on investment decision and profitability of companies, directly impacting their stock performance. Moreover, government bonds with longer maturities are viewed as closer substitutes to equities, further increasing the degree of linkage between the two. The paper finds that interest rates have an insignificant influence for the short term (<8 days), strong negative impact at medium term (16-32 days) and somewhat a weaker connection at a long horizon (>32 days). As expected, interest rate hikes are linked to negative stock returns as they directly raise the cash flow discount factor lowering asset prices. Interestingly, the relationship is found to be stronger and more stable for France and Germany, and exhibit a weaker pattern for southern countries (Portugal and Spain). This relationship is found to be strengthening since the post 2008 financial crisis. Additionally, as argued by the European Central Bank (2014) German government bonds, referred to as Bunds, are a relevant proxy for the risk-free interest rate in the eurozone.

III. Data and Methodology

1. Data

The data collected for this study comes from multiple sources and covers the timeframe from October 2004 until March 2025 with weekly observations. The stocks included for analysis are the historical components for that period of the Stoxx 50, an index tracking the 50 largest market capitalisations of the Eurozone. This is to ensure that those are large and liquid companies, meaning that the signals are exploitable and not subjected to hard to arbitrage barriers as it is often the case for machine learning strategies as pointed by Avramov et al. (2023). Companies that were not traded for the entire period (IPO after October 2004 or delisted before March 2025) or that were no longer traded on a stock exchange of the Eurozone have been excluded (relisting in the US or UK). This results in 64 individual companies listed in the Appendix D.

The main source of data is the Bloomberg Terminal, it was used to retrieve fundamental, basic technical and macroeconomic data (Bloomberg L.P., 2025). The data collected spanned from January 2004 until March 2025 for data cleaning purposes as explained later in the methodology. Most of the advanced technical indicators were calculated using the Python package "Technical Analysis (TA)" using the variables: "Volume", "High Price", "Low Price", "Opening Price" and "Closing Price", following the procedure in Breitung (2023). Indicators that showed to be relevant for the task at hand based on Breitung (2023) and other academic references were selected. Technical measures that were not scale free were made relative using the approach in Wolff and Echterling (2024) to ensure better comparability and transferability of learned patterns for the model.

The Kenneth R. French Data Library was used to acquire the daily values for the Fama French Five Factors for European Equities. The dataset includes the daily returns for the Market (Mkt-Rf), Small Minus Big (SMB), High Minus Low (HML), Robust Minus Weak (RMW) and Conservative Minus Aggressive (CMA) factors, in addition to the risk-free rate (Rf). These daily factor returns were compounded into weekly returns to match the rest of the data (French, n.d.).

The ECB data portal was used to access monthly data about the inflation rate measured by the Harmonised Index of Consumer (HICP) and the monthly change in monetary aggregate M1/M2/M3 (European Central Bank, 2025). Moreover, the composite business and consumer confidence indicators were both retrieved from the OECD Data Explorer (OECD, 2025). All measures were lagged by one month to account for the delay in publication.

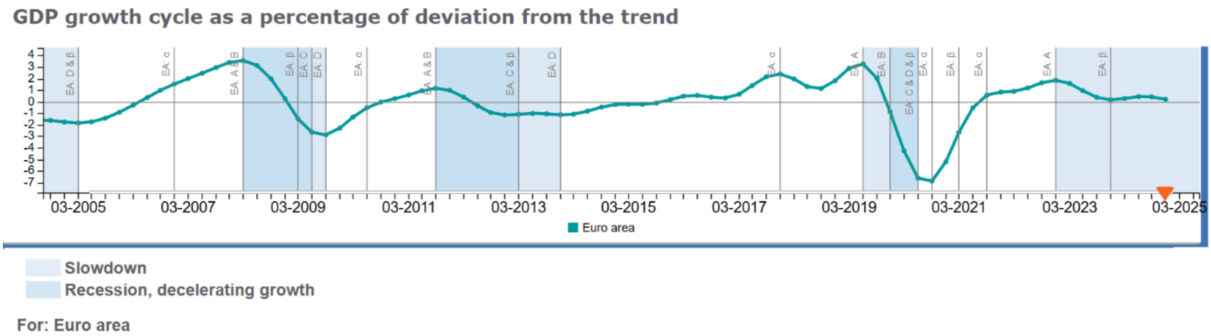
Additionally, the Euro High Yield Index Effective Yield was retrieved from FRED, Federal Reserve Bank of St. Louis. This index tracks the bond yield of euro denominated high yield companies, which refers to debt issued by companies with a below investment grade credit rating based on an average of the ratings from Moody’s (<BBB), S&P (<BBB) and Fitch (<Baa3)[3,4].

Lastly, the European Business Cycle Clock published by Eurostat is used to identify the different economic cycle phases for the euro area. It is calculated by combining three synthetic indicators: the growth cycle coincident indicator (GCCI), the business cycle coincident indicators (BCCI) and the acceleration cycle coincident indicator (ACCI). The output indicates if the euro area is in one of the six possible states which are: expansion with decelerating growth, expansion with acceleration growth, recovery, slowdown, recession with accelerating growth and recession with decelerating growth. Those measures are furthermore aggregated at a higher level into three categories: expansion, slowdown and recession (Eurostat, 2025). Those will be the classification used throughout this paper.

All the features included in this study are listed under the Appendix E with the Bloomberg field code, Python function or manipulations used to obtain them. Additionally, the literature based on which their selection was made is indicated.

[3] (Chen, 2025)
[4] (What to Know Before Saying Hi to High-Yield Bonds, 2023)

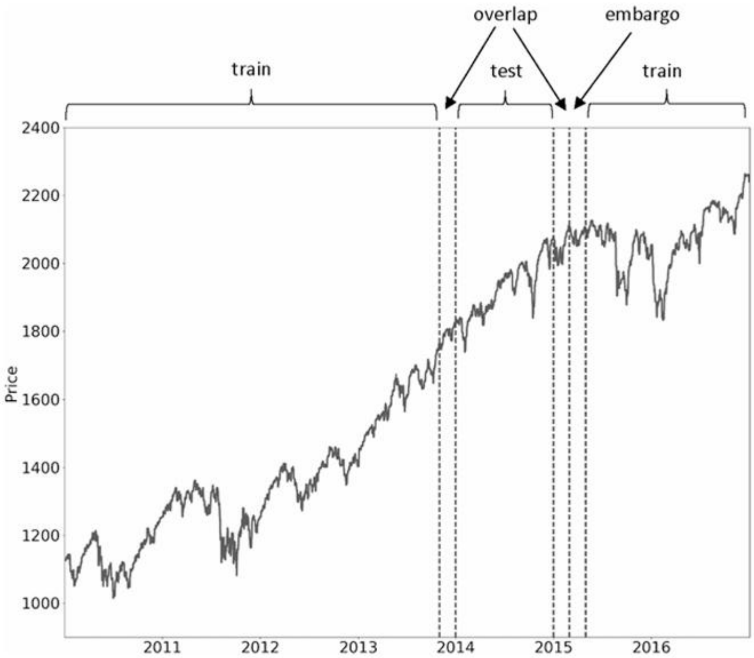
Figure 2: *Economic Cycles as Measured by the Business Cycle Clock*



Note. Eurostat. (2016, December 9). *Business cycle clock.* https://ec.europa.eu/eurostat/cache/bcc/bcc.html

The data ready for analysis spans from October 2004 to March 2025. The dataset is further divided into three sets: training (October 20004 to December 2016), validation (January 2017 to December 2018) and testing (January 2019 to March 2025). Those splits were selected to ensure that both the training and testing set included all the economic cycles as measured by the Business Cycle Clock (Eurostat, 2025), so that the model could learn all these different patterns and testing could happen in all the economic conditions. The validation set is used as a replacement for the testing set while design choices are being actively made, once everything is arranged it will be merged with the training and the entire model will be retrained on the combined data before performing predictions on the test data. This is to ensure that no information from the testing set is leaked ahead. Hyperparameter tuning will be based on cross-validation. The specific technique used, which is standard in the finance machine learning industry, is Purged 5-Fold with Embargo. This approach is presented as most suitable by López de Prado (2018), this is because standard K-Fold performs poorly on financial data and can lead to overfitting. This happens because normal K-Fold will randomly shuffle observations into 5 groups (4 training and 1 testing, which are switched at each iteration), meaning that it will be asked to predict features at time t while being partly trained with data from t+1. This leakage can lead to inflate cross validation results and overfit to irrelevant features. To address this problem purged K-FOLD with embargo divides the training set into 5 sequential time components where overlapping information is purged from the training blocks. Additionally, the observations that comes directly after the testing set can also leak information. Therefore, an embargo is implemented only for training observations directly after the testing set with a size set at $h \approx 0.01T$. This ensures that the model operates in a more realistic setting and avoids overfitting when choosing the optimal hyperparameters. Five splits is deemed sufficient as Bergmeir & Benítez (2012) suggests that any value above would not yield significantly better results, but still would come at the expense of higher computational needs.

Figure 3: *Example of Purged K-Fold with Embargo for One-Fold*



Note. Figure from López de Prado (2018) page 108.

2. Methodology

2.1 Data Cleaning

The first and substantial part of this work is to ensure the data is fit prior to any further analysis. Due to the various data sources and types at hand extensive cleaning was required. To begin with as fundamental data is updated infrequently at every reporting deadline all the intermediary weekly observations were marked as missing values by Bloomberg. The financial reports of the considered companies are published on a half-year or quarterly basis, therefore, missing values for only features expressing this update pattern are imputed using forward fill for a maximum period of 27 weeks. This aspect leads to the insertion of rows which do not represent a valid weekly observation but rather the publication of fundamental data. For this reason, after the first step is executed all observations which do not represent a Friday (end of the trading week) are removed to ensure that only rows at weekly interval remain. Once the operations above are finished it is possible to start computing new variables. Momentum returns are computed for the 52, 40-, 26-, 10- and 4-week intervals as well as the weekly returns. For the latter lag 1 and 2 values are also included. The above measures were computed using the last price (PX_LAST) to compute capital appreciation gains. For trading strategies dividend gains are irrelevant as it requires the model to hold to a position long enough for it to be eligible for the payout (ex-dividend date) and according to dividend irrelevance theory the dividend payout is associated with a drop in price as investors reevaluate the new company value (deducted from payout cash)(Modigliani & Miller, 1958).

The feature “BID_ASK_SPREAD%” may be containing many missing values because Bloomberg only started explicitly measuring it later, nonetheless it can be accurately filled using other variables. As a result, for all observation with missing values for “BID_ASK_SPREAD%” those are retrieved by applying the formula below:

$$\frac{PX\_ASK - PX\_BID}{PX\_LAST} = \text{BID\_ASK\_SPREAD\%}$$

Moving forward all the technical indicators that were identified as relevant by Breitung (2023) and other academic papers are obtained using the Technical Analysis (TA) Python Package. The package only requires five inputs: opening, closing, highest, lowest prices and volume, and can return a broad array of sophisticated technical indicators. Additionally, to ensure comparability and transferability of the learned patterns, indicators which are scale specific are standardised following the procedure in Wolff & Echterling (2024). For further details please refer to table 25 under the Appendix E.

Now macroeconomic variables can be left joined on dates. Furthermore, key features identified by Breitung (2023) are computed on a 26 period rolling basis (half a year): average weekly returns, weekly returns skewness, weekly returns kurtosis, max spread between maximum and minimum weekly returns, Sharpe Ratio and Adjusted Sharpe Ratio. After the first stage of data cleaning observations before the 23/08/2004 are cutoff as they were only used for the data cleaning purposes (forward fill) or to compute variables which need significant lookback periods (technical indicators).

Finally, all remaining missing values can be imputed using a Rolling MICE Forest approach. This technique builds on Miceforest which is a random forest version of Multiple Imputation by Chained Equations (MICE) suggested by (Shah et al., 2014). However, it has a severe problem as it does not consider temporality and thus all observations from time t →T would be used for imputation. Financial data is usually in the form of time series, using posterior values to make imputation causes the risk of leaking future information and causing a "look-ahead bias" (Freyberger et al., 2022). This could lead the model to overfit as it captures this foreshadowing rather than genuine signal. To solve this Stam (2022) introduced a new rolling window approach to Miceforest, which would work like its predecessor to the exception that the model would only be fitted on observations that occurred before the imputed values, in this way preserving the spatial integrity. Stam (2022) finds that proper treatment of missing financial values significantly affects portfolio construction and asset pricing. In this study Rolling MICE Forest was used with 100 estimators (trees) and median as the initial strategy. Additionally, for values that remained missing a median and zero fallback option were implemented which respect the chronological nature of financial data. The median imputation was calculated based on the observations leading to t and excluding future datapoints.

Lastly, new variables are created including the binary target variable. For every security at time t a value of 1 is attributed if at time t+1 its weekly return is above or equal to the median weekly return at time t+1 of all the considered companies. This ensures class balance. Other variables which were found to be important by previous research and could not be directly retrieved from Bloomberg are computed. These include: Net Income to Market Cap, Cash From Operation to Market Cap, difference in volatility between short and long term (30D Vol – 36D Vol) and log of standardised moving averages, following the procedure in (Wolff & Echterling, 2024). Observations before 1/10/2004 are further dropped as they were mostly used to train the Rolling MICE Forest in its first iterative rounds. For more details, the complete list of all the variables including the manipulation performed can be found under the Appendix E.

Figure 4: *Median Weekly Returns*

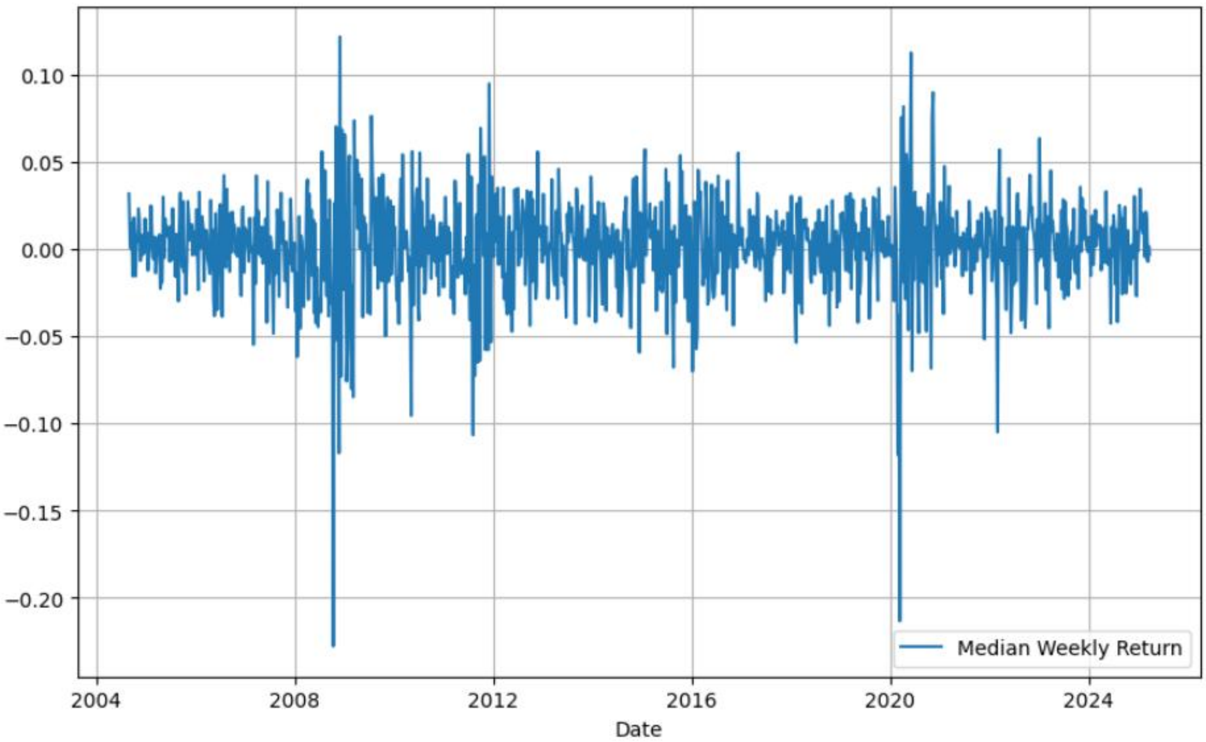


Figure 5: *Cumulative Median Return*

2.2 Machine Learning Models

The machine learning models will be used in this work for a binary classification task. During the data preparation each stock at every time step was labelled 1 or 0 depending on if it outperformed or not the median returns for the subsequent period. Using the training data the models will try to determine the relationship between the features and the dependent variable. As the task at hand is a binary classification problem the models will output a probability of belonging to class 1, expressed as a value between 0 and 1. Predictions above the threshold of 0,5 are then classified as label 1 “outperformer” and otherwise labelled 0 “underperformer”. The models used for this work are as follows Lasso, Ridge, Elastic Net (ENet), Random Forest (RF) and Light GBM. They were selected for their relative ease of implementation and their good performance (Wolff & Echterling, 2024). Additionally, two ensemble models are assembled. Those combine the predictions made by the base learners and usually perform better when the base learners are diverse enough. This is the reason why both linear and tree-based models are considered. The first Simple Ensemble (SE) will simply take the average of the predicted probabilities. The second, a Performance-Based Ensemble (PBE) will take the weighted average by training AUC of those predicted probabilities, following the procedure presented by Bella et al. (2013). The study will mostly focus on the predictions of those two ensembles.

2.3 Hyperparameter Tuning

An important aspect of machine learning is hyperparameter tuning. These allow to steer the model in the right direction to obtain the desired outcome and most importantly avoid overfitting. As mentioned previously to evaluate the combinations of hyperparameters Purged 5-Fold with Embargo is used, as it reflects a more realistic setting and avoids inflating the cross-validation score. Traditionally, two techniques were employed to find the best combination of hyperparameters, either Grid Search or Random Search. The first, one would go through every possible association and keep the one with the highest cross-validation score. On the other side Random Search would just pick random combinations, speeding the process and requiring substantially less computational needs, but nonetheless providing as equally or even better models (Bergstra & Bengio, 2012). However, now there is a third option which is Bayesian Hyperparameter Search. It merges the speed and computational efficiency of Random Search with the exhaustivity of Grid Search. What it essential tries to achieve is map the hyperparameter search space using probability densities for good and bad scores. This allows the search to adaptively learn which hyperparameters lead to better performance and focus on those instead of wasting trials on bad combinations. This approach is implemented with the Python Package “Optuna” (Akiba et al., 2019). The number of trials is set to 200 for linear models, 400 for Random Forest and 800 for Light GBM. These values are decided based on the dimension of the hyperparameter space. All the hyperparameter search space, selected values and Optuna plots can be accessed in the Appendix B.

2.4 Portfolios

For this study different portfolios will be constructed for analysis. The benchmarks will be comprised of the Stoxx 50 index and an equal weighted (EW) portfolio of all the 64 considered companies. The first offers directly a point of comparison against the returns that would have been obtained by investing in the index. The second is a common benchmark portfolio, as the works of DeMiguel et al. (2009) uncovered how difficult it is to beat this seemingly naïve approach in terms of risk adjusted returns. This is partly explained by the fact that this strategy exposes itself more to small cap stocks, which were found to perform better on average.

