Advancing Markowitz: Asset Allocation Forest

Luis Ormonde Bettencourt[1], Alla Petukhina[3], and Anastasija Tetereva *[1,2]

[1]Erasmus University Rotterdam

[2]Tinbergen Institute

[3]Hochschule für Technik und Wirtschaft Berlin

December 2023

Abstract

We propose a novel approach to model the optimal portfolio weights as a nonparametric function of market conditions by introducing an Asset Allocation Forest (AAF). Instead of maximizing the Sharpe ratio (SR) globally, AAF considers local, market conditions-specific optimization problems. We construct a trading strategy for a multi-asset portfolio composed of stocks, bonds, credits, high-yields, and commodities, and illustrate that localizing the portfolio weights by using random forests makes it possible to effectively capture market regimes in a data-driven way. The results demonstrate that the proposed methodology outperforms an equally weighted portfolio, as well as the Sharpe Ratio (SR) maximization strategy and Hidden Markov Model (HMM). These results hold after taking into account trading costs. By constructing Accumulated Local Effects (ALE) plots, we find evidence of 'flight-to-safety' by shifting from the risky assets to the less risky bonds in the periods of higher volatility. Moreover, our model shows a preference for bonds during inflationary periods.

Keywords: black-box methods, interpretable machine learning, portfolio allocation

JEL Classification: C14, C51, C53, C58

*Corresponding author: Anastasija Tetereva, Erasmus School of Economics, Erasmus University Rotterdam, Burg. Oudlaan 50, Rotterdam. Email: tetereva@ese.eur.nl.

1 Introduction

The inclusion of different asset classes in portfolio allocation allows investors to achieve further diversification gains when compared to limiting themselves to a single asset class. However, the heterogeneity between the different classes imposes a new challenge in finding common features or factors that can drive such allocations. For example, one can group equities relative to their dividend yield, respective market capitalization, or leverage, but the same characteristics cannot be applied to commodities or gold. Even if focusing solely on equities and bonds, one cannot assume a yield-to-maturity or duration for a stock.

A proposed alternative, and a common research topic in literature, is to assess how the different asset classes respond to different economic environments ((Ilmanen et al., 2014) and (Yang et al., 2009)), namely focusing on the impact on volatility, covariances, and expected returns. This research will focus on capturing the different dynamics of the (US-based) asset classes: equities, bonds, credits, high-yields, and commodities, through the application of an adapted version of the random forest method (Breiman, 2001), while using macroeconomic and market variables to delimit the different economic environments.

In the development of the AAF strategy, the research strategically addresses two predominant challenges for investors. The first challenge sourced by non-linear dependencies of asset returns on macroeconomic variables (see e.g. (Flannery & Protopapadakis, 2002), (Boucher & Tokpavi, 2019), (Simonian & Wu, 2019a)) to portfolio construction. Traditional approach to addressing the challenge of non-linearity in asset returns and macroeconomic variables is the employment of Hidden Markov Models (HMM) for regimes modeling. While HMMs offer a mechanism to navigate through distinct market regimes, they are not without their limitations. Specifically, HMMs operate under the assumption of state independence and often encounter challenges in parameter estimation, which can constrain their applicability and predictive accuracy in certain contexts. Consequently, the AAF, employs a tree-based model, which dynamically adjusts portfolio weights in response to macroeconomic variables, with an optimization objective tethered to the maximization of the Sharpe ratio. This approach facilitates a market conditions-specific optimization, ensuring a tailored response to localized economic scenarios.

The second challenge pertains to the mitigation of high weight fluctuations, which inherently induce high trading costs. To navigate through this, a turnover penalty is seamlessly integrated into the optimization rule. Furthermore, the incorporation of Bagging in Asset Allocation Forests serves to stabilize portfolio weights, thereby judiciously managing associated trading costs.

The proposed adaptation of the random forest method will be denominated as AAF and will include portfolio optimization under the classic mean-variance framework of Markowitz (1952). The benchmarks defined to evaluate the performance of the model are the naive $\frac{1}{N}$ (equal-weight) strategy and the classical SR optimization where the parameters are estimated directly from the historical data (pure SR optimization), or through an HMM. This analysis will be done both in and out-of-sample and perform statistical tests to verify the significance of the results. Furthermore, we will also take into account transaction costs by measuring the turnover of the strategies, and by including penalties for large allocation changes over time in the model.