The portfolios formed for analysis will be based on the outputs of the machine learning models. They will be rebalanced at each time step t and carried to t+1. First, the value weighted (VW), based on market capitalisation, and equal weighted (EW) will be constructed. They will allow to compare portfolio forming strategies based on simpler methods versus mean variance, this will tell how much of the potential gains can be attributed to the portfolio formation technique or the machine learning predictions. Secondly, the mean-variance portfolio will be constructed using the probabilities of outperformance subtracted by 0,5 as a proxy for expected returns and the covariance matrix will be computed using a 52-week past window. For each portfolio type different variants will be introduced. Long only vs long short, allowing to compare if the short leg is also driving most of the returns and constrained vs unconstrained to test if introducing limits to the portfolio weights can result in better risk adjusted returns.

Finally, for all the portfolios detailed above three pools of assets will be considered. The top 5 (7,81%) or 10 (15,63%) stocks with the highest probability of outperformance at each time step and the bottom 5 or 10 for long-short portfolios. This is because as highlighted by Wolff & Echterling (2024) the highest/lowest predictions are more accurate and generate more economic gains. However, a more concentrate portfolio is a trade-off between risk and returns, therefore, those different level of investments in the top/bottom assets will allow to identify where the optimal balance is. Lastly, all assets will also be considered, where for long only portfolios that will only be assets labelled 1, which are the assets with a probability of outperformance superior or equal to the median at each time step.

2.5 Evaluation Metrics

2.5.1 Models

Different evaluation metrics will be used to assess the performance of the machine learning prediction models. The first measure which is standard for classification tasks is accuracy. It measures how many observations were correctly labelled over the total observations. Additionally, this measure will also be applied for the 5 or 10 highest or lowest predictions. For instance, it looks at 5 stocks with the highest predicted probability of outperformance at every time step and determines the accuracy of the labels assigned to those specific observations. This will give important insights into whether this approach is appropriate for portfolio rankings as observed in previous studies. It is important to note that what is considered good accuracy depends on the context. For financial data good predictive ability is very difficult to obtain as any strong signal would be directly identified and traded upon closing the temporary arbitrage opportunity. In their paper Wolff & Echterling (2024) obtain goods results with accuracies ranging from 50,43% to 50,81%. Therefore, anything in that range or above will be considered satisfactory. As a reminder random guessing or labelling every observation with either 1 or 0 should yield an accuracy of 50%.

The second metric used is the F1 Score. It is an accuracy measure that balances precision and recall. Which are respectively how many of the predicted positives are actually positive and how many of the actual positive did the model correctly identify. It is an important metric which allows to verify that the model does not overly focus on predicting a particular class at the detriment of the other.

The third measure is the Area Under the Curve (AUC) of the Receiver Operating Characteristic curve (ROC). It plots the true positive (recall) versus the false positive rate (FDR = 1 – Specificity[5]). The returned value is between 0 and 1, where 1 represents perfect

[5] Specificity measures how good the model is at identifying the negative class correctly. How many of the actual negative class have been correctly predicted by the model.

classification, 0,5 random guessing and anything below worse than guessing. It's a useful measure that mostly cares about the ranking of the predictions, by evaluating how well the model separates the two classes regardless of the classification threshold.

Finally, McNemar's p-value will be calculated. It is a statistical test which tests whether the accuracy is statistically above the naïve threshold of 50%.

2.5.2 Portfolios

Financial measures are meaningless if they are not contextualised against a benchmark. Therefore, all the metrics selected to evaluate the portfolios will be assessed not on an absolute basis, but rather compared to the two benchmarks. Firstly, returns will be considered in three manners: total, annualised and weekly average returns. The latter will substitute the earlier measures for performance measurement during cycles analysis, as those are inaccurate in that context. Second, volatility measured as the standard deviation of portfolio returns will be used as a proxy for risk, as it is common in financial literature. Thirdly, the Sharpe Ratio will be considered. It is a risk adjusted metric expressed as follows:

$$Sharpe\ Ratio = \frac{\left(\overline{R_p} - \overline{R_f}\right)}{\sigma_p} * \sqrt{52}$$

Where $\overline{R_p}$ is the average portfolio returns, $\overline{R_f}$ the average risk-free rate, σ_p the standard deviation of portfolio returns and $\sqrt{52}$ used to annualise the weekly measures (Sharpe, 1966). However, a big shortcoming of the Sharpe Ratio is its assumption that returns follow a normal distribution. Consequently, it can lead to overoptimistic values although the portfolio is exposed to severe tail risk. Resultingly, Pezier (2004) suggest an adjusted version to account for the skewness and kurtosis of returns:

$$ASR = SR(1 + \frac{S}{6}SR - \frac{(K-3)}{24}SR^2)$$

Where "ASR" is Adjusted Sharpe Ratio, "SR" is Sharpe Ratio, "S" is skewness and "K" is kurtosis (normal value 3). Fourthly, maximum drawdown will be included. It is a common financial metrics used to measure downside risk. It expresses the largest drop from a peak to a through in a portfolio's value before a new peak is achieved. Fifthly, turnover which refers to the trading activity required to rebalance the portfolios will be calculated for all the investing strategies. Avramov et al. (2023) pointed out that machine learning based strategies usually rely on a large turnover, which after accounting for fees can erode significantly returns. Average turnover is calculated as follows:

$$\tau = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{2}\sum_{i=1}^{N}\left|w_{i,t} - w_{i,t-1}\right|$$

Where T is the number of periods, N is the number of considered assets and $w_{i,t}$ is the weight of asset i at time t. Additionally, Wolff & Echterling (2024) presented an approach to evaluate the robustness of a trading strategy to fees by computing the Break-Even Transaction Costs (BTC) against a benchmark. It essentially measures the maximum transaction fee per unit of turnover before underperforming the benchmark.

$$BTC = \frac{\overline{R_p} - \overline{R_b}}{\tau}$$

Where $\overline{R_p}$ is the average portfolio returns, $\overline{R_b}$ is the average benchmark returns and τ the average turnover. It is important to note that in this study the unit of turnover is for 100% turnover and the benchmark against which it will be compared is the equal weighted portfolio given its status of hard to beat. Lastly, the concentration of the portfolios will be measured using the Herfindahl–Hirschman Index (HHI). The value range is between 0 and 1, measures on the lower end represent less concentration and vice e versa for values closer to 1, with 1 being a complete monopoly. It is calculated by summing the squared weights for the different assets:

$$HHI = \sum_{i=1}^{n} w_i^2$$

2.5.3 Risk Factors

As presented previously exposure to risk will be measured using the Fama French Five Factor model. Additionally, it allows to capture any abnormal returns, alpha, not explained by exposure to risk factors. This assessed by regressing the returns of the different risk factors, as provided by the Kenneth R. French Data Library (French, n.d.), on the portfolio excess returns (portfolio returns minus the risk-free rate). The regression techniques employed is Ordinary Least Squares (OLS) and the importance of the factors is assessed based on their estimated coefficient and associated p-value. Alpha is capture in the intercept given the equation manipulation. Its significance is also assessed based on its p-value.

Moreover, in their study Wolff & Echterling (2024) implemented a novel approach to measure the evolution of risk factors exposure. This is done by computing the rolling factor regression with a past window of 52 weeks. It essentially performs a regression at every time step t, regressing on the observation between t and t-51. This is in order to identify if factor exposure is stable or does it fluctuate over time and economic cycles.

2.5.4 Feature Importance

Finally, Feature Importance will be evaluated. It will provide insights into the relevance of including macroeconomic variables which was recommended by Breitung (2023) and is one of the primary concerns of this paper. This will be done by calculating the Shapley values. Those are based on the work of Shapley (1953) in game theory, which determine the contribution of each participant to the game's outcome. Lundberg & Lee (2017) adapted the concept to machine learning by suggesting SHAP (SHapley Additive exPlanations). It essentially considers each feature as a participant and evaluates its contribution to the model's output. A higher value suggests a higher contribution and thus a higher importance.

IV. Results

This section outlines the empirical results of this work. Section 1 analyses the classification accuracies, the correlation of predictions and their evolution across time or economic cycles. Section 2 will examine the economic results of the assembled portfolios against the benchmarks. Mean variance portfolios with constraints will be included to determine if those can lead to better risk adjusted returns and what is the optimal value. Additionally, all portfolios will be examined during the different economic cycles. Section 3 dives into the portfolio's exposure to risk factors and whether they can explain the results obtained. Finally, section 4 will look at feature importance and the relevance of including macroeconomic variables. Further results are available under the Appendix C.

1. Prediction Results

Table 1: *Model Accuracies*

Model	Accuracy	F1 Score	AUC	McNemar P-Value	Top 5 Accuracy	Bottom 5 Accuracy	Top 10 Accuracy	Bottom 10 Accuracy
Lasso	.5135	.5163	.5168	.0001***	.5305	.5298	.5295	.5182
Ridge	.5130	.5167	.5167	.0001***	.5311	.5305	.5283	.5185
ENet	.5134	.5162	.5168	.0001***	.5305	.5298	.5295	.5182
Random Forest	.5058	.5313	.5140	.0474**	.5342	.5218	.5194	.5228
Light GBM	.5119	.5345	.5111	.0003***	.5151	.5034	.5252	.5052
Simple Ensemble	.5121	.5339	.5146	.0003***	.5323	.5311	.5255	.5077
Performance Based Ensemble	.5120	.5343	.5135	.0003***	.5360	.5243	.5237	.5077

Note. *Significant at the at the 10% level, **Significant at the 5% level and *** Significant at the 1%* level.

Beginning with the analysis of the predictions accuracies and associated measures for the different models. Surprisingly, the best performing models are the linear ones, with Lasso achieving the highest accuracy and AUC. Light GBM also does well achieving the highest score on balanced metrics such as F1 Score. Overall, all models achieve a good performance in comparison with previous studies suggesting that they were successfully able to identify signals. This is confirmed with all the models having accuracies significantly higher than 50% at the 1% level, except for Random Forest which is observed to be performing the worse. Nonetheless, it achieves a 5% significance level for its accuracy and does comparatively well on other metrics, particularly accuracy for the top 5 were it scores the highest suggesting a good ability for ranking.

The Performance Based Ensemble was weighted as follows: 16,25% for each linear model, 21,05% for Random Forest and 30,2% for Light GBM. Those weights are implied from the training AUC, the values displayed above are the test AUC. Despite the slightly different weights each Ensemble gives to its base learners both perform sensibly the same.

Finally, the relatively high accuracies for top/bottom observations supports the fact that this strategy is suitable to construct ranked portfolios.

Table 2: *Correlations of Predictions across Different Models*

	Lasso	Ridge	ENet	Random Forest	Light GBM	Simple Ensemble	Performance Based Ensemble
Lasso	1						
Ridge	.9941	1					
ENet	.9999	.9942	1				
Random Forest	.3002	.2994	.3002	1			
Light GBM	.1895	.1893	.1894	.2860	1		
Simple Ensemble	.3784	.3780	.3783	.3723	.8081	1	
Performance Based Ensemble	.3002	.3000	.3001	.3449	.8861	.9212	1

In general, all the models are relatively uncorrelated, checking the diverse base learner criteria and translating into better performance for the Ensembles. However, as expected the three linear models are highly correlated, particularly Lasso and ENet. ENet shares many characteristics with Lasso, but given the tuned l1 ratio of 0,24 it would have been expected for the model to perform more similarly to Ridge rather than Lasso.

Table 3: *Prediction Accuracies over the Years*

	Lasso	Ridge	ENet	Random Forest	Light GBM	Simple Ensemble	Performance Based Ensemble
2019	.5195	.5198	.5195	.5294	.5358	.5403	.5400
2020	.5012	.5006	.5009	.5006	.4898	.4886	.4865
2021	.5130	.5127	.5130	.5056	.4956	.4991	.4971
2022	.5114	.5111	.5114	.5108	.5210	.5201	.5231
2023	.5096	.5078	.5096	.5030	.5252	.5204	.5204
2024	.5201	.5204	.5201	.4991	.5057	.5048	.5069
2025	.5391	.5378	.5391	.4466	.5052	.5091	.5039

When looking at the evolution of prediction accuracies for the models over the years it is hard to spot any patterns. This is consistent with the finding of Breitung (2023), that refitting the model is unnecessary and that a static one performs better. This is also in line with the observation that markets tend to be more efficient after 2001 but since no particular trend can be noticed.

Table 4: *Model Accuracies during Expansion Cycles*

Model	Accuracy	F1 Score	AUC	McNemar P-Value	Top 5 Accuracy	Bottom 5 Accuracy	Top 10 Accuracy	Bottom 10 Accuracy
Lasso	.5064	.4666	.5090	.1115	.5333	.5236	.5278	.4965
Ridge	.5061	.4657	.5090	.1238	.5347	.5250	.5271	.4965
Elastic Net	.5063	.4665	.5090	.1155	.5333	.5236	.5278	.4965
Random Forest	.5092	.5244	.5193	.0392**	.5444	.5347	.5299	.5312
LightGBM	.5067	.5146	.5090	.1001	.5111	.5125	.5333	.5035
Simple Ensemble	.5081	.5094	.5107	.0603*	.5306	.5444	.5243	.5167
Performance Based Ensemble	.5076	.5112	.5104	.0738*	.5278	.5333	.5299	.5083

Note. *Significant at the at the 10% level, **Significant at the 5% level and *** Significant at the 1%* level.

Table 5: *Model Accuracies during Slowdown Cycles*

Model	Accuracy	F1 Score	AUC	McNemar P-Value	Top 5 Accuracy	Bottom 5 Accuracy	Top 10 Accuracy	Bottom 10 Accuracy
Lasso	.5179	.5522	.5240	.0002***	.5265	.5432	.5303	.5342
Ridge	.5172	.5535	.5239	.0003***	.5265	.5432	.5290	.5348
Elastic Net	.5179	.5522	.5240	.0002***	.5265	.5432	.5303	.5342
Random Forest	.5059	.5404	.5115	.1201	.5355	.5123	.5155	.5161
LightGBM	.5168	.5456	.5142	.0004***	.5226	.5058	.5161	.5097
Simple Ensemble	.5157	.5489	.5197	.0009***	.5316	.5252	.5245	.5071
Performance Based Ensemble	.5165	.5483	.5179	.0005***	.5381	.5277	.5135	.5110

Note. *Significant at the at the 10% level, **Significant at the 5% level and *** Significant at the 1%* level.

Table 6: *Model Accuracies during Recession Cycles*

Model	Accuracy	F1 Score	AUC	McNemar P-Value	Top 5 Accuracy	Bottom 5 Accuracy	Top 10 Accuracy	Bottom 10 Accuracy
Lasso	.5258	.5394	.5235	.0186**	.5385	.4846	.5346	.5423
Ridge	.5264	.5397	.5235	.0165**	.5385	.4846	.5308	.5423
Elastic Net	.5258	.5394	.5235	.0186**	.5385	.4846	.5346	.5423
Random Forest	.4862	.5134	.5014	.8754	.4692	.5077	.4846	.5154
LightGBM	.5108	.5707	.5067	.1954	.4923	.4385	.5346	.4885
Simple Ensemble	.5120	.5690	.5100	.1695	.5462	.4923	.5385	.4615
Performance Based Ensemble	.5090	.5693	.5086	.2386	.5692	.4538	.5500	.4846

Note. *Significant at the at the 10% level, **Significant at the 5% level and *** Significant at the 1%* level.

Observing model performance across the different economic cycles offers interesting insights. The periods are based on the European Business Clock (Eurostat, 2025) for the test set. The expansion periods span from 01/01/2019 to 31/03/2019 and 01/04/2022 to 30/09/2022, the slowdown periods are from 01/04/2019 to 30/09/2019 and 01/10/2022 to 31/03/2025, lastly the recession period is from 01/10/2019 until the 31/03/2020.

First it can be seen that models do not perform consistently through all the cycles. During the expansion phase none of the base learners has an accuracy significant at the 10% level, except for Random Forest (RF) for which it is its best performing period. Linear models also score poorly on F1 Score, suggesting that it makes either a lot of false positive predictions or it misses a lot of the true positive cases. Nonetheless, all base models still achieve high accuracies for top/bottom predictions supporting their relevance for portfolio rankings. The two Ensembles perform relatively better with both accuracies significant at the 10% level and emphasising the advantages of combining the predictions of diverse base learners.

During economic slowdown this is the period in which all models, except RF, attain the highest significance at a 1% level. RF still does well on F1 Score and AUC. Overall accuracy for top/bottom observation remains very good for all models.

Lastly through recessions the models do not perform very well, where none are significant apart from the linear models. Those attain their highest accuracies but are only significant at the 5% level suggesting that there is a high standard error. RF performs extremely bad on almost all metrics. The Simple Ensemble display better results than the Performance Based Ensemble likely driven by its higher exposure to linear models.

In conclusion, the results are in line with the observations made in previous studies that machine learning models perform better during times of market stress (Avramov et al., 2023). However, it is interesting to note that the increased performance was mostly seen during slowdowns compared to recession phases. Part of the explanation lies in the way those cycles were defined in previous studies where it would be a binary value assigned based on economic indicators or determined based on the median of a volatility index. Consequently, the added granularity in this study highlights that those better performances happen during economic slowdown, but not during recessions. Regarding the models it is important to note that they seem to all be performing relatively well during different cycles supporting the added value of combining them. Finally, no substantial difference can be observed between the two Ensembles apart during recessions as the Simple Ensemble displayed a higher accuracy but still insignificant. Consequently, the rest of the analysis will focus on the Simple Ensemble as it achieves slightly better results and was found to be the best performing in the literature (Krauss et al., 2017; Wolff & Echterling, 2024).

2. Portfolio Performance

Table 7: *Portfolio Performance Metrics for Benchmarks, Naïve and Unconstrained Mean Variance Portfolios*

Panel A: Benchmarks	Ann Returns (%)	Total Returns (%)	Vol (%)	SR	ASR	Turn (%)	BTC vs EW_ben (%)	MDD (%)	HHI
EW_ben	10.63	88.06	20.55	.47	.37	nan	nan	-35.64	nan
SX5E_ben	10.02	81.61	20.12	.45	.37	nan	nan	-33.65	nan
Panel B: Naïve Portfolios	**Ann Returns (%)**	**Total Returns (%)**	**Vol (%)**	**SR**	**ASR**	**Turn (%)**	**BTC vs EW_ben (%)**	**MDD (%)**	**HHI**
EW_ALL	11.73	100.04	21.20	.51	.38	18.46	5.95	-36.06	0.03
EW_TOP5	14.45	132.52	23.76	.58	.33	43.91	8.70	-34.32	0.20
EW_TOP10	12.11	104.27	22.03	.52	.38	37.17	3.97	-35.28	0.10
VW_ALL	10.82	90.02	20.90	.48	.37	21.07	0.88	-35.78	0.05
VW_TOP5	14.32	130.85	23.69	.58	.38	50.11	7.36	-34.39	0.27
VW_TOP10	13.55	121.25	21.86	.58	.42	43.97	6.63	-34.69	0.14
Panel C: Unconstrained Mean Variance Portfolios	**Ann Returns (%)**	**Total Returns (%)**	**Vol (%)**	**SR**	**ASR**	**Turn (%)**	**BTC vs EW_ben (%)**	**MDD (%)**	**HHI**
MV_ALL_L	4.96	35.21	29.71	.23	.21	63.85	-8.88	-46.89	1.00
MV_ALL_LS	-0.07	-0.43	18.88	-.04	-.04	59.16	-18.09	-44.96	0.50
MV_TOP5_L	4.95	35.11	29.71	.23	.21	63.89	-8.90	-46.93	1.00
MV_TOP5_LS	-0.10	-0.64	18.88	-.04	-.05	59.14	-18.16	-44.88	0.50
MV_TOP10_L	4.96	35.24	29.71	.23	.21	63.85	-8.88	-46.87	1.00
MV_TOP10_LS	-0.05	-0.33	18.88	-.04	-.04	59.14	-18.07	-44.86	0.50

Note. “Ann Returns” is Annualised Returns, “Vol” is Volatility, “SR” is Sharpe Ratio, “ASR” is Adjusted Sharpe Ratio, “Turn” is Turnover, “BTC vs EW_ben” is the Break-even Transaction Cost against the Equal Weighted Benchmark, “MDD” is Maximum Drawdown and “HHI” is Herfindahl-Hirschman Concentration Index.

Next the results obtained from the formed portfolios are examined. Naïve portfolios, weighted based on either market cap (value) or equally, outperform the benchmarks on returns, Sharpe Ratio (SR) and Adjusted Sharpe Ratio (ASR) to the exception of the equal weighted invested in the top 5 (EW_TOP5). Given it had the highest SR the fact that its ASR is the lowest suggests strong exposure to tail risks. This outperformance also did not come at the cost of a substantially higher risk, as volatility is only slightly up and ranges for maximum drawdown are similar. Overall strategies invested in the top 5 or 10 performed better, however, this came at the expense of volatility (risk). Which is line with the usual expected trade-off, leading the different naïve portfolios to be comparable in terms of risk adjusted metrics. However, their higher turnover suggests that some of the profits will be eroded after accounting for transaction fees, but their “cost margin” as expressed by “BTC”, is still relatively high confirming that it could materialise into a profitable trading strategy. In conclusion, the fact that those portfolios purely assembled based on the signals captured by the machine learning models beat all the benchmarks suggests that the signals exploited generate meaningful economic value.

On the contrary the portfolios constructed based on Mean Variance (MV) without constraints achieve extremely poor results. They consistently substantially underperform benchmarks and naïve portfolios on all measures, generating low returns with increased risk exposure at the cost of a higher turnover, maximum drawdowns at least 10% more severe and more concentrated portfolios as indicated by the Herfindahl-Hirschman Concentration Index (HHI). Additionally, it is apparent that exclusively Long Only portfolios were able to generate positive returns, whereas Long Short portfolios are not even able to provide positive returns over a period which has been characterised by upward market trends.

Table 8: *Portfolio Performance Metrics for Benchmarks and Constrained Mean Variance Portfolios*

Panel A: Benchmarks	Ann Returns (%)	Total Returns (%)	Vol (%)	SR	ASR	Turn (%)	BTC vs EW_ben (%)	MDD (%)	HHI
EW_ben	10.63	88.06	20.55	.47	.37	nan	nan	-35.64	nan
SX5E_ben	10.02	81.61	20.12	.45	.37	nan	nan	-33.65	nan
Panel B: Mean Variance Portfolios with 30% Constraint on Weights	**Ann Returns (%)**	**Total Returns (%)**	**Vol (%)**	**SR**	**ASR**	**Turn (%)**	**BTC vs EW_ben (%)**	**MDD (%)**	**HHI**
MV_ALL_L_C_30	13.62	121.54	24.24	.55	.30	48.61	6.14	-34.38	.28
MV_ALL_LS_C_30	1.97	12.90	14.47	.03	.03	53.42	-16.22	-35.63	.26
MV_TOP5_L_C_30	13.77	123.43	24.26	.55	.30	48.66	6.45	-34.13	.28
MV_TOP5_LS_C_30	1.98	13.00	14.48	.03	.03	53.42	-16.19	-35.61	.26
MV_TOP10_L_C_30	13.62	121.52	24.23	.55	.30	48.60	6.14	-34.39	.28
MV_TOP10_LS_C_30	1.96	12.89	14.48	.03	.03	53.43	-16.22	-35.65	.26
Panel C: Mean Variance Portfolios with 50% Constraint on Weights	**Ann Returns (%)**	**Total Returns (%)**	**Vol (%)**	**SR**	**ASR**	**Turn (%)**	**BTC vs EW_ben (%)**	**MDD (%)**	**HHI**
MV_ALL_L_C_50	13.36	118.41	26.46	.52	.30	56.12	4.86	-36.74	.50
MV_ALL_LS_C_50	-0.08	-0.48	18.88	-.04	-.04	59.15	-18.11	-44.96	.50
MV_TOP5_L_C_50	13.30	117.72	26.46	.52	.30	56.14	4.75	-36.88	.50
MV_TOP5_LS_C_50	-0.15	-0.90	18.88	-.05	-.05	59.15	-18.22	-44.99	.50
MV_TOP10_L_C_50	13.33	118.14	26.46	.52	.30	56.11	4.81	-36.77	.50
MV_TOP10_LS_C_50	-0.08	-0.52	18.88	-.04	-.04	59.14	-18.12	-44.97	.50
Panel D: Mean Variance Portfolios with 70% Constraint on Weights	**Ann Returns (%)**	**Total Returns (%)**	**Vol (%)**	**SR**	**ASR**	**Turn (%)**	**BTC vs EW_ben (%)**	**MDD (%)**	**HHI**
MV_ALL_L_C_70	9.75	78.57	27.14	.39	.30	59.42	-1.48	-41.96	.58
MV_ALL_LS_C_70	-0.10	-0.59	18.88	-.04	-.04	59.16	-18.14	-44.98	.50

MV_TOP5_L_C_70	9.77	78.72	27.14	.39	.30	59.45	-1.46	-41.97	.58
MV_TOP5_LS_C_70	-0.10	-0.60	18.88	-.04	-.04	59.14	-18.14	-44.92	.50
MV_TOP10_L_C_70	9.77	78.70	27.14	.39	.30	59.41	-1.46	-41.94	.58
MV_TOP10_LS_C_70	-0.07	-0.45	18.89	-.04	-.04	59.17	-18.10	-44.90	.50

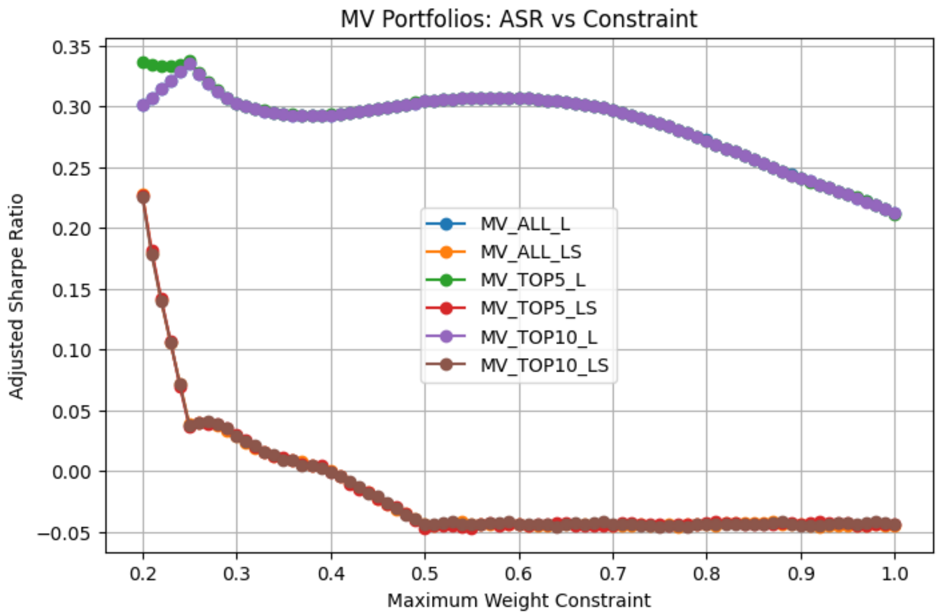
Note. "Ann Returns" is Annualised Returns, "Vol" is Volatility, "SR" is Sharpe Ratio, "ASR" is Adjusted Sharpe Ratio, "Turn" is Turnover, "BTC vs EW_ben" is the Break-even Transaction Cost against the Equal Weighted Benchmark, "MDD" is Maximum Drawdown and "HHI" is Herfindahl-Hirschman Concentration Index.

From the table above a clear pattern emerges. As hypothesised adding tougher constraints to mean variance (MV) portfolios leads to better results. The MV portfolios constructed with a 30% limit on the weights outperformed those formed with a 50% or 70% constraint. This is consistent across all metrics, as it achieves higher returns and risk adjusted returns, and even positive returns for Long Short strategies. In parallel it manages to be exposed to less risk with a lower volatility, maximum drawdown, turnover and logically less concentration. Furthermore the 30% and 50% constrained portfolios perform better than the benchmarks in terms of returns and risk adjusted returns. This supports the hypothesis made that constraints on the portfolio weights can lead to better results as it limits the risk associated with noisy or wrong estimates.

Moreover, the results confirm that Long Only strategies substantially outperform Long Short ones. This is in contradiction with what was observed in previous studies (Avramov et al., 2023; Breitung, 2023). Two hypotheses can be made related to that. First, given that the sample of companies used for this study are the historical members of the Stoxx 50, it can be argued that those have been some of the most successful companies in Eurozone over the last few decades. Therefore, by allowing the algorithm to take short positions, there is more chance of a wrong classification rather than a potential drop in price that could be exploited. The second explanation could be that similarly to imposing constraints on weights, Long Only is also a type of constraint, which in fact was the one the most put forward by Jagannathan & Ma (2003). Therefore, the outperformance of Long Only portfolios might be due to them being extra constrained. However, the first hypothesis seems more plausible as even with a constraint of 30% the performances of Long Short Portfolios are still substantially lower. Nonetheless, Long

Short strategies attain extremely low volatilities and lower concentration. However, their usually negative returns still make them unattractive even for risk averse investors.

Figure 6: *MV Portfolios: ASR vs Constraint*



The plot above compares the Adjusted Sharpe Ratio (ASR) obtained for different levels of weight constraints. As stated above it can be clearly seen that Long Only portfolios outperform Long Short ones and always attain values above 0. It is also interesting to notice that the portfolios based on different considered assets within the two portfolio strategies perform extremely similarly, making them indistinguishable on the graph. This suggests that the universe of asset considered for the portfolio formation does not impact the risk adjusted metrics and suggests a constant relationship between risk and returns. Finally, it can be observed that the best values for ASR are obtained at the lower end, once again confirming the benefits of imposing constraints on the weights. For all the portfolios their top 10% ASRs happened in the range 20-28%, noting that for MV_TOP5 a constraint of 20% is equivalent to the equal weighted portfolio.

Table 9: *Portfolio Performance Metrics for Selected Portfolios during Expansion Cycle*

Panel A: Benchmarks	**AVG Returns (%)**	**Vol (%)**	**SR**	**ASR**	**Turn (%)**	**BTC vs EW_ben (%)**	**MDD (%)**	**HHI**
EW_ben	.30	21.86	.70	.67	nan	nan	-22.40	nan
SX5E_ben	.26	20.89	.62	.60	nan	nan	-24.07	nan
Panel B: Naïve Portfolios	**AVG Returns (%)**	**Vol (%)**	**SR**	**ASR**	**Turn (%)**	**BTC vs EW_ben (%)**	**MDD (%)**	**HHI**
EW_ALL	.28	22.95	.62	.59	20.40	-.10	-24.35	.03
EW_TOP5	.17	25.00	.34	.29	48.12	-.28	-29.09	.20
EW_TOP10	.17	22.93	.38	.36	37.43	-.34	-25.94	.10
VW_ALL	.25	22.20	.58	.55	24.62	-.20	-24.76	.05
VW_TOP5	.20	25.47	.40	.36	55.18	-.18	-24.84	.28
VW_TOP10	.25	22.69	.55	.51	44.46	-.12	-23.17	.16
Panel C: Mean Variance Portfolios with 30% Weight Constraint	**AVG Returns (%)**	**Vol (%)**	**SR**	**ASR**	**Turn (%)**	**BTC vs EW_ben (%)**	**MDD (%)**	**HHI**
MV_ALL_L_C_30	.08	25.75	.14	.13	54.14	-.41	-32.24	.28
MV_TOP5_L_C_30	.08	25.80	.15	.14	54.33	-.41	-32.24	.28
MV_TOP10_L_C_30	.08	25.74	.14	.13	54.13	-.42	-32.24	.28

Note. “AVG Returns” is Weekly Average Returns, “Vol” is Volatility, “SR” is Sharpe Ratio, “ASR” is Adjusted Sharpe Ratio, “Turn” is Turnover, “BTC vs EW_ben” is the Break-even Transaction Cost against the Equal Weighted Benchmark, “MDD” is Maximum Drawdown and “HHI” is Herfindahl-Hirschman Concentration Index.

Table 10: *Portfolio Performance Metrics for Selected Portfolios during Slowdown Cycle*

Panel A: Benchmarks	AVG Returns (%)	Vol (%)	SR	ASR	Turn (%)	BTC vs EW_ben (%)	MDD (%)	HHI
EW_ben	.37	13.58	1.08	1.02	nan	nan	-8.37	nan
SX5E_ben	.37	14.56	1.02	0.99	nan	nan	-10.12	nan
Panel B: Naïve Portfolios	**AVG Returns (%)**	**Vol (%)**	**SR**	**ASR**	**Turn (%)**	**BTC vs EW_ben (%)**	**MDD (%)**	**HHI**
EW_ALL	.43	13.78	1.30	1.18	17.26	.38	-8.75	.03
EW_TOP5	.62	17.15	1.61	1.23	40.39	.62	-10.85	.20
EW_TOP10	.54	15.60	1.49	1.22	36.90	.45	-10.34	.10
VW_ALL	.43	14.47	1.23	1.15	18.72	.33	-10.63	.05
VW_TOP5	.62	16.91	1.63	1.34	46.03	.54	-11.14	.26
VW_TOP10	.51	16.07	1.37	1.26	43.65	.33	-10.32	.14
Panel C: Mean Variance Portfolios with 30% Weight Constraint	**Ann Returns (%)**	**Vol (%)**	**SR**	**ASR**	**Turn (%)**	**BTC vs EW_ben (%)**	**MDD (%)**	**HHI**
MV_ALL_L_C_30	.66	17.73	1.67	1.47	44.75	.64	-10.06	.28
MV_TOP5_L_C_30	.66	17.73	1.68	1.47	44.66	.65	-10.01	.28
MV_TOP10_L_C_30	.66	17.72	1.67	1.47	44.74	.64	-10.00	.28

Note. "AVG Returns" is Weekly Average Returns, "Vol" is Volatility, "SR" is Sharpe Ratio, "ASR" is Adjusted Sharpe Ratio, "Turn" is Turnover, "BTC vs EW_ben" is the Break-even Transaction Cost against the Equal Weighted Benchmark, "MDD" is Maximum Drawdown and "HHI" is Herfindahl-Hirschman Concentration Index.

Table 11: *Portfolio Performance Metrics for Selected Portfolios during Recession Cycle*

Panel A: Benchmarks	AVG Returns (%)	Vol (%)	SR	ASR	Turn (%)	BTC vs EW_ben (%)	MDD (%)	HHI
EW_ben	-.93	39.04	-1.31	-1.34	nan	nan	-35.64	nan
SX5E_ben	-.86	37.05	-1.28	-1.36	nan	nan	-33.65	nan
Panel B: Naïve Portfolios	**AVG Returns (%)**	**Vol (%)**	**SR**	**ASR**	**Turn (%)**	**BTC vs EW_ben (%)**	**MDD (%)**	**HHI**
EW_ALL	-.93	39.46	-1.29	-1.33	15.87	.00	-36.06	.03
EW_TOP5	-.66	42.92	-0.86	-0.94	42.31	.63	-34.32	.20
EW_TOP10	-.81	41.18	-1.09	-1.17	37.69	.32	-35.28	.10
VW_ALL	-.97	38.23	-1.38	-1.42	16.97	-.21	-35.78	.05
VW_TOP5	-.90	41.08	-1.21	-1.23	46.08	.06	-34.39	.24
VW_TOP10	-.79	39.77	-1.09	-1.22	43.51	.33	-34.69	.13
Panel C: Mean Variance Portfolios with 30% Weight Constraint	**AVG Returns (%)**	**Vol (%)**	**SR**	**ASR**	**Turn (%)**	**BTC vs EW_ben (%)**	**MDD (%)**	**HHI**
MV_ALL_L_C_30	-.54	42.36	-0.72	-0.80	42.91	.92	-32.88	.28
MV_TOP5_L_C_30	-.54	42.35	-0.72	-0.80	42.97	.91	-32.88	.28
MV_TOP10_L_C_30	-.54	42.35	-0.72	-0.80	42.96	.92	-32.88	.28

Note. "AVG Returns" is Weekly Average Returns, "Vol" is Volatility, "SR" is Sharpe Ratio, "ASR" is Adjusted Sharpe Ratio, "Turn" is Turnover, "BTC vs EW_ben" is the Break-even Transaction Cost against the Equal Weighted Benchmark, "MDD" is Maximum Drawdown and "HHI" is Herfindahl-Hirschman Concentration Index.

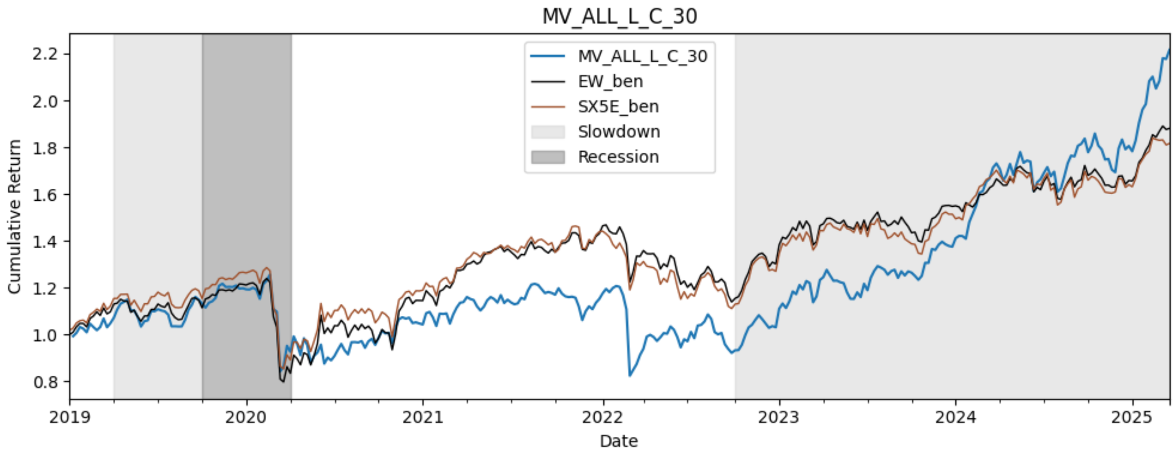
The tables above provide interesting insights into the performance of the portfolios across the different economic cycles. What can be first noticed is that the performance nor the top performing strategy is constant across cycles. During the expansion periods the equal weighted benchmark (EW_ben) outperforms all the other strategies across all metrics. Once again proving to be a reference point hard to beat. Additionally, given its statical weights it also has extremely low trading fees associated. On the contrary 30% constrained Mean Variance portfolios perform the worse across all metrics and despite the benefits of limiting the weights.

During slowdown periods all models including benchmarks substantially perform the best. For all strategies the Sharpe Ratio and Adjusted Sharpe Ratio are above 1. Interestingly, for all portfolios volatility and maximum drawdown are substantially lower compared to other periods. Most importantly, MV portfolios outperform all the others, yielding results as much as 50% higher than the benchmarks. This is line with the observation made earlier that machine learning models were the most accurate during that cycle.