The research does find that the AAF can lead to higher performance, even when taking into account transaction costs, and with some degree of statistical significance. One advantage found over the classical SR maximization is the use of explanatory variables to prioritize which returns to take into account, while in the latter case, all returns are equally taken into consideration. This is evident in the weight allocation for commodities in 2022, where the classical SR completely neglects the asset class due to past performance, while the AAF is able to increase the exposure in a period when commodities were the best-performing class.

Because a drawback of Machine Learning methods is the lack of transparency in what drives the model, the second value proposition of this research is to assess how the explanatory variables affect the output of the model. This will be done through the use of Accumulated Local Effects plots (Apley & Zhu, 2020), which allows visualization of the impact a single variable has on the output within small intervals. Moreover, we aim to tackle the topic of Machine Learning methods being a 'black box' and show that interpretation is still possible, while also inferring the economic states from how the different asset classes behave with respect to the input variables.

Additionally, the paper presents a Literature Review section with respect to regime identification methods and the application of Machine Learning methods in Finance and proceeds to present the Data taken into consideration. After going through the Methodology and Empirical results, a final section is reserved for the conclusion and discussion.

2 Asset Allocation Forest

2.1 Literature review

This research follows the current trend of applying Machine Learning (ML) methods to topics where econometric methods traditionally prevail. For example, Gu et al. (2020) estimate a variety of ML methods (including random forests) in order to perform cross-section return predictions of stocks and find them to have a superior forecasting performance than traditional linear models such as OLS. Another example is the case of Wu et al. (2022), which outperforms linear methods in predicting both the level and the trend of the stock-bond correlation by using Machine Learning.

This trend is further explained by Athey and Imbens (2019), and de Prado (2022) where they point out limitations that econometric approaches have in comparison with ML methods. For instance, an econometric model typically implies the assumption of a statistical distribution and estimates the model accordingly, which encompasses the risk of misspecification which consequentially turns the results invalid. Whereas ML methods do not require such assumptions, adapting instead the model to the data provided. Moreover, applying the traditional OLS regression has the advantage that the relations between dependent and independent variables are transparent, however, these are generally assumed to be linear, which often is not the case. Meanwhile, ML methods such as Decision Trees are able to take into account both non-linear relations and interactions between variables, by continuously splitting the original data in order to have more homogeneous subsets.

The context of building diversified portfolios is based on estimating properly the expected (excess) returns and a variance-covariance matrix of the respective assets, which are then used as inputs to maximize the SR (Markowitz (1952) and Sharpe (1964)). This research aims to explore how the resulting allocations differ under different frameworks delimited by macroeconomic and market indicators. The assumption that such means and variances are time-variant and dependent on the economic context has been vastly explored in the literature (Guidolin and Timmermannc (2007)). Considering the example of the correlation between equities and bonds, a cornerstone in asset allocation, Yang et al. (2009) shows that in the US this correlation is lower during recessions when compared to expansions and that a higher correlation also follows after higher inflation and short rate. Not only the correlation varies according to the macroeconomic variables, but it is also an example of an interaction between two variables (inflation and short rates), a dynamic better captured by Machine Learning as opposed to linear regressions. Additionally, Wu et al. (2022) use of ML further shows that other factors such as industrial production, equity volatility, and real yields also have predictive power on the correlation between (US) stocks and bonds.

From the perspective of expected (excess) returns, the work of (Fama & French, 1989) establishes that features such as dividend yield, default, and term spreads have forecasting power over the returns of both stocks and bonds. Similarly to Chen et al. (1986), these variables are said to reflect business conditions, with the example of default spreads being the highest during recessions. As Chen et al. (1986) suggest, the lower income from poor economic conditions requires higher expected returns on the assets in order for a switch from consumption to investment to happen, thus higher default spreads are associated with higher expected returns.

Typically, in the modeling of financial time series, the prevailing market state - such as a bull or bear regime, or a period of heightened volatility - exerts a notable influence on the dynamics of the time series. However, this market state is often unobservable, rendering it a latent variable in modeling contexts. The Hidden Markov Model (HMM), introduced by Baum et al. (1970), serves as a tool to model this concealed market state, providing a framework to analyze and predict the underlying states that influence observable financial time series dynamics. These models do not require explanatory variables, being instead estimated solely using the variable of interest. The usual procedure is to identify the regimes of a particular macroeconomic variable and change allocations based on the predicted regime. An example of this application within multi-asset allocation is given by Kritzman et al. (2012), where the HMM are applied to identify regimes in inflation, economic growth, or market turbulence, and tilt a standard multi-asset portfolio according to the prediction of the next regime.