Lastly, during the recession all the portfolios perform poorly with highly negative returns. MV still performs the best with almost twice as good results in terms of returns and risk adjusted metrics compared to the benchmarks. This echoes the observation made by Avramov et al. (2023) that ML strategies prove to be resilient during periods of market stress and are characterised by lower maximum drawdowns which makes them a good hedge. Furthermore, it is important to point that this is also the period in which they have the lowest turnover, achieving a turnover lower than certain naïve portfolios. This suggests that in times of recession the model holds more to certain positions. Nonetheless, the turnover remains high.

In conclusion, all the portfolios perform the best during slowdown cycles. Additionally, the equal weighted benchmark performs the best during expansion periods and 30% constrained MV portfolios performs best during slowdowns or recession, in line with previous literature. Suggesting investing in EW_ben during expansion phases and in MV_C_30 during economic slowdowns or recession.

Figure 7: *Cumulative Return Plot for MV ALL Long Only Constrained with 30%*



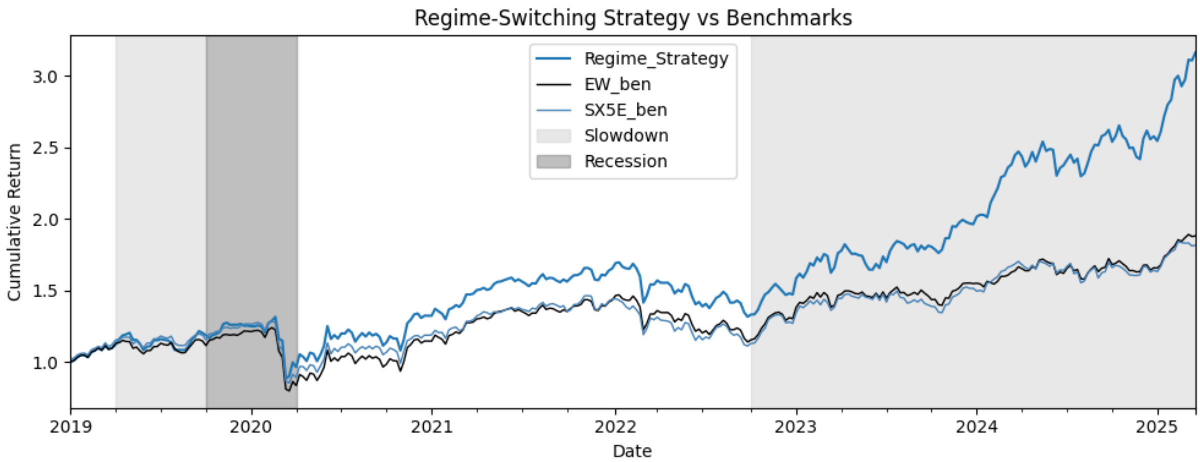
Here is a visual representation of the performance of the MV_ALL_L_C_30. As mentioned previously it can be seen that it performs the best during periods of slowdown particularly the second one from the 01/10/2022 to the 31/03/2025. Moreover, during the recession phase it drops slightly less than the benchmarks and during the expansion period its spread with the benchmarks grows consistently and it is more volatile.

Table 12: *Portfolio Performance Metrics for Regime Switching Strategy vs Others*

Portfolios	Ann Returns (%)	Total Returns (%)	Vol (%)	SR	ASR	MDD (%)
EW_ben	10.63	88.06	20.55	0.47	0.37	-35.64
SX5E_ben	10.02	81.61	20.12	0.45	0.37	-33.65
VW_TOP10	13.55	121.25	21.86	0.58	0.42	-34.69
MV_ALL_L_C_30	13.62	121.54	24.24	0.55	0.30	-34.38
Regime Switching Strategy	20.22	216.14	22.45	0.82	0.41	-32.88

The table above presents the performance metrics for the benchmarks and portfolios found to be performing well. Those are then compared to the Regime Switching Strategy, which consists of being invested in the equal weighted benchmark during periods of expansion and invested in the MV_ALL_L_C_30 during phases of slowdown or recession. It can be observed that this strategy yields the highest returns, Sharpe Ratio and lowest maximum drawdown. However, its Adjusted Sharpe Ratio drops significantly suggesting that those higher returns are a compensation for a stronger exposition to tail risks.

Figure 8: *Cumulative Return Plot for Regime Switching Strategy*



As explained previously such as a strategy would capitalise on a lower decline during the recession, bounce back from a higher point during the expansion period following the returns of the equal weighted benchmark and then lastly further growing its spread during the slowdown cycle through its investment in the 30% constrained MV portfolio.

3. Risk Exposure

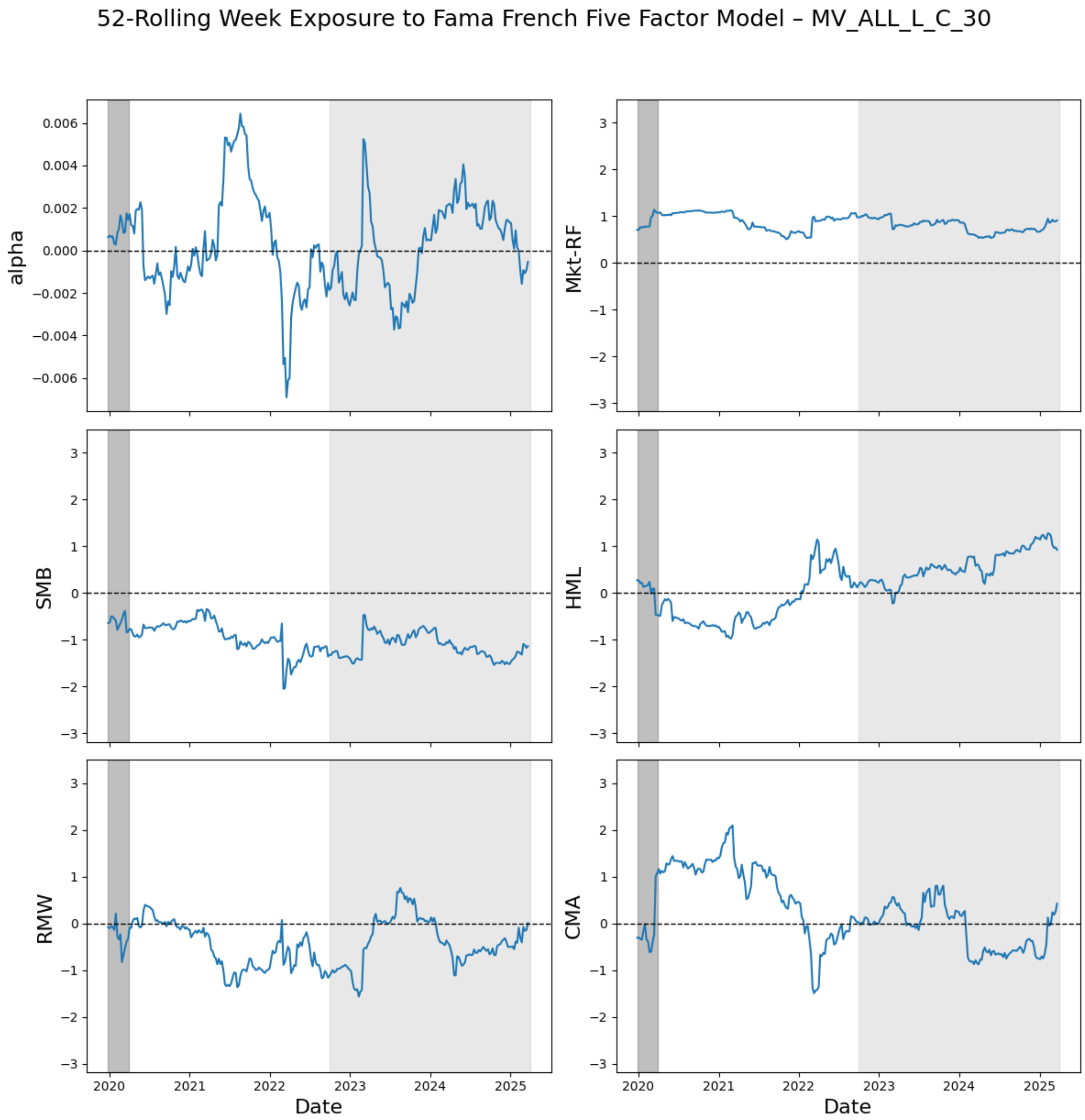
Table 13: *Fama French Five Factor Exposure*

Panel A: Benchmarks	**Alpha**	**Mkt-RF**	**SMB**	**HML**	**RMW**	**CMA**	**R2**
EW_ben	-.0002	.87***	-0.64***	.42***	-.02	.12	.90
SX5E_ben	-.0003	.87***	-0.74***	.20***	.11	-.06	.90
Panel B: Naïve Portfolios	**Alpha**	**Mkt-RF**	**SMB**	**HML**	**RMW**	**CMA**	**R2**
EW_ALL	-.0001	.89***	-0.70***	.41***	-.11	.08	.88
EW_TOP5	.0004	.87***	-0.98***	.35***	-.42**	-.01	.73
EW_TOP10	.0000	.86***	-0.85***	.39***	-.28*	.06	.81
VW_ALL	-.0002	.89***	-0.77***	.21***	.01	-.01	.87
VW_TOP5	.0004	.88***	-1.00***	.33***	-.26	-.06	.72
VW_TOP10	.0003	.86***	-0.93***	.22**	-.26*	.07	.77
Panel C: Mean Variance Portfolios with 30% Weight Constraint	**Alpha**	**Mkt-RF**	**SMB**	**HML**	**RMW**	**CMA**	**R2**
MV_ALL_L_C_30	.0005	.85***	-1.02***	.15	-.62***	.12	.65
MV_TOP5_L_C_30	.0005	.85***	-1.03***	.16	-.61***	.10	.65
MV_TOP10_L_C_30	.0005	.85***	-1.02***	.15	-.62***	.12	.65
Panel D: Custom Strategy	**Alpha**	**Mkt-RF**	**SMB**	**HML**	**RMW**	**CMA**	**R2**
Regime Switching Strategy	.0014*	.88***	-0.86***	.40***	-.13	.11	.81

Note. *Significant at the at the 10% level, **Significant at the 5% level and *** Significant at the 1%* level.

An essential question is whether the performance of the ML strategies is attributable to common risk factors, measured by the Fama French Five Factor model, or does the portfolio generate abnormal returns not explained by exposure to risk captured in the term alpha. Hence an OLS factor regression is performed. First it can be mentioned that no model displays significant alpha, suggesting that the returns of the portfolios are in line with the expectations of asset pricing theory, which argues that returns should be explained by risk factors. This is further supported with relatively high values for R2. Nonetheless, the only exception is the Regime Switching Strategy introduced based on the results presented formerly. Meaning that its higher returns are not just attributable to higher risk exposure. Regarding the risk factors all the model load positively to the Market and negatively to the Small Minus Big factors. This is logical as those are all Long Only strategies. Additionally, it is expected that the exposure to SMB will be negative as the stocks considered are the historical members of the Stoxx 50 and therefore, are large capitalisations. Furthermore, all portfolios except the MVs, but still including the Regime Switching Strategy have a strong significant exposure to the High Minus Low factor, suggesting that those strategies privilege stocks with high market to book ratios. Finally, the MV portfolios and some naïve portfolios have a negative exposure to Robust Minus Weak factor, indicating that these approaches focus more on companies with weak profit margins. This could be explained by exposure to specific industries characterised by lower margins.

Figure 9: *52 Rolling Week Exposure to Fama French Five Factor Model for MV_ALL_L_C_30*



The graphs above offer a visual representation of the rolling exposure to risk factors for the MV_ALL_L_C_30 portfolio. It can be seen that the Market and SMB factors are respectively consistently above and below zero. The alpha makes big swings without any patterns. The same can be said about the other risk factors suggesting that those fluctuations imply that the model does not change risk exposure strategy during the different cycles.

4. Feature Importance

Finally, feature importance is analysed to understand how the machine learning models come to their conclusions. Particular attention is given to macroeconomic factors as this study is the first to include them in this context. Linear models are found to give a lot of importance to the Euribor 3- and 6-months rates as well as the Term Spread. Additionally, Lasso and Elastic Net have almost identical feature importances which is not surprising given how similar their predictions are. These two additionally include monthly changes in M1 and M2 money aggregates, suggesting in general linear models give more importance to indicators linked to monetary policies. Moreover, apart from monetary indicators the models rely a lot on Relative Strength Indices (RSI), price-based technical indicators and on the contrary do not give importance to fundamentals.

Regarding Random Forest it does not bring attention to any macroeconomic indicator, maybe this is the reason why it underperformed the other models. Instead, it puts emphasis on valuations (ex: TRAIL12M_NI_to_MktCap, CFO_to_MktCap) and momentum/growth features. In general, it favours more fundamental and earnings-based signals.

Lastly, Light GBM gives a lot of importance to market state variables, particularly those derived from the Stoxx 50 such as the dividend yield, Price to Earnings and Price to Book ratios. Additionally, it gives the second most importance to the V2X Index, which tracks the volatility of the Stoxx 50. Essentially focusing more on indicators of macro sentiment. Apart from that it ranks high momentum/previous weeks' returns and volatility features.

Overall macroeconomic indicators are found to be important to several models, confirming their relevance. The SHAP values plots are available under the Appendix C.

V. Conclusion

In conclusion, this study provides robust empirical evidence related to using machine learning to predict the probability of outperformance of a security. First, consistent with Wolff & Echterling (2024) it is found that the best performing models are the linear ones, supporting the observation made by Avramov et al. (2023) that the non-linearities are mostly relevant to predict hard to arbitrage stocks. Additionally, models' accuracies peak during slowdown cycles.

The results obtained from the portfolios also shed light on the benefits of imposing weight constraints in the mean variance (MV) optimisation problem. Naïve portfolios (equal or value weighted) performed the best compared to the benchmarks and unconstrained MV. However, after imposing different levels of constraints it was found that the 30% constrained MV portfolios performed the best supporting the hypothesis made about the benefit in terms of risk adjusted returns of imposing limits to the weight assets can take within MV, thus exposing them less to the risk of erroneous estimates for the expected returns or the covariance matrix. Interestingly, it was found that the equal weighted benchmark performs the best during expansion phases, whereas MV 30% constrained portfolios performed the best during recession or slowdown periods. Emphasising again the results obtained in previous studies about their lower maximum drawdowns and utility as a hedge during market crises. All portfolios had their best performing period during the slowdown cycle, previous studies found similar findings but for recessions. This could be explained by the increased granularity in this study, suggesting that actually the better performances are present during slowdown phases. A Regime Switching Strategy that would invest in the equal weighted benchmark during periods of expansions and in MV_ALL_L_C_30 during recessions or slowdowns substantially outperformed other strategies, achieving annualised returns of 20,22% and a Sharpe Ratio of 0,82.

When evaluating the portfolios exposure to risk factors measured through the Fama French Five Factor model, all portfolios were significantly positively exposed to the market factor and negatively to the Small Minus Bis, which is logical given that those are all long only strategies in large European capitalisations. The only strategy with a significant alpha at the 10% level is the Regime Switching Strategy, putting forward again its ability to generate superior results. Apart from that benchmarks, naïve portfolios and the Regime Switching Strategy also had significant positive exposure to the High Minus Low factor and the other MV portfolios had a significant negative exposure to the Robust Minus Weak factor.

Additionally, it was found that long short strategies returned extremely poor results which is inconsistent with previous findings, which found the short leg to be the most profit generating. The most probable reason is due to the sample of companies considered for this work. Those are the historical components of the Stoxx 50. Therefore, those have been some of the most successful companies in the last decades in the eurozone. Consequently, there was not much short signal to exploit for the model. The stocks considered are also likely to be the reason why this study attains lower results in terms of returns and Sharpe Ratio. This is because as they are some of the largest businesses in Europe they are scrutinised by teams of analysts, making them less likely to be mispriced.

Finally, macroeconomic factors were found to play a significant role. Linear models particularly pick-up indicators linked to monetary policies, such as Euribor 3- and 6-months rates, Term Spread or monthly changes in monetary aggregate M1 and M2. On the other side Light GBM gave more attention market state variables linked to the Stoxx 50. Random Forest was the only model which ignored macroeconomic indicators, maybe this is the reason why it underperformed its peers.

This study aims to be the starting point for further research in the field. Recommended directions could be exploring what values for the risk aversion parameter (lambda) yields the best results. Additionally, implementing the Black-Litterman asset allocation model could allow to include views, and a level of confidence associated with the expected returns. Furthermore, additional research could be pursued to identify the optimal constraint range during the different economic cycles. Finally, it could be considered to expand the stock universe and features. A particularly interesting direction would be to include media sentiment data as in the work of Zheng & Lucey (2025).

VI. References

1. Academic

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Avramov, D., Cheng, S., & Metzker, L. (2023). Machine Learning vs. Economic Restrictions: Evidence from Stock Return Predictability. *Management Science*, *69*(5), 2587–2619. https://doi.org/10.1287/mnsc.2022.4449

Bella, A., Ferri, C., Hernández-Orallo, J., & Ramírez-Quintana, M. J. (2013). On the effect of calibration in classifier combination. *Applied Intelligence*, *38*(4), 566–585.

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, *191*, 192–213.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*(2).

Black, F. (1972). Capital market equilibrium with restricted borrowing. *The Journal of Business*, *45*(3), 444–455.

Black, F., & Litterman, R. (1990). Asset allocation: Combining investor views with market equilibrium. *Goldman Sachs Fixed Income Research*, *115*(1), 7–18.

Boons, M. (2016). State variables, macroeconomic activity, and the cross section of individual stocks. *Journal of Financial Economics*, *119*(3), 489–511. https://doi.org/10.1016/j.jfineco.2015.05.010

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/BF00058655

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

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification And Regression Trees* (1st ed.). Routledge. https://doi.org/10.1201/9781315139470

Breitung, C. (2023). Automated stock picking using random forests. *Journal of Empirical Finance*, *72*, 532–556. https://doi.org/10.1016/j.jempfin.2023.05.001

Celebi, K., & Hönig, M. (2019). The impact of macroeconomic factors on the German stock market: Evidence for the crisis, pre-and post-crisis periods. *International Journal of Financial Studies*, *7*(2), 18.

Chen, L., Pelger, M., & Zhu, J. (2023). Deep learning in Asset Pricing. *Management Science*.

Cornell, B. (2021). ESG preferences, risk and return. *European Financial Management*, *27*(1), 12–19.

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, *22*(5), 1915–1953.

Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The 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. (2004). The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives*, *18*(3), 25–46.

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

Ferrer, R., Bolós, V. J., & Benítez, R. (2016). Interest rate changes and stock returns: A European multi-country study with wavelets. *International Review of Economics & Finance*, *44*, 1–12. https://doi.org/10.1016/j.iref.2016.03.001

Flannery, M. J., & Protopapadakis, A. A. (2002). Macroeconomic factors do influence aggregate stock returns. *The Review of Financial Studies*, *15*(3), 751–782.

Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, *111*(1), 1–25. https://doi.org/10.1016/j.jfineco.2013.10.005

Freyberger, J., Höppner, B., Neuhierl, A., & Weber, M. (2022). *Missing data in asset pricing panels* (w30761). National Bureau of Economic Research.

Green, R. C., & Hollifield, B. (1992). When will mean-variance efficient portfolios be well diversified? *The Journal of Finance*, *47*(5), 1785–1809.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Huang, C.-F. (2012). A hybrid stock selection model using genetic algorithms and support vector regression. *Applied Soft Computing*, *12*(2), 807–818. https://doi.org/10.1016/j.asoc.2011.10.009

Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, *58*(4), 1651–1683.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945-1964. *The Journal of Finance*, *23*(2), 389–416.

Jenson, M., Scholes, M., & Black, F. (1972). The Capital Asset Pricing Model: Some Empirical Tests. *Jensen, Michael. Studies in the Theory of Capital Markets*.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *30*.

Keating, C., & Shadwick, W. F. (2002). A universal performance measure. *Journal of Performance Measurement*, *6*(3), 59–84.

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, *259*(2), 689–702. https://doi.org/10.1016/j.ejor.2016.10.031

Kubat, M., Bratko, I., & Michalski, R. (1998). A review of machine learning methods. In *Machine learning and data mining: Methods and applications* (pp. 3–69). Springer.

Li, A. W., & Bastos, G. S. (2020). Stock Market Forecasting Using Deep Learning and Technical Analysis: A Systematic Review. *IEEE Access*, *8*, 185232–185242. https://doi.org/10.1109/ACCESS.2020.3030226

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

López de Prado, M. M. (2018). *Advances in financial machine learning*. Wiley.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and Machine Learning. *Expert Systems with Applications*, *165*.

Maréchal, M. (2024). *The Impact of Political Indicators on Asset Pricing Predictions: An LSTM-based approach for ASEAN Markets* [Thesis]. Tilburg University.

Markowitz, H. M. (1952). Portfolio Selection. *The Journal of Finance*, *7*(1), 77–91.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, 867–887.

Modigliani, F., & Miller, M. H. (1958). The Cost of Capital, Corporation Finance and the Theory of Investment. *The American Economic Review*, *48*(3), 261–297.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica 34 (4)*, 768–783.

Pezier, J. (2004). Risk and risk aversion,[w:] alexander c., shedy e. *The Professional Risk Managers' Handbook. A Comprehensive Guide to Current Theory and Best Practices*, *1*, 19.

Salehpour, A., & Samadzamini, K. (2023). Machine Learning Applications in Algorithmic Trading: A Comprehensive Systematic Review. *International Journal of Education and Management Engineering*, *13*(6), 41–53. https://doi.org/10.5815/ijeme.2023.06.05

Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison of Random Forest and Parametric Imputation Models for Imputing Missing Data Using MICE: A CALIBER Study. *American Journal of Epidemiology*, *179*(6), 764–774. https://doi.org/10.1093/aje/kwt312

Shapley, L. S. (1953). A value for n-person games. In *Contributions to the theory of games* (Vol. 2, pp. 307–317). Princeton University Press. https://doi.org/10.1515/9781400881970-018

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

Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business*, *39*(1), 119–138.

Stam, M. H. (2022). *Asset Pricing and the applications of Machine Learning in missing data treatment* [PhD Thesis]. Tilburg University.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Treynor, J. L. (1961). *Toward a theory of market value of risky assets*.

Tsai, C.-F., & Hsiao, Y.-C. (2010). Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. *Decision Support Systems*, *50*(1), 258–269. https://doi.org/10.1016/j.dss.2010.08.028

Wolff, D., & Echterling, F. (2024). Stock picking with machine learning. *Journal of Forecasting*, *43*(1), 81–102. https://doi.org/10.1002/for.3021

Yu, J.-R., Paul Chiou, W.-J., Lee, W.-Y., & Lin, S.-J. (2020). Portfolio models with return forecasting and transaction costs. *International Review of Economics and Finance*, *66*, 118–130.

Zheng, Y., & Lucey, B. (2025). Supervised learning models, statistical models or hybrid models? A prediction of clean energy stock based on fear and fundamental factors. *Finance Research Open*, *1*(2), 100006. https://doi.org/10.1016/j.finr.2025.100006

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

2. Databases

Bloomberg L.P. (2025) *Fundamental and Technical Data for the Historical Components of the Stoxx 50, as well as Macroeconomic Data for the Eurozone. 1/1/2004 to 31/3/2025*. Retrieved June 3, 2025 from Bloomberg terminal.

European Central Bank. (2025, June 3). *HICP – Overall index, Euro area (changing composition), Monthly* [Data set]. ECB Data Portal. Retrieved June 3, 2025, from https://data.ecb.europa.eu/data/datasets/ICP/ICP.M.U2.N.000000.4.ANR data.ecb.europa.eu

European Central Bank. (2025, May 30). *Monetary aggregate M1 reported by MFIs, central government and post office giro institutions in the euro area (annual growth rate), Euro area (changing composition), Monthly* [Data set]. ECB Data Portal. Retrieved June 3, 2025, from https://data.ecb.europa.eu/data/datasets/BSI/BSI.M.U2.Y.V.M10.X.I.U2.2300.Z01.A data.ecb.europa.eu

European Central Bank. (2025, May 30). *Monetary aggregate M2 reported by MFIs, central government and post office giro institutions in the euro area (annual growth rate), Euro area* *(changing composition), Monthly* [Data set]. ECB Data Portal. Retrieved June 3, 2025, from https://data.ecb.europa.eu/data/datasets/BSI/BSI.M.U2.Y.V.M20.X.I.U2.2300.Z01.A data.ecb.europa.eu

European Central Bank. (2025, May 30). *Monetary aggregate M3 reported by MFIs, central government and post office giro institutions in the euro area (annual growth rate), Euro area (changing composition), Monthly* [Data set]. ECB Data Portal. Retrieved June 3, 2025, from https://data.ecb.europa.eu/data/datasets/BSI/BSI.M.U2.Y.V.M30.X.I.U2.2300.Z01.A

Eurostat. (2025). *Business Cycle Clock* [Dataset]. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Business_Cycle_Clock#Data_sources

French, K. R. (n.d.). *Fama/French European 5 Factors [Daily]* [Data set]. Kenneth R. French Data Library. Retrieved June 3, 2025, from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Ice Data Indices, LLC, ICE BofA Euro High Yield Index Effective Yield [BAMLHE00EHYIEY], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/BAMLHE00EHYIEY , June 3, 2025.

OECD. (2025). *Composite leading Indicators* [Dataset]. https://data-explorer.oecd.org/vis?lc=en&pg=0&snb=1&vw=tb&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_STES%40DF_CLI&df[ag]=OECD.SDD.STES&df[vs]=&pd=2004-01%2C2025-03&dq=EA19.M.CCICP%2BBCICP. . .AA. . .H&ly[rw]=TIME_PERIOD&to[TIME_PERIOD]=false

3. Others

Chen, J. (2025, May 10). *Investment grade credit rating details: What does it mean?* Investopedia. https://www.investopedia.com/terms/i/investmentgrade.asp

European Central Bank. (2014). *Euro area risk-free interest rates: measurement issues, recent developments and RELEVANCE to monetary policy*. https://www.ecb.europa.eu/pub/pdf/other/art1_mb201407_pp63-77en.pdf

What to know before saying Hi to High-Yield Bonds. (2023, January 17). FINRA.org. https://www.finra.org/investors/insights/what-to-know-high-yield-bonds

VII. Appendix

Appendix A: AI Declaration

This report was prepared with the assistance of generative AI tools. Artificial intelligence was used in a limited capacity to support the writing process, particularly for idea structuring, editing, and enhancing clarity of expression. Additionally, AI tools were used during the coding and debugging phases of the project. All analysis, interpretation of results, and final conclusions were made by the author.

Appendix B: Hyperparameter Search

1. Ridge

Table 14: *Hyperparameter Search Space for Ridge*

Hyperparameter	Search Space	Selected Value
Alpha (Shrinkage Parameter)	(1e-4 – 1e2) with Log Sale	0.1771350052000001

Note. 200 trials with best cross validation accuracy of 0.5137348786154593

Figure 10: *Optimisation History Plot Ridge*

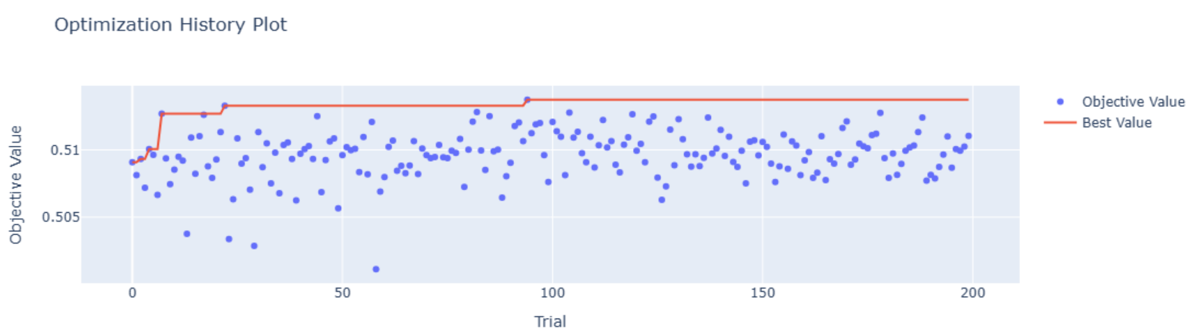


Figure 11: *Parallel Coordinate Plot Ridge*

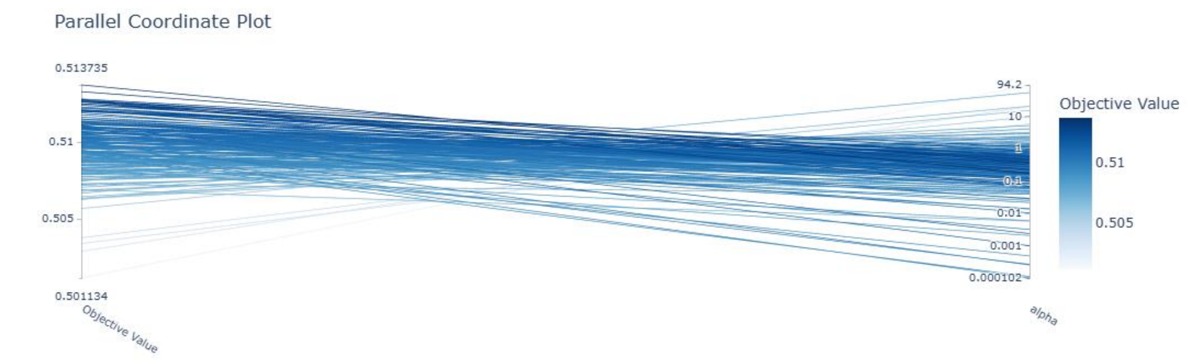
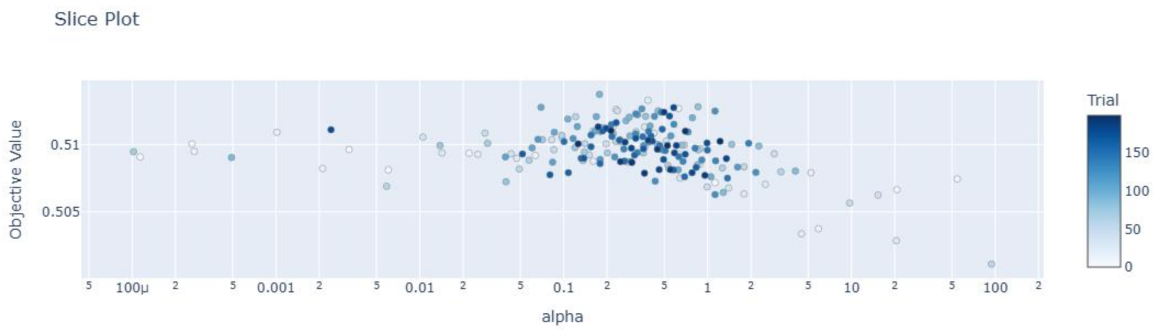


Figure 12: *Slice Plot Ridge*



2. Lasso

Table 15: *Hyperparameter Search Space for Lasso*

Hyperparameter	Search Space	Selected Value
Alpha (Shrinkage Parameter)	(1e-4 – 1e2) with Log Sale	0.00019926243381685035

Note. 200 trials with best cross validation accuracy of 0.5285827724588005

Figure 13: *Optimisation History Plot Lasso*

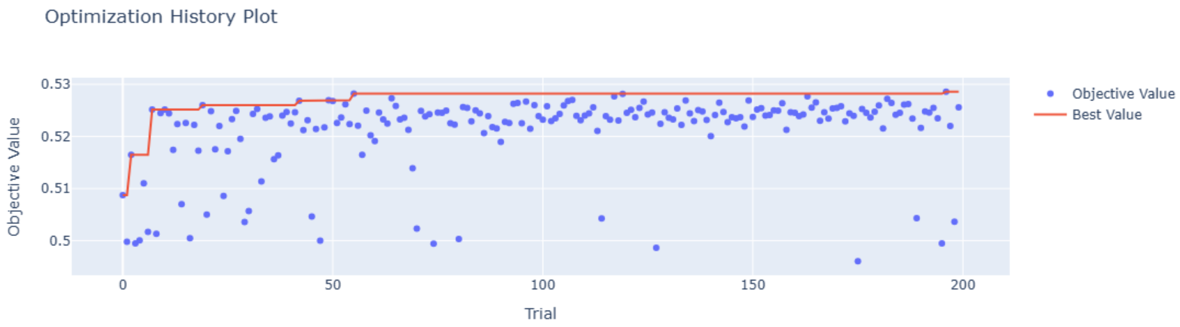


Figure 14: *Parallel Coordinate Plot Lasso*

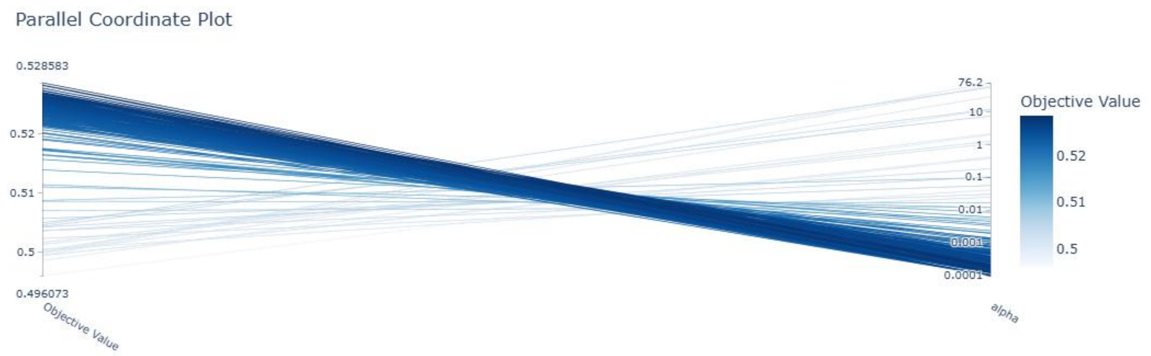
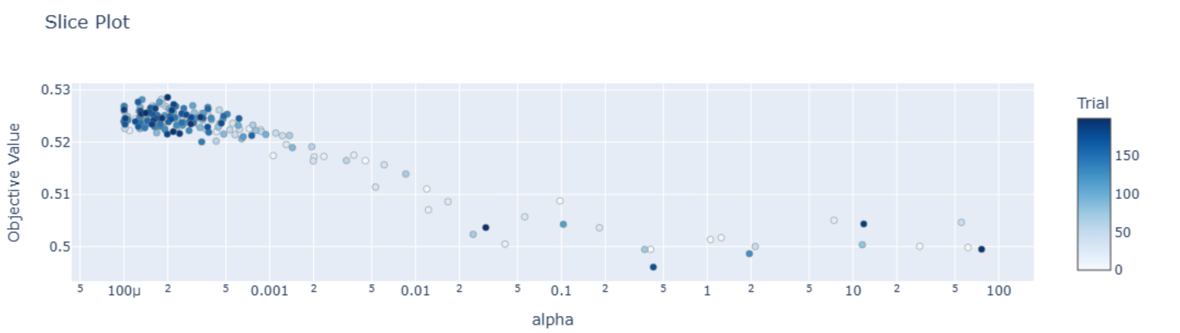


Figure 15: *Slice Plot Lasso*



3. Elastic Net

Table 16: *Hyperparameter Search Space for ENet*

Hyperparameter	Search Space	Selected Value
Alpha (Shrinkage Parameter)	(1e-4 – 1e2) with Log Sale	0.0012846119338193636
L1 Ratio	(0 – 1)	0.24217426451275467
Epochs	100	100

Note. 200 trials with best cross validation accuracy of 0.512495852676667.