Another example is given by Uysal and Mulvey (2021), where a HMM is used to identify stock market regimes, alongside a random forest that aims to predict the same stock market regimes and economic recessions. Upon identification of these periods, the historical covariance matrices of each state are used to compute the expected covariance matrix, which is the weighted average of the historical covariance matrices. Moreover, the weights correspond to the predicted probability of the respective state. The resulting matrix is then plugged in to form risk parity portfolios.

These models are however dependent on several assumptions, such as the number of states, the Markov chain dynamic of switching between states, and the statistical distributions of the variables, which are sources for model misspecification. Even if the model is correctly specified, Dacco and Satchell (1999) shows that regime-switching models struggle to outperform the forecast of a random walk since a single regime misclassification has a negative effect on the subsequent forecast.

In our case, the aim is not in identifying economic regimes or states, nor forecasting the value of a particular economic or market indicator. Instead, the focus is on adapting the Machine Learning method random forest (Breiman, 2001) to split a series of multi-asset returns based on a set of macroeconomic and market indicators. While the simple version of a random forest model tries to predict a target variable, the output of the proposed adaptation will deliver the asset allocation implied by the means and covariances, delimited by the macroeconomic and market variables used as input.

An example of a similar extension of the random forests is the Macroeconomic Random Forest of Coulombe (2021). Instead of using a full sample or pre-specified subsets to estimate an auto-regressive model, the model is estimated multiple times on different subsets defined by a variation of the random forest algorithm, implying that parameters are dependent on the (macroeconomic) features. In this research context, we allow for expected returns, volatilities, and correlations of the different asset classes to vary in different economic contexts in order to build optimal portfolios.

The present study adds to the extensive body of literature that emphasizes robust portfolio strategies aimed at minimizing transaction costs and turnover rates. This focus has steered research towards exploring constrained portfolio policies. In a seminal work, (Jagannathan & Ma, 2003) consider portfolios with non-negative constraints. Despite considering a lesser class of portfolios, they demonstrate a better out-of-sample performance. Furthermore, (Fan, Zhang & Yu, 2012) introduce Gross Exposure Constraints, which work similarly but allow negative allocation weights. Another significant trend in this significant trend in this field is the development of robust covariance matrix estimators. Thus, (DeMiguel & Nogales, 2009) construct a portfolio optimization procedure based on M- and S-estimation technique and analyze the stability of the estimator analytically; they also demonstrate empirically that their approach reduces portfolio *turnover*, whereas it slightly improves the out-of-sample performance. (Fan, Wang & Zhong, 2019) construct an elementwise covariance estimator through an M-estimation procedure with Huber loss, providing statistical high-probability guarantees.

3 Methodology

3.1 Mean-Variance and Tangency portfolios

Suppose we have an opportunity to invest into N assets and $r_1, \ldots, r_N$ denote their excess returns. Let $X = (r_1, \ldots, r_N)^\top$ be the multivariate return vector with mean μ and covariance Σ. Then a portfolio with allocation weights $w = (w_1, \ldots, w_N)^\top$ has returns with expectation $\mu_P(w) \stackrel{\text{def}}{=} w^\top \mu$ and variance $\sigma_P^2(w) \stackrel{\text{def}}{=} w^\top \Sigma w$. One of fundamental portfolios on the efficient frontier (Markowitz (1952)) is a tangency portfolio introduced in Sharpe (1964). Investor maximizes a portfolio's Sharpe Ratio (SR) $SR \stackrel{\text{def}}{=} \frac{w^\top \mu - rf}{\sqrt{w^\top \Sigma w}}$. Thus, at each time t, the investor solves the following optimization problem:

$$\begin{aligned} \max_{w_t} \frac{w_t^\top \mu_t - rf}{\sqrt{w_t^\top \Sigma_t w_t}} \\ \text{s.t.} w_t^\top 1_N = 1, \quad w_{i,t} \geq 0 \end{aligned} \tag{1}$$