Figure 16: *Optimisation History Plot ENet*

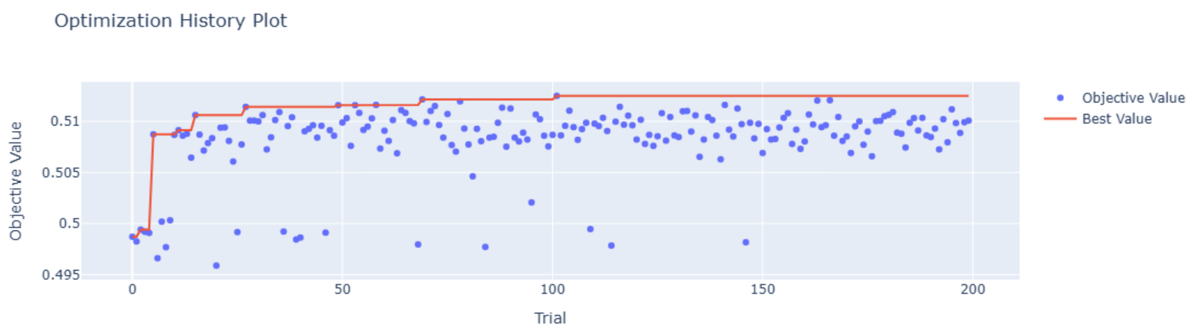


Figure 17: *Parallel Coordinate Plot ENet*

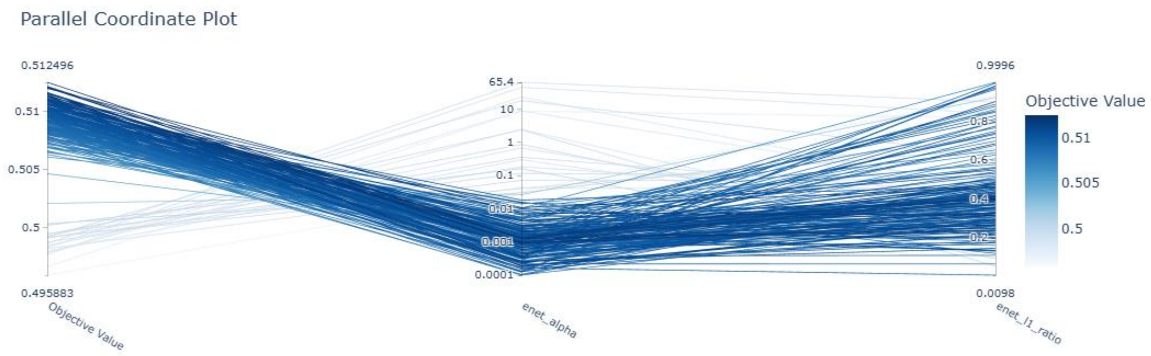


Figure 18: *Slice Plot ENet*

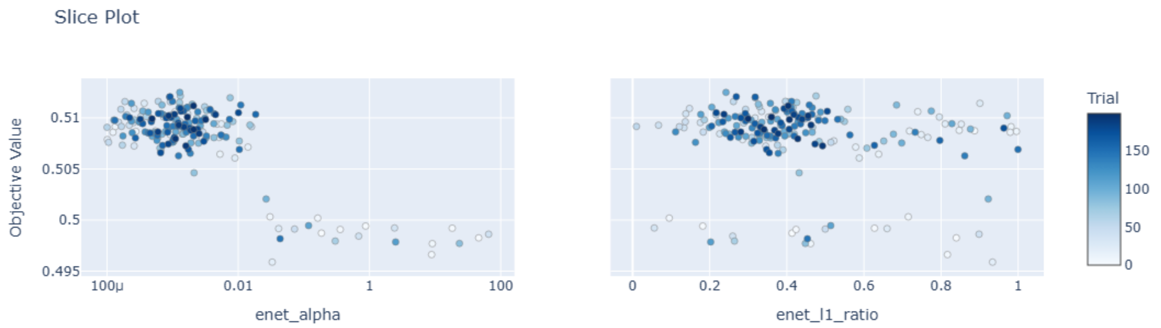
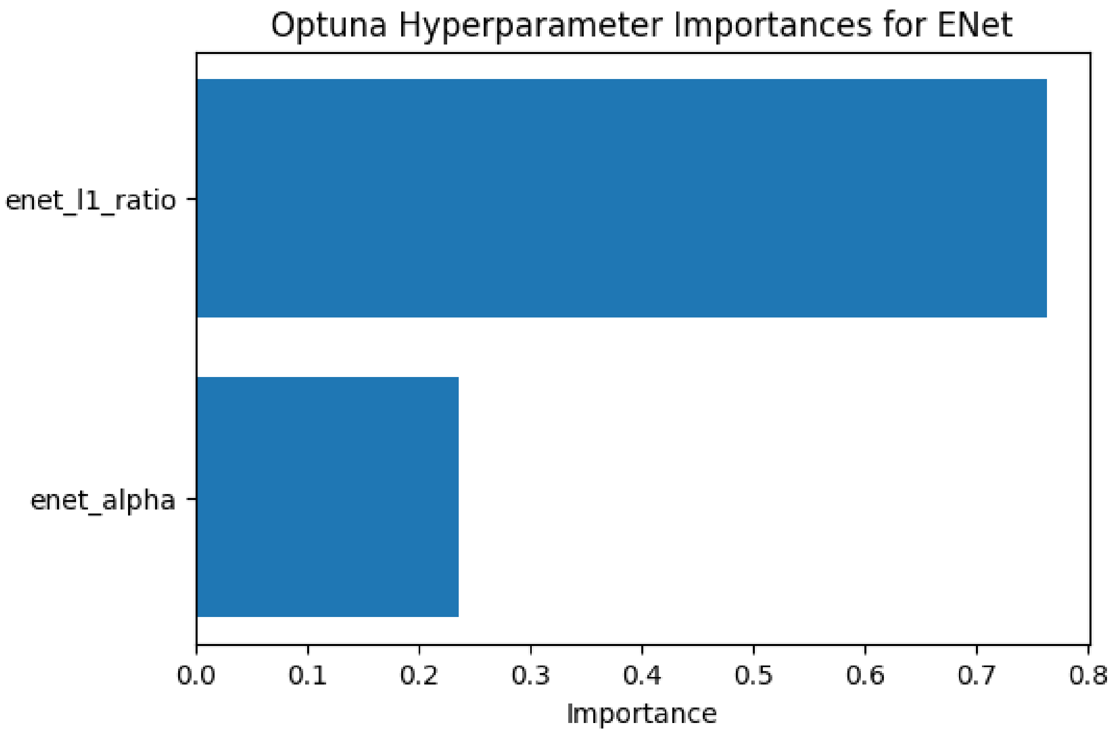


Figure 19: *Optuna Hyperparameter Importances for ENet*



4. Random Forest

Table 17: *Hyperparameter Search Space for Random Forest*

Hyperparameter	Search Space	Selected Value
Number of Estimators (Trees)	(200 - 500) with Step 50	250
Maximum Depth	(3 - 13)	12
Minimum Samples per Leaf	(150 - 1650) with Step 5	180
Minimum Impurity Decrease	(1e-6 - 0.1) with Log Scale	3.0307257109774056e-05
Maximum Features	'sqrt'	'sqrt'

Note. 400 trials with best cross validation accuracy of 0.5178932872040868

Figure 20: *Optimisation History Plot Random Forest*

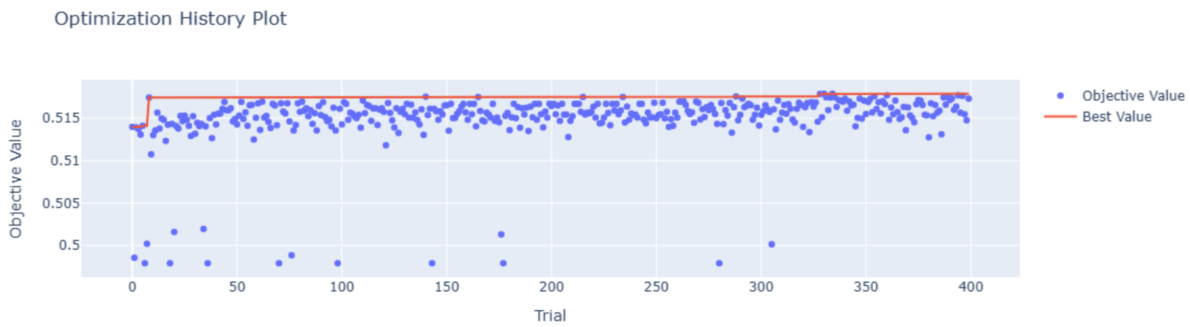


Figure 21: *Parallel Coordinate Plot Random Forest*

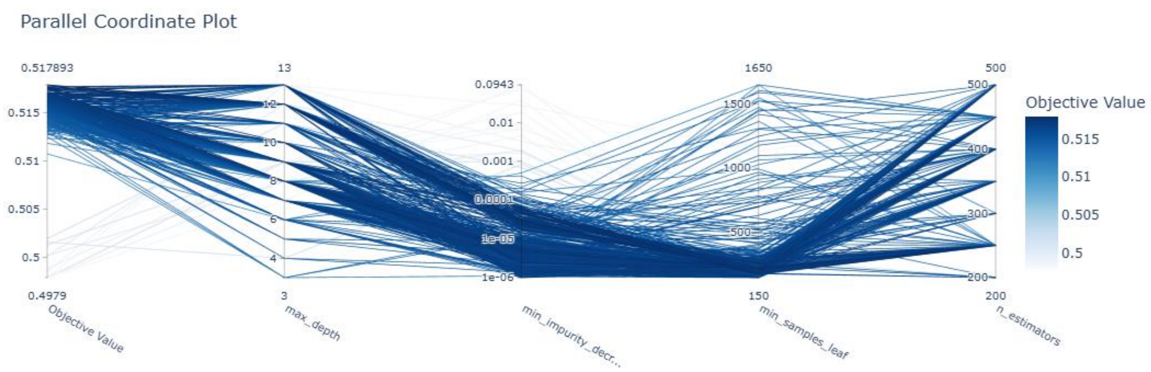
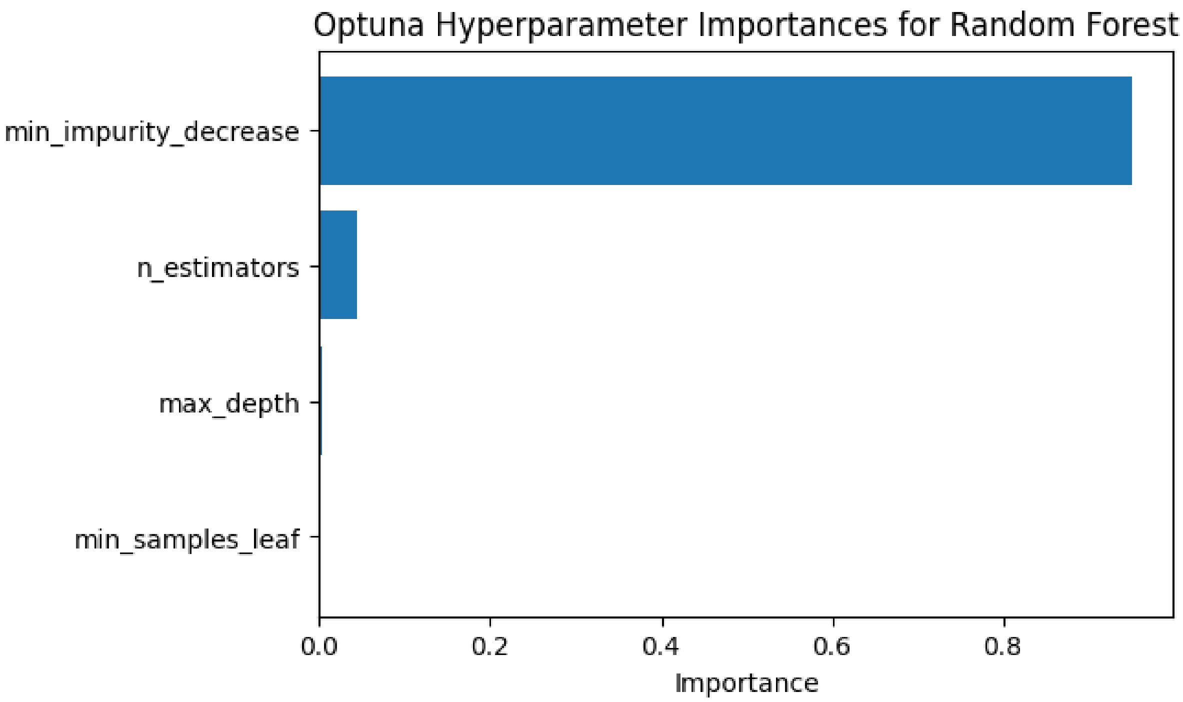


Figure 22: *Split Plot Random Forest*

Figure 23: *Optuna Hyperparameter Importances for Random Forest*



5. Light GBM

Table 18: *Hyperparameter Search Space for Light GBM*

Hyperparameter	**Search Space**	**Selected Value**
Number of Estimators (Trees)	(200 – 1000) with Step 50	950
Maximum Depth	(5 – 15)	15
Number of Leaves	(8 – 256)	230
Minimum Samples per Leaf (min_child_samples)	(150 - 1650) with Step 10	150
Minimum Gain to Split	(1e-6 - 1) with Log Scale	1.3805195901727264e-06
Bagging Frequency	(1 – 5)	3
Bagging Fraction	(0.5 – 0.9)	0.89994945639258
L1 Penalty	(1e-4 – 1e2) with Log Scale	2.4448095019841722
Feature Fraction	(0.4 – 0.9)	0.8941287837897733
Learning Rate	(0.01 – 0.3)	0.11043692417684253

Note. 800 trials with best cross validation accuracy of 0.5361856259672034.

Figure 24: *Optimisation History Plot Light GBM*

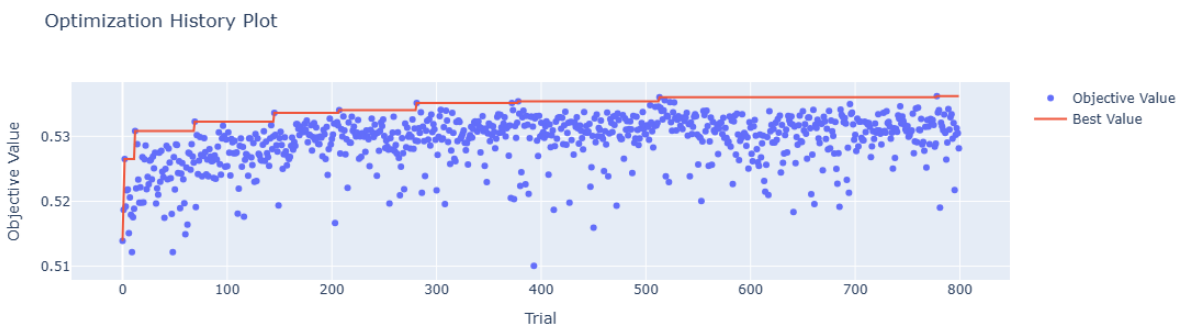


Figure 25: *Parallel Coordinate Plot Light GBM*

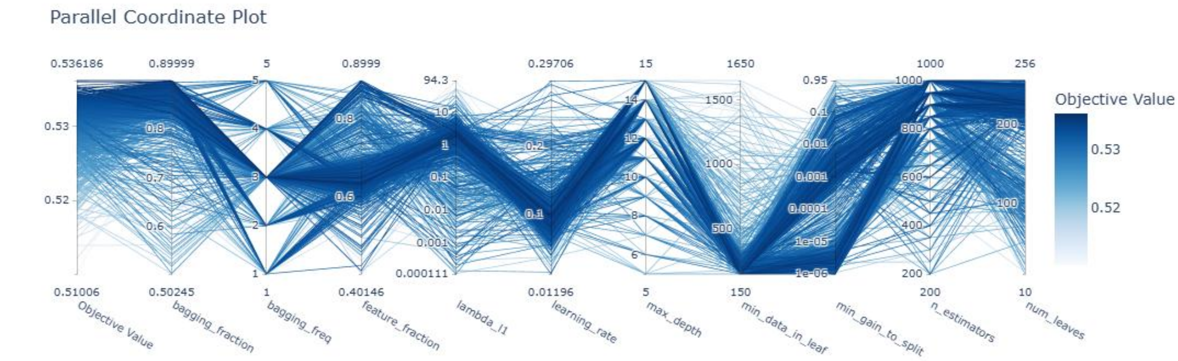


Figure 26: *Slice Plot 1 Light GBM*

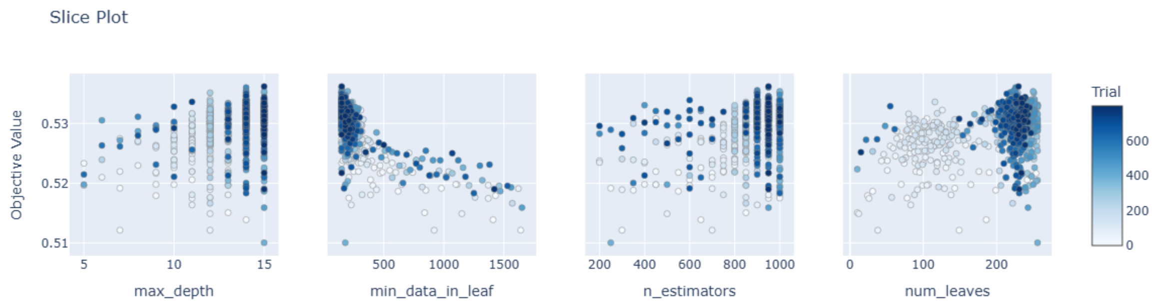


Figure 27: *Slice Plot 2 Light GBM*

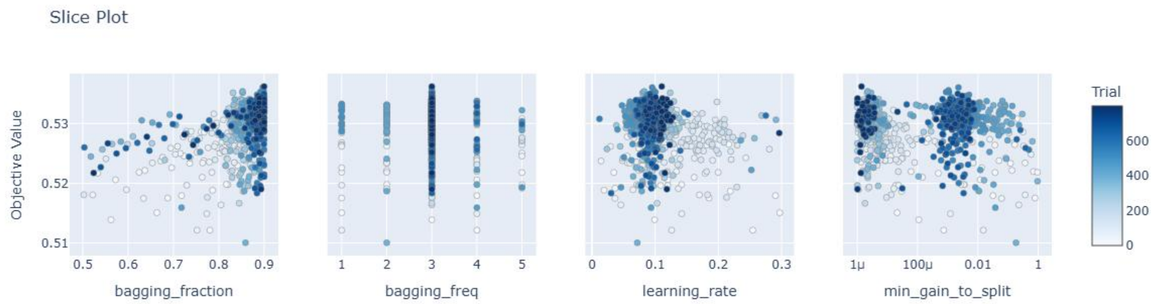


Figure 28: *Slice Plot 3 Light GBM*

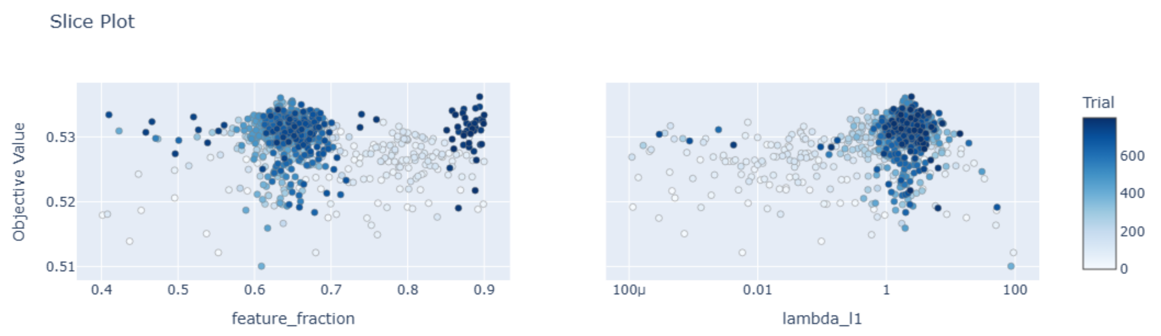
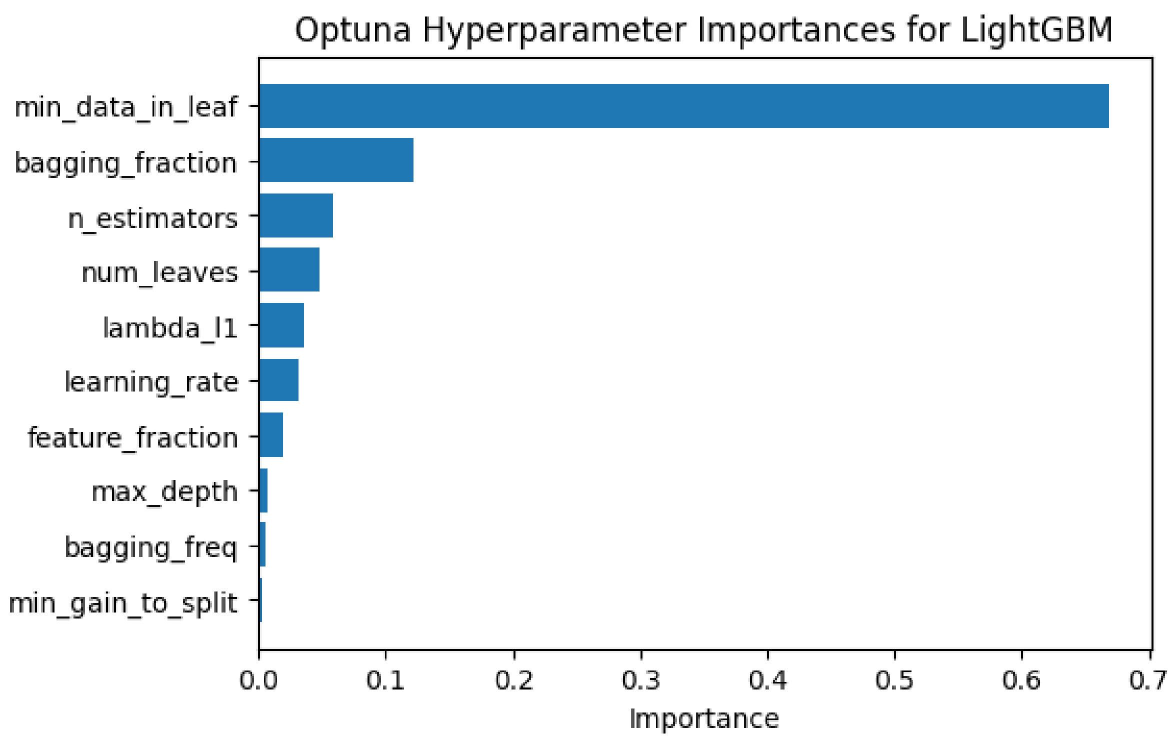


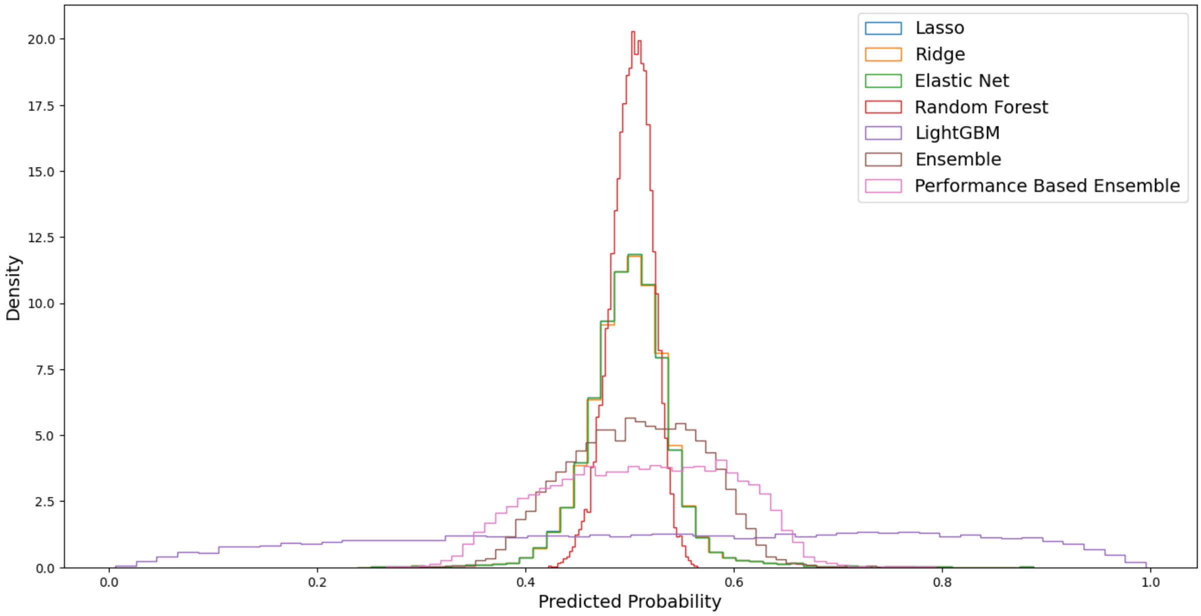
Figure 29: *Optuna Hyperparameter Importances for Light GBM*



Appendix C: Additional Results

1. Distribution of Predicted Probabilities

Figure 30: *Distribution of Predicted Probabilities*



2. Fama French Five Factor Exposure during Economic Cycles

Table 19 : *Fama French Five Factor Exposure during Expansion Cycle*

Panel A: Benchmarks	**Alpha**	**Mkt-RF**	**SMB**	**HML**	**RMW**	**CMA**	**R2**
EW_ben	.0005	.86***	-.53***	.59***	0.01	-.24	.91
SX5E_ben	.0003	.90***	-.63***	.21**	0.11	-.08	.91
Panel B: Naïve Portfolios	**Alpha**	**Mkt-RF**	**SMB**	**HML**	**RMW**	**CMA**	**R2**
EW_ALL	.0003	.90***	-.54***	.55***	-0.16	-.24	.90
EW_TOP5	-.0001	.83***	-.92***	.31	-0.68**	-.35	.69
EW_TOP10	-.0003	.82***	-.65***	.47***	-0.44**	-.33	.81
VW_ALL	.0001	.91***	-.65***	.33***	0.01	-.26	.88
VW_TOP5	.0001	.87***	-.93***	.31	-0.48*	-.34	.68
VW_TOP10	.0006	.84***	-.75***	.27	-0.48**	-.29	.78
Panel C: Mean Variance Portfolios with 30% Weight Constraint	**Alpha**	**Mkt-RF**	**SMB**	**HML**	**RMW**	**CMA**	**R2**
MV_ALL_L_C_30	-.0002	.80***	-.99***	-.02	-1.01***	-.03	.57
MV_TOP5_L_C_30	-.0002	.80***	-.99***	-.00	-1.01***	-.07	.57
MV_TOP10_L_C_30	-.0002	.80***	-.99***	-.02	-1.01***	-.03	.57

Note. *Significant at the at the 10% level, **Significant at the 5% level and *** Significant at the 1%* level.

Table 20 : *Fama French Five Factor Exposure during Slowdown Cycle*

Panel A: Benchmarks	Alpha	Mkt-RF	SMB	HML	RMW	CMA	R2
EW_ben	-.0007	.78***	-0.64***	.39***	.05	.14	.84
SX5E_ben	-.0002	.76***	-0.81***	.17**	.10	-.25*	.84
Panel B: Naïve Portfolios	**Alpha**	**Mkt-RF**	**SMB**	**HML**	**RMW**	**CMA**	**R2**
EW_ALL	-.0001	.74***	-0.80***	.35***	.06	.14	.80
EW_TOP5	.0008	.76***	-1.03***	.68***	-.11	-.04	.65
EW_TOP10	.0003	.74***	-0.99***	.53***	.03	.11	.71
VW_ALL	.0003	.72***	-0.92***	.11	.04	-.10	.79
VW_TOP5	.0011	.74***	-1.09***	.56***	.07	-.05	.65
VW_TOP10	.0003	.75***	-1.07***	.32**	.13	.18	.68
Panel C: Mean Variance Portfolios with 30% Weight Constraint	**Alpha**	**Mkt-RF**	**SMB**	**HML**	**RMW**	**CMA**	**R2**
MV_ALL_L_C_30	.0012	.80***	-0.99***	.61***	-.06	.05	.62
MV_TOP5_L_C_30	.0012	.80***	-0.99***	.62***	-.05	.05	.63
MV_TOP10_L_C_30	.0012	.80***	-0.99***	.61***	-.06	.05	.62

Note. *Significant at the at the 10% level, **Significant at the 5% level and *** Significant at the 1%* level.

Table 21 : *Fama French Five Factor Exposure during Recession Cycle*

Panel A: Benchmarks	Alpha	Mkt-RF	SMB	HML	RMW	CMA	R2
EW_ben	.0019	0.88***	-0.95***	.37	-.11	0.99*	.94
SX5E_ben	.0008	0.87***	-0.89***	.22	-.14	0.76	.93
Panel B: Naïve Portfolios	**Alpha**	**Mkt-RF**	**SMB**	**HML**	**RMW**	**CMA**	**R2**
EW_ALL	.0017	0.90***	-0.88***	.33	.06	1.00*	.94
EW_TOP5	.0039	1.04***	-0.87**	-.15	.06	1.46*	.91
EW_TOP10	.0039	0.93***	-0.98**	.24	.04	1.38*	.92
VW_ALL	.0001	0.92***	-0.72**	.04	-.04	1.09*	.93
VW_TOP5	-.0004	1.04***	-0.73*	-.33	-.16	1.19	.90
VW_TOP10	.0042	0.84***	-1.19***	.37	.08	1.36	.89
Panel C: Mean Variance Portfolios with 30% Weight Constraint	**Alpha**	**Mkt-RF**	**SMB**	**HML**	**RMW**	**CMA**	**R2**
MV_ALL_L_C_30	.0018	1.11***	-0.83*	-.70	-.60	1.20	.87
MV_TOP5_L_C_30	.0018	1.11***	-0.83*	-.70	-.61	1.20	.87
MV_TOP10_L_C_30	.0018	1.11***	-0.83*	-.70	-.61	1.20	.87

Note. *Significant at the at the 10% level, **Significant at the 5% level and *** Significant at the 1%* level.

3. SHAP Values Plots

Figure 31: *SHAP Values Plot for Ridge*

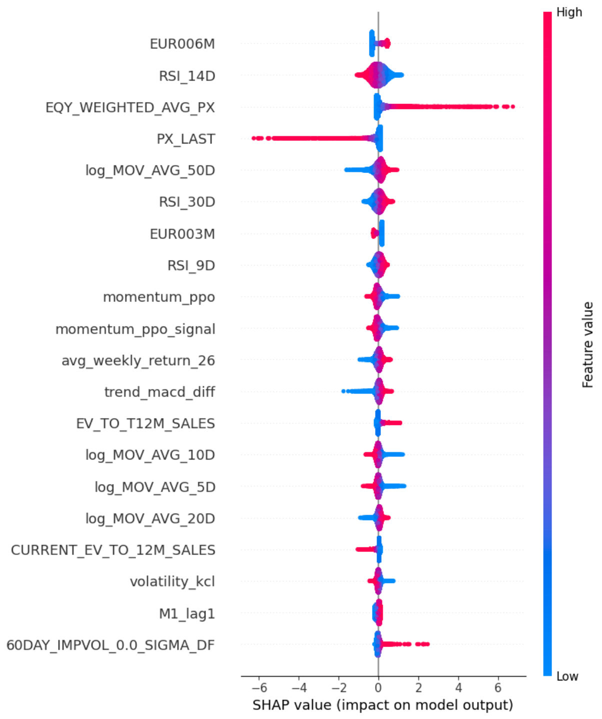


Figure 32: *SHAP Values Plot for Lasso*

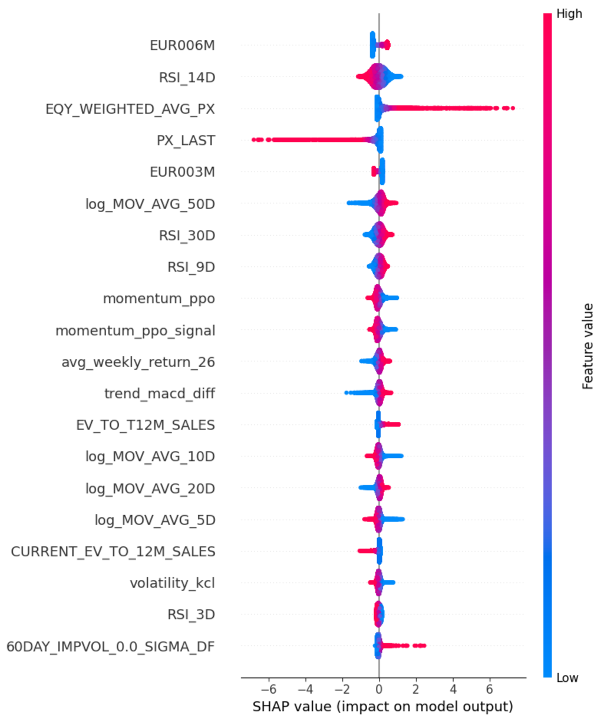


Figure 33: *SHAP Values Plot for Elastic Net*

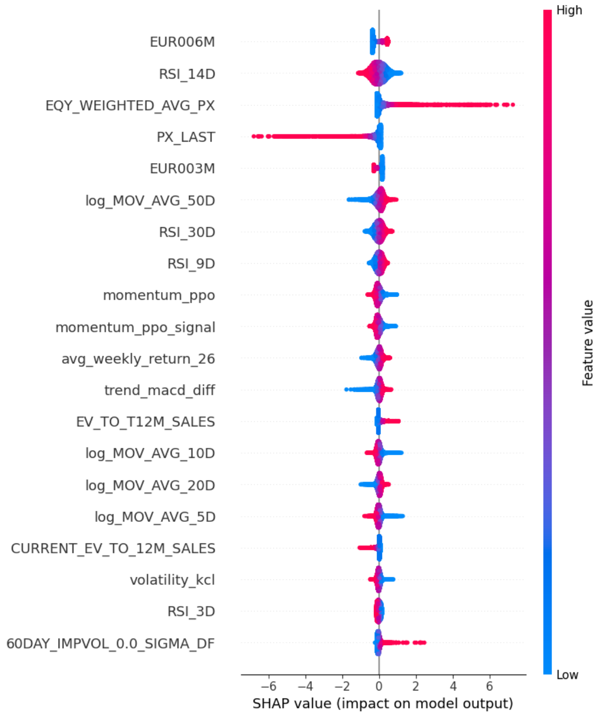
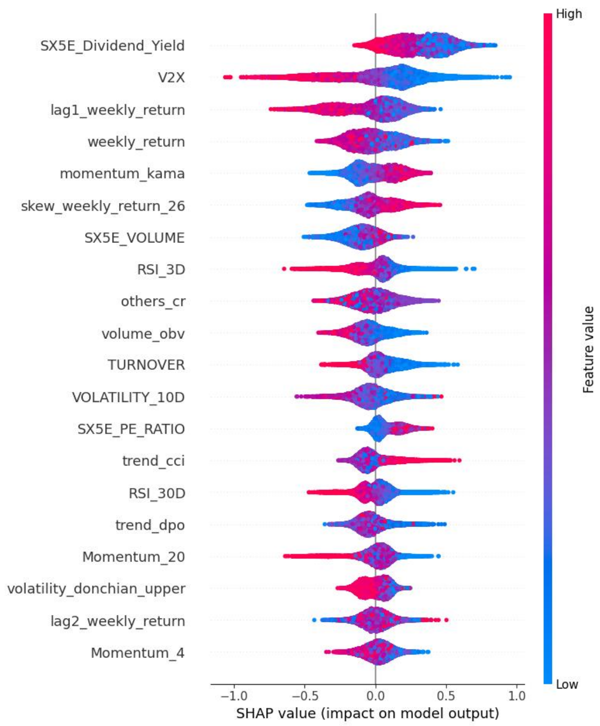


Figure 34: *SHAP Values Plot for Light GBM*



Appendix D: Selected Companies

Table 22: *List of Selected Companies*

Company Name	Bloomberg Ticker	Country	Market Cap as of 28/03/2025 in Billions of Euros (€)
Anheuser-Busch InBev SA/NV	ABI BB Equity	Belgium	116,55
Credit Agricole SA	ACA FP Equity	France	51,27
Koninklijke Ahold Delhaize NV	AD NA Equity	Netherlands	31,84
Adidas AG	ADS GY Equity	Germany	39,94
Aegon Ltd	AGN NA Equity	Netherlands	10,13
Ageas SA/NV	AGS BB Equity	Belgium	10,48
Air Liquide SA	AI FP Equity	France	101,97
Sanofi SA	SAN FP Equity	France	127,55
Airbus SE	AIR FP Equity	France	131,73
Allianz SE	ALV GY Equity	Germany	137,20
ASML Holding NV	ASML NA Equity	Netherlands	246,38
BASF SE	BAS GY Equity	Germany	42,34
Bayer AG	BAYN GY Equity	Germany	22,15
Banco Bilbao Vizcaya Argentaria SA	BBVA SM Equity	Spain	73,34
Bayerische Motoren Werke AG	BMW GY Equity	Germany	48,20
Danone SA	BN FP Equity	France	48,13
BNP Paribas SA	BNP FP Equity	France	88,83
Carrefour SA	CA FP Equity	France	8,99
AXA SA	CS FP Equity	France	88,57
Deutsche Boerse AG	DB1 GY Equity	Germany	51,09
Deutsche Bank AG	DBK GY Equity	Germany	43
Vinci SA	DG FP Equity	France	69,13
Deutsche Post AG	DHL GY Equity	Germany	47,80
Deutsche Telekom AG	DTE GY Equity	Germany	170,19
EssilorLuxottica SA	EL FP Equity	France	121,97
Endesa SA	ELE SM Equity	Spain	26,09
Enel SpA	ENEL IM Equity	Italy	76,60
Eni SpA	ENI IM Equity	Italy	45
E.ON SE	EOAN GY Equity	Germany	36,60
Fresenius SE & Co KGaA	FRE GY Equity	Germany	22,46
Generali	G IM Equity	Italy	51,21
Societe Generale SA	GLE FP Equity	France	34,05
Iberdrola SA	IBE SM Equity	Spain	96,74
Infineon Technologies AG	IFX GY Equity	Germany	40,65
ING Groep NV	INGA NA Equity	Netherlands	57,60
Intesa Sanpaolo SpA	ISP IM Equity	Italy	85,85
Industria de Diseno Textil SA	ITX SM Equity	Spain	144,08

Kering SA	KER FP Equity	France	24,59
Mercedes-Benz Group AG	MBG GY Equity	Germany	53,68
LVMH Moet Hennessy Louis Vuitton SE	MC FP Equity	France	293,63
Muenchener Rueckversicherungs-Gesellschaft AG in Muenchen	MUV2 GY Equity	Germany	78,60
Nordea Bank Abp	NDA FH Equity	Finland	41,86
Nokia Oyj	NOKIA FH Equity	Finland	27,37
L'Oreal SA	OR FP Equity	France	184,39
Orange SA	ORA FP Equity	France	31,52
Koninklijke Philips NV	PHIA NA Equity	Netherlands	22,30
Repsol SA	REP SQ Equity	Spain	14,42
Pernod Ricard SA	RI FP Equity	France	23,743
Hermes International SCA	RMS FP Equity	France	25,716
Renault SA	RNO FP Equity	France	14,05
RWE AG	RWE GY Equity	Germany	24,54
Safran SA	SAF FP Equity	France	104,09
Banco Santander SA	SAN SM Equity	Spain	95,73
SAP SE	SAP GY Equity	Germany	303,13
Cie de Saint-Gobain SA	SGO FP Equity	France	48,19
Siemens AG	SIE GY Equity	Germany	172,8
Schneider Electric SE	SU FP Equity	France	125
Telefonica SA	TEF SQ Equity	Spain	24,51
Telecom Italia SpA/Milano	TIT IM Equity	Italy	6,99
TotalEnergies SE	TTE FP Equity	France	134,41
UniCredit SpA	UCG IM Equity	Italy	82,87
Vivendi SE	VIV FP Equity	France	2,93
Volkswagen AG	VOW GY Equity	Germany	49,26
Wolters Kluwer NV	WKL NA Equity	Netherlands	34,62