where $\Sigma_t \stackrel{\text{def}}{=} \mathbf{E}_t\{(X-\mu)(X-\mu)^\top\}$, $\mu_t \stackrel{\text{def}}{=} \mathbf{E}_t(X)$ are the sample covariance matrix or the vector of mean returns, $w_t = (w_{1,t}, \dots, w_{N,t})^\top$ and rf is the risk-free rate and hence $\mathbf{E}_t$ is the expectation operator conditional on the information set available at time t. Several constraints are introduced in (1). First, the weights must sum to one. The second constraint is that the weights must be non-negative, which on the one hand provides a realistic framework for asset managers who are restricted from taking short positions. On the other hand, it also serves as a method for applying shrinkage. As shown by Jagannathan and Ma (2003), these constraints can be viewed as equivalent to optimizing under constraints on the covariance matrix. For example, large negative (positive) weights correspond to large positive (negative) covariances between assets. Therefore, incorporating these constraints effectively shrinks these covariances towards zero. While it is possible for an optimal allocation to include short selling, constraining the weights could result in a suboptimal solution. However, this potential bias is offset by the reduction in the variance of the estimate, ultimately leading to more accurate portfolios with improved out-of-sample performance. It's also important to recognise that relying solely on historical data to estimate expected returns may not be optimal for forecasting. This in turn can be detrimental to mean-variance performance, as pointed out by Chopra and Ziemba (1993).

In current research, instead of directly maximizing expected SR, we use the following modified optimization problem:

$$
\begin{aligned}
\underset{w_t}{max} \, & \frac{w_t^T \mu_t - rf - \lambda |w_t - w_{t-1}|}{\sqrt{w_t^T \Sigma_t w_t}} \\
\text{s.t.} & \sqrt{w_t^T \Sigma_t w_t} = \sigma_{t,EW} \\
& w_t^\top 1_N = 1, \quad w_{i,t} \geq 0
\end{aligned} \tag{2}
$$

where λ is a penalty for changes in portfolio weights (trading costs) and $\sigma_{t,EW}$ is the volatility of the equally weighted portfolio, $i = 1, \dots, N$. We add two additional constraints. The first additional constraint in (2) aims to match the volatility of the resulting portfolio to that of an equal- weighted (EW) portfolio ($\sigma_{t,EW}$). The motivation behind this objective is to ensure that the strategies have a similar level of volatility, thus facilitating a direct comparison of their cumulative return performance. An example of a common, albeit naive, allocation used as a benchmark in the literature is an EW portfolio ($\frac{1}{N}$). When the primary focus is on maximising SR, it is often the case that lower volatility is prioritised, potentially resulting in lower returns. An important consideration when rebalancing the portfolio over time is the inclusion of transaction costs as a critical component of the investment strategy. While the concept of continuously rebalancing to maximise SR at all times seems promising, it requires the introduction of a penalty to reflect the transaction costs associated with buying or selling parts of the portfolio. Transaction costs can have a significant impact on performance, as highlighted in the Detzel, Novy-Marx and Velikov (2023). To address this, rebalancing should only occur if the new optimal portfolio weights promise higher excess returns than the current portfolio weights, taking into account a penalty for transaction costs. There is no consensus in the literature on an appropriate quantification of transaction costs, i.e. the penalty. The introduction of the term λ serves as a penalty for changes in portfolio weights. In the numerator, the term w_{t-1} corresponds to the estimated weight allocation at time $t-1$. Therefore, the magnitude of the difference between this allocation and the new weights reflects the total changes required to achieve the new allocation. By multiplying these changes by a positive λ and subtracting it from the numerator, allocations that minimise turnover are favoured. In this context, λ effectively acts as an implicit transaction cost per percentage unit of the portfolio. Historically, transaction costs have typically been quoted in the literature at fifty basis points - bps (DeMiguel, Garlappi & Uppal, 2009a), but contemporary transaction costs are closer to five bps, as noted by De Nard, Ledoit and Wolf (2021). We consider three possible values for $\lambda \in [0, 0.002, 0.005]$, where 0 means no transaction costs, 20 bps is a moderate assumption for transaction costs, and 50 bps is a conservative assumption, also used by DeMiguel et al. (2009a). Furthermore, De Nard et al. (2021) emphasise that rebalancing on a monthly basis significantly mitigates the challenges posed by transaction costs. Given the limitations imposed by data availability on macroeconomic variables, this research restricts portfolio rebalancing to a monthly frequency, while still accounting for transaction costs.

3.2 Construction of the Asset Allocation Tree

In modern machine learning, tree-based algorithms are particularly popular. They offer the advantages of dealing with complex non-linearities, effectively handling high-dimensional data, mitigating overfitting concerns, and often requiring minimal tuning. This makes tree-based algorithms a preferred machine learning tool compared to neural networks, which can be highly sensitive to the choice of hyperparameters and prone to overfitting if not configured correctly. Therefore, tree-based algorithms are a promising approach for constructing time-varying models, particularly portfolios with time-varying weights. Tree-based models can provide an alternative to Markov switching models and are advantageous for several reasons. First, and most importantly, they allow optimal portfolio weights to be a non-parametric function of covariates such as market conditions. Therefore, in Samitas and Armenatzoglou (2014), the regression tree model is found to outperform the Markov switching model. Furthermore, it is a widely used machine learning technique with an intuitive economic interpretation, using recursive binary splitting on given criteria Loh (2011). Moreover, tree-based models are a suitable solution for the problem at hand, as they can capture complex and non-linear relationships between macroeconomic variables (Carrizosa, Molero-Río & Romero Morales, 2021).

In this paper, we introduce the Asset Allocation Trees (AAT) and Forests, a combination of the local tree-based model and the portfolio construction methods described above. By defining portfolio weights as a non-parametric function of market conditions, more optimal time-varying trading strategies can be developed. In other words, we propose to model the optimal portfolio weights as a function of market conditions, i.e. $w_t^* = f(Z_t, \mu_t, \Sigma_t)$, where $Z_t \in \mathbb{R}^J$ is the state vector of J macroeconomic splitting variables at time t. The function $f(\cdot)$ is given by a tree-based model. To simplify the notation, we will use $w_t^* = f(Z_t)$ instead of $w_t^* = f(Z_t, \mu_t, \Sigma_t)$.

This approach makes it easy to flexibly model evolving weights in a data-driven manner and capture the interactions of different regimes with respect to different market conditions. This is an advantage over latent regime-switching models, where only a few regimes can be estimated and the number of regimes is difficult to determine. Tree-based methods partition the space of predictors, in our case macroeconomic variables, into rectangles by a sequence of splits and provide local conditional optimal portfolios. Let $Z_t = (Z_{1,t}, \ldots, Z_{J,t})$ denote a vector of J market state variables at time t. A tree $\mathcal{T}$ consists of K splits that partition the predictor space into $K+1$ regions (leaf nodes $R_1, \ldots, R_{K+1}$) on the state of J splitting variables $\mathcal{J}$. The state vector of these J splitting variables is denoted by $Z_t \in \mathbb{R}^J$. Now $\mathcal{T}$ assigns each possible value of Z_t to one of $K+1$ terminal nodes denoted by $R_1, \ldots, R_{K+1}$.

Typically, a tree is constructed by finding splits that solve some predefined optimization problem, such as minimizing the mean squared error. In the context of AAT, the target variable is the SR of a portfolio consisting of equities, bonds, credits, commodities and high-yields. Unlike standard regression problems, the objective is not to minimize the mean squared error, but to find the optimal portfolio under specific macroeconomic and market conditions Z_t.

More specifically, AAT searches for an optimal variable at an optimal time that maximizes the weighted average of the SR resulting from the split. In other words, an optimal split is the one that solves the following maximization problem:

$$\max_{c \in \mathbb{R},\ j \in \mathcal{J}} \left(\frac{1}{n_{left}} \mathrm{SR}^*_{t|Z_{j,t}<c}(w_t) + \frac{1}{n_{right}} \mathrm{SR}^*_{t|Z_{j,t}\geq c}(w_t) \right) \tag{3}$$

where $\mathrm{SR}^*_{t|Z_{j,t}<c}(w_t)$ and $\mathrm{SR}^*_{t|Z_{j,t}\geq c}(w_t)$ are the maximum SR of portfolios in corresponding leaves computed by solving (2), and n_{left} and n_{right} are the number of observations in the leaves; and c is a potential threshold chosen from the values of all splitting covariates $Z_{j,t}$, $j = 1, \ldots, J$. This means that the optimization procedure (2) of the left and right leaves is solved based on the returns for which $Z_{j,t} < c$ and $Z_{j,t} \geq c$ respectively. In other words, the optimal weight w_t^* is the one obtained by solving (2) using only the observations in that subset. Following this process to estimate a single tree will return a set of rules that partition Z_t into different subsets corresponding to the leaves of the tree. When the APT is constructed, every leaf corresponds to a set of optimal portfolio weights given the specific market conditions. To make an investment from t to $t+1$, one selects the optimal weights $w_t^* = f(Z_t)$ from the leaf that corresponds to Z_t and uses this weights to make an investment for the next period. For example, Figure 1 shows a tree that uses inflation and TED as splitting variables. If at time t inflation is smaller or equal than 4.13% and TED is larger than 0.12, one should use the weight vector $(9.5, 23.5, 32.9, 0, 34.1)^\top$ to invest from t to $t+1$ for equities, bonds, credits, commodities, and high-yields correspondingly.

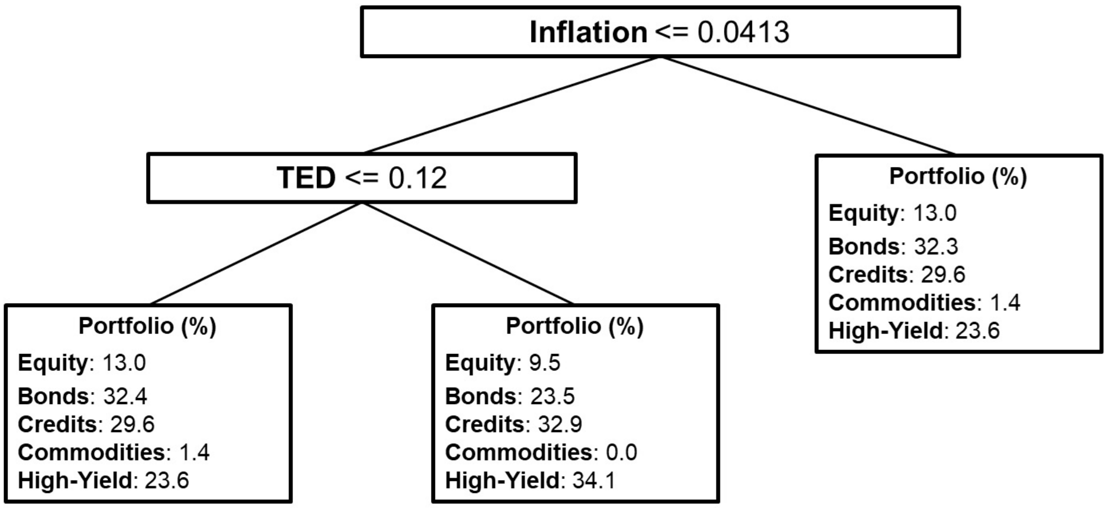


Figure 1: Asset allocation tree on CAPE and credit subindex.

A shortcoming of individual trees is that the greedy algorithm recursively computes local optimal solutions at each split, instead of a global optimal tree, which would be computationally infeasible. Given the large number of possible trees, it would be computationally prohibitive to evaluate all possible combinations of variables and thresholds at each split, and all possible sequences of such splits. As an alternative, Breiman (2001) suggests the use of random forests, where multiple trees are estimated, introducing randomness into the estimation, with the final model being the average of these trees. In the context of optimal AAF, averaging the weights resulting from different trees and renormalising their sum to be equal to one.

The first component of randomness in the random forest is called bootstrap aggregation or bagging by Breiman (1996), where each tree is estimated on a random bootstrapped version of the training set. Breiman (1996) further argues that bagging leads to a significant improvement when applied to unstable models, where a small change in the training set can lead to large changes in the estimated model. This is the case for single trees, where a small change in the training set can lead to different splitting variables and/or thresholds when searching for the optimal combination. In addition, when building portfolios and reestimating the model over time, it is preferable to promote stability, as large changes in portfolio weights lead to higher transaction costs.

A second source of randomness in random forests is the decorrelation between the estimated trees. As suggested by Breiman (2001), instead of going through all the variables when searching for the optimal split, a random subset of these variables is considered when computing each split, resulting in each tree considering different sets of splitting variables, thus decorrelating the trees. This reduces the computational complexity of the estimation by considering fewer splits, while also diversifying each tree with its own decision path rather than repeating the same sequence of locally optimal solutions. Since the prediction of the random forest is the average of the predictions of its trees, the variance of the prediction is the variance of the average predictions, making the random forest a more efficient estimator.

The number of features randomly selected for each tree is also itself a hyperparameter, with the standard practice in regression being the square root of the total number of variables. The second constraint is the minimum number of observations in a leaf. In our case, we want the subsets to have more than 36 observations (equivalent to 3 years of data) in order to have a consistent estimation of mean and the variance-covariance matrix. In addition, the resulting SR should both be higher than that of the original subset to prevent the implied allocations from being worse than the original.

4 Empirical Results

4.1 Data

In this research, the following asset classes are taken into consideration: equities, high-yields, credits, bonds, and commodities. We consider the monthly returns between August 1983 and January 2023. The starting