Figure 35: *Sum of Market Cap by Country*

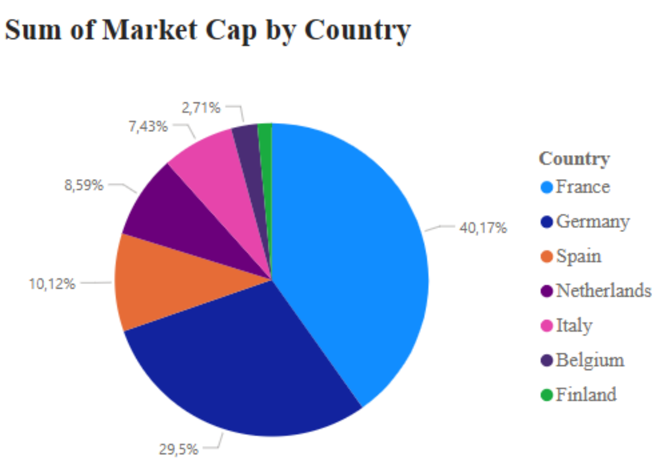


Figure 36: *Count of Companies by Country*

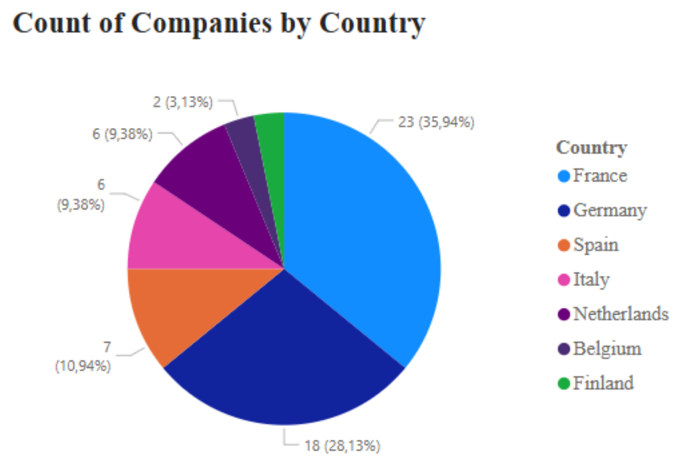
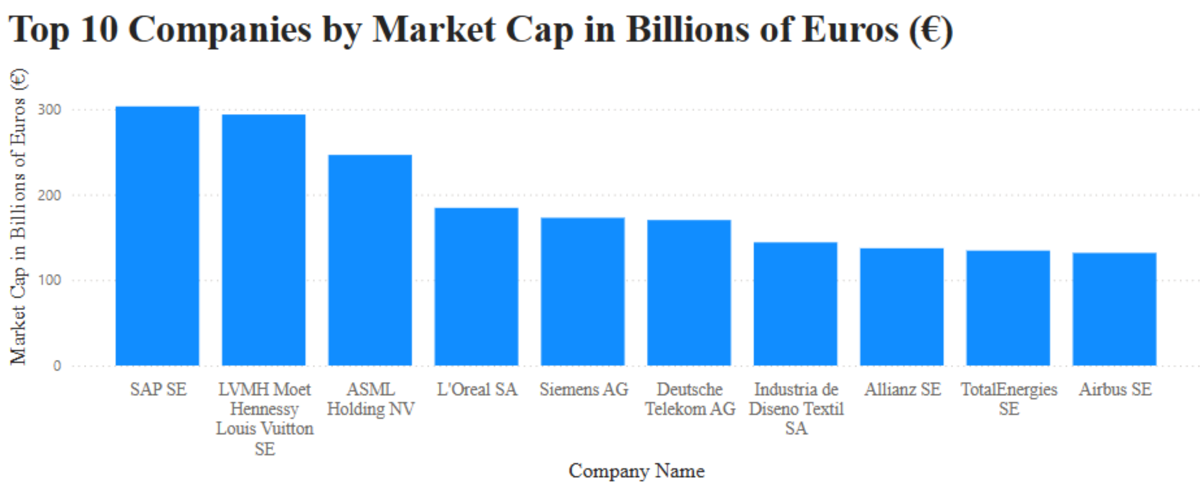


Figure 37: *Top 10 Companies by Market Cap in Billions of Euros (€)*



Appendix E: Features

Table 23: *Overall Features*

Feature	Bloomberg Field Code	Reference	Note
Last/Closing Price	PX_LAST	Wolff & Echterling (2024); Salehpour & Samadzamini (2023)	
Opening Price	PX_OPEN		
High Price	PX_HIGH		

Low Price	PX_LOW		
Bid Price	PX_BID		
Ask Price	PX_ASK		
Cumulative Total Return (Gross Dividend)	CUMULATIVE_TOT _RETURN_GROSS_ DVDS	Wolff & Echterling (2024)	

Table 24: *Fundamental Indicators*

Feature	Bloomberg Field Code	Reference	Note
Current Market Capitalisation	CUR_MKT_CAP	Avramov et al. (2023); Wolff & Echterling (2024)	
Book Value per Share	BOOK_VAL_PER_SH	Avramov et al. (2023); Huang (2012); Wolff & Echterling (2024)	
Earnings per Share Growth (1 & 3 years)	EPS_GROWTH BASIC_EPS_3YR_ AVG_GROWTH	Tsai & Hsiao (2010); Wolff & Echterling (2024)	
Earnings Variability (EPS Std Dev)	STANDARD_ERROR _BASIC_EPS	Wolff & Echterling (2024)	Imp – Wolff
Total Debt to Total Asset (Financial Leverage)	TOT_DEBT_TO_TOT _ASSET	Wolff & Echterling (2024)	
Return on Invested Capital (ROIC)	RETURN_ON_INV _CAPITAL	Wolff & Echterling (2024)	Imp – Wolff
Net Income	NET_INCOME TRAIL_12M_NET_INC	Wolff & Echterling (2024)	
Enterprise Value to Sales	EV_TO_T12M_SALES CURRENT_EV_TO_ 12M_SALES	Wolff & Echterling (2024)	
Free Cash Flow	CF_FREE_CASH_FLOW TRAIL_12M_FREE_ CASH_FLOW		
Free Cash Flow Yield	FREE_CASH_FLOW_ YIELD		
Free Cash Flow per Share	TRAIL_12M_FREE_ CASH_FLOW_PER_SH		

	FREE_CASH_FLOW_PER_SH	
Cash Flow per Share	CASH_FLOW_PER_SH	
Cash from Operating Activities	CF_CASH_FROM_OPER TRAIL_12M_CASH_FROM_OPER	Wolff & Echterling (2024)
Free Cash Flow to Equity	FREE_CASH_FLOW_EQUITY TRAILING_12_MONTH_FCF_TO_EQUITY	Wolff & Echterling (2024)
Enterprise Value to Free Cash Flow	EV_TO_T12M_FREE_CASH_FLOW	Wolff & Echterling (2024)
Dividend Yield	DIVIDEND_INDICATED_YIELD DIVIDEND_12_MONTH_YIELD	Wolff & Echterling (2024)
Operating Margin	OPER_MARGIN TRAIL_12M_OPER_MARGIN	Wolff & Echterling (2024); Tsai & Hsiao (2010); Huang (2012)
Profit Margin	PROF_MARGIN TRAIL_12M_PROF_MARGIN	Wolff & Echterling (2024)
Asset Growth	ASSET_GROWTH	Wolff & Echterling (2024); Tsai & Hsiao (2010); Avramov et al. (2023)
Cash Flow From Investing Activities	CF_CASH_FROM_INV_ACT TRAIL_12M_CASH_FROM_INV_ACT	Wolff & Echterling (2024)
Employee Growth	EMPL_GROWTH	Wolff & Echterling (2024)
Sales Growth	SALES_GROWTH TRAILING_12M_SALES_GROWTH	Wolff & Echterling (2024); Tsai & Hsiao (2010)
Return On Assets (ROA)	RETURN_ON_ASSET	Tsai & Hsiao (2010); Huang (2012)

Operating Income Sequential Growth	OPERATING_INCOME _SEQ_GROWTH	Tsai & Hsiao (2010)
Operating Income or Loss	IS_OPER_INC TRAIL_12M_OPER_INC	Tsai & Hsiao (2010)
Operating Income or Loss per Share	OPER_INC_PER_SH	
Net Income Growth	NET_INC_GROWTH TRAIL_12M_NET_ INC_GROWTH	Tsai & Hsiao (2010)
Continuing Income Growth	CONT_INC_GROWTH	Tsai & Hsiao (2010)
Cash Flow to Total Debt	CFO_TO_TOT_DEBT	Tsai & Hsiao (2010)
Quick Ratio	QUICK_RATIO	Tsai & Hsiao (2010); Huang (2012)
Asset Turnover	ASSET_TURNOVER	Tsai & Hsiao (2010)
Fixed Asset Turnover	NET_FIX_ASSET_TURN	Tsai & Hsiao (2010)
Price to Earnings Ratio	PE_RATIO	Huang (2012)
Price to Book Ratio	PX_TO_BOOK_RATIO	Huang (2012)
Price to Sales Ratio	PX_TO_SALES_RATIO	Huang (2012)
Return on Equity (ROE)	RETURN_COM_EQY	Huang (2012)
Total Debt to Equity	TOT_DEBT_TO_TOT _EQY TOT_DEBT_TO_COM _EQY	Huang (2012)
Current Ratio	CUR_RATIO	Huang (2012)
Total Number of Analysts making recommendations for the security	TOT_ANALYST_REC	Avramov et al. (2023)
Altman Z Score	ALTMAN_Z_SCORE	Avramov et al. (2023)
Income (Loss) from Continuing Operations	IS_INC_BEF_XO_ITEM	Tsai & Hsiao (2010)
Analysts' Recommendation Consensus (1-5)	EQY_REC_CONS	Avramov et al. (2023)
Net Income to Market Capitalisation	NI_to_MktCap = NET_INCOME / CUR_MKT_CAP TRAIL12M_NI_to_MktCap = TRAIL_12M_NET_INC / CUR_MKT_CAP	Wolff & Echterling (2024)

Cash from Operating Activities to Market Capitalisation	CFO_to_MktCap = CF_CASH_FROM_OPER / CUR_MKT_CAP TRAIL12M_CFO_to_MktCap = TRAIL_12M_CASH_FROM_OPER / CUR_MKT_CAP	Wolff & Echterling (2024)	

Table 25: *Technical Indicators*

Feature	Bloomberg Field Code	Reference	Note
Relative Share Price Momentum	REL_SHR_PX_MOMENTUM	Wolff & Echterling (2024); Avramov et al. (2023)	Imp – Wolff
Standardised Log Moving Average 200 Days	Log(PX_LAST/ MOV_AVG_200D)	Wolff & Echterling (2024); Avramov et al. (2023); Li & Bastos (2020); Breitung (2023)	Imp – Wolff Imp – Breitung
Standardised Log Moving Average 100 Days	Log(PX_LAST/ MOV_AVG_100D)	Wolff & Echterling (2024); Avramov et al. (2023); Li & Bastos (2020); Breitung (2023)	
Standardised Log Moving Average 50 Days	Log(PX_LAST/ MOV_AVG_50D)	Wolff & Echterling (2024); Avramov et al. (2023); Li & Bastos (2020); Breitung (2023)	
Standardised Log Moving Average 20 Days	Log(PX_LAST/ MOV_AVG_20D)	Wolff & Echterling (2024); Avramov et al. (2023); Li & Bastos (2020); Breitung (2023)	
Standardised Log Moving Average 10 Days	Log(PX_LAST/ MOV_AVG_10D)	Wolff & Echterling (2024); Avramov et al. (2023); Li & Bastos (2020); Breitung (2023)	
Standardised Log Moving Average 5 Days	Log(PX_LAST/ MOV_AVG_5D)	Wolff & Echterling (2024); Avramov et al. (2023); Li & Bastos	

		(2020); Breitung (2023)	
Relative Strength Index 30 Days	RSI_30D	Wolff & Echterling (2024); Salehpour & Samadzamini (2023); Li & Bastos (2020)	Imp – Wolff
Relative Strength Index 14 Days	RSI_14D	Wolff & Echterling (2024); Salehpour & Samadzamini (2023); Li & Bastos (2020)	Imp – Wolff
Relative Strength Index 9 Days	RSI_9D	Wolff & Echterling (2024); Salehpour & Samadzamini (2023); Li & Bastos (2020)	Imp – Wolff
Relative Strength Index 3 Days	RSI_3D	Wolff & Echterling (2024); Salehpour & Samadzamini (2023); Li & Bastos (2020)	Imp – Wolff
360 Days Volatility	VOLATILITY_360D	Wolff & Echterling (2024); Breitung (2023)	Imp – Wolff
180 Days Volatility	VOLATILITY_180D	Wolff & Echterling (2024); Breitung (2023)	Imp – Wolff
90 Days Volatility	VOLATILITY_90D	Wolff & Echterling (2024); Breitung (2023)	
30 Days Volatility	VOLATILITY_30D	Wolff & Echterling (2024); Breitung (2023)	
10 Days Volatility	VOLATILITY_10D	Wolff & Echterling (2024); Breitung (2023)	
Difference Between Short and Long Term Volatility	Diff_vola = VOLATILITY_30D - VOLATILITY_360D	Breitung (2023)	Imp - Breitung
12 Months Implied Volatility at 0.0 Units of Sigma	12MTH_IMPVOL_0.0_SIGMA_DF		
6 Months Implied Volatility at 0.0 Units of Sigma	6MTH_IMPVOL_0.0_SIGMA_DF		
3 Months Implied Volatility at 0.0 Units of Sigma	3MTH_IMPVOL_0.0_SIGMA_DF		

60 Days Implied Volatility at 0.0 Units of Sigma	60DAY_IMPVOL_0.0_SIGMA_DF	
30 Days Implied Volatility at 0.0 Units of Sigma	30DAY_IMPVOL_0.0_SIGMA_DF	
Total Number of Shares Traded	PX_VOLUME	Salehpour & Samadzamini (2023)
Equity Weighted Average Price	EQY_WEIGHTED_AVG_PX	Li & Bastos (2020)
Total Amount Traded in the Security's Currency	TURNOVER	Avramov et al. (2023)
Average Bid Ask Spread Percentage	AVERAGE_BID_ASK_SPREAD_%	Avramov et al. (2023)
News Sentiment Daily Average	NEWS_SENTIMENT_DAILY_AVG	Salehpour & Samadzamini (2023)
Beta	BETA_RAW_OVERRIDABLE	Wolff & Echterling (2024)
Beta during Bearish Markets	BETA_MINUS	Wolff & Echterling (2024)
Beta during Bullish Markets	BETA_PLUS	Wolff & Echterling (2024)
Momentum 52 Weeks	Momentum_52 =	Wolff & Echterling (2024)
Momentum 40 Weeks	Momentum_40 =	Breitung (2023)
Momentum 26 Weeks	Momentum_26 =	Wolff & Echterling (2024)
Momentum 20 Weeks	Momentum_20 =	Breitung (2023)
Momentum 10 Weeks	Momentum_10 =	Breitung (2023)
Momentum 4 Weeks	Momentum_4 =	Wolff & Echterling (2024)
Weekly Return		Breitung (2023)
Lagged 1 Weekly Return		Breitung (2023)
Lagged 2 Weekly Return		Breitung (2023)
On Balance Volume	volume_obv = ta.volume.on_balance_volume / df['VOLUME']	Breitung (2023); Li & Bastos (2020)
Volume Force Index	volume_fi = ta.volume.force_index / df['PX_LAST']	Breitung (2023)
Volume Simple Moving Average	volume_sma_em = ta.volume.volume_weighted_average_price / df['PX_LAST']	Breitung (2023)
Negative Volume Index	volume_nvi = ta.volume.negative_volume_index/ df['PX_LAST']	Breitung (2023)

Average True Ranges	volume_atr = ta.volatility.average_true_range/ df['PX_LAST']	Breitung (2023); Li & Bastos (2020)	
Kelter Channel Higher	volatility_kcc	Breitung (2023)	
Kelter Channel Lower	volatility_kcl	Breitung (2023)	
Moving Average Convergence Divergence	trend_macd	Breitung (2023); Li & Bastos (2020)	
MACD Histogram – Signal line	trend_macd_diff	Breitung (2023)	
Trend Mass Index	trend_mass_index	Breitung (2023)	
Commodity Channel Index	trend_cci	Breitung (2023)	
Detrended Price Oscillator	trend_dpo	Breitung (2023)	
	trend_kst_diff	Breitung (2023)	
Parabolic SAR Up	trend_psar_up	Breitung (2023)	
Parabolic SAR Down	trend_psar_down	Breitung (2023)	
Ultimate Oscillator	momentum_uo	Breitung (2023)	
Kaufman Adaptive Moving Average	momentum_kama	Breitung (2023)	
Percentage Price Oscillator	momentum_ppo	Breitung (2023)	
PPO signal	momentum_ppo_signal	Breitung (2023)	
PPO Histogram	momentum_ppo_hist	Breitung (2023)	
Others	others_cr = (df['PX_LAST'] - df['PX_OPEN']) / (df['PX_OPEN'] + 1e-9)	Breitung (2023)	
Bollinger Upper Band	volatility_bb_upper	Li & Bastos (2020)	
Bollinger Lower Band	volatility_bb_lower	Li & Bastos (2020)	
Stochastic Oscillator	momentum_stoch	Li & Bastos (2020)	
Stochastic Oscillator Signal	momentum_stoch_signal	Li & Bastos (2020)	
William's %R	momentum_williams_r	Li & Bastos (2020)	
Rate of Change	momentum_roc	Li & Bastos (2020)	
Chailin's Volatility	volume_chaikin_volatility	Li & Bastos (2020)	
Donchain Channel Upper	volatility_donchian_upper	Li & Bastos (2020)	
Donchain Channel Lower	volatility_donchian_lower	Li & Bastos (2020)	
Money Flow Index	volume_mfi	Li & Bastos (2020)	
Accumulation/Distribution Line	volume_adl = ta.volume.acc_dist_index / df['VOLUME']	Li & Bastos (2020)	
Rolling 26 Weeks Average Weekly Return	avg_weekly_return_26	Breitung (2023)	Imp - Breitung
Rolling 26 Weeks Skewness of Weekly Return	skew_weekly_return_26	Breitung (2023)	Imp – Breitung
Rolling 26 Weeks Kurtosis of Weekly Return	kurtosis_weekly_return_26	Breitung (2023)	Imp – Breitung

Rolling 26 Weeks Maximum Spread between Highest and Lowest Weekly Return	max_spread_weekly_return_26	Breitung (2023)	Imp – Breitung
Rolling 26 Weeks Sharpe Ratio	sharpe_ratio_26	Breitung (2023)	Imp – Breitung
Rolling 26 Weeks Adjusted Sharpe Ratio	adj_sharpe_ratio_26	Breitung (2023)	Imp - Breitung

Table 26: *Market State Variables/Macro Economic Indicators*

Feature	Bloomberg Field Code	Reference	Note
Stoxx 50 Price to Earnings Ratio	SX5E [PE_RATIO]	Avramov et al. (2023)	
Stoxx 50 Trading Volume	SX5E [VOLUME]	Flannery & Protopapadakis (2002)	
Stoxx 50 Dividend Yield	SX5E [Dividend_Yield]	Avramov et al. (2023); Boons (2016)	
Stoxx 50 Price to Book Ratio	SX5E [PB_RATIO]		
10 Years German Government Bond Yield	GDBR10	Boons (2016); Celebi & Hönig (2019); Ferrer et al. (2016)	
1 Year German Government Bond Yield	GDBR1	Boons (2016); Ferrer et al. (2016)	
Euribor 6 Months Yield	EUR006M	Ferrer et al. (2016)	
Euribor 3 Months Yield	EUR003M	Celebi & Hönig (2019); Ferrer et al. (2016)	
Euribor 1 Months Yield	EUR001M	Ferrer et al. (2016)	
10 Years US Government Bond Yield	USGG10YR	Flannery & Protopapadakis (2002); Boons (2016)	
3 Months US Government Bond Yield	USGG3M	Avramov et al. (2023); Flannery & Protopapadakis (2002); Boons (2016)	
Stoxx 50 Volatility Index	V2X	Boons (2016)	
Euro High Yield Index			
Composite Business Confidence Indicator Lagged by 1 Month		Celebi & Hönig (2019)	

Composite Consumer Confidence Indicator Lagged by 1 Month		
Monthly Change in Harmonised Index of Consumer (inflation) Lagged by 1 Month		Flannery & Protopapadakis (2002)
Monthly Change in Monetary Aggregate M1 Lagged by 1 Month		Flannery & Protopapadakis (2002); Celebi & Hönig (2019)
Monthly Change in Monetary Aggregate M2 Lagged by 1 Month		Flannery & Protopapadakis (2002)
Monthly Change in Monetary Aggregate M3 Lagged by 1 Month		Celebi & Hönig (2019)
Term Spread	GDBR10 - GDBR1	Avramov et al. (2023); Boons (2016)
US EU Spread	USGG10YR - GDBR10	
Credit Spread	Euro High Yield Index - GDBR10	Avramov et al. (2023); Boons (2016)

* “Imp” means it was found to be important by a certain paper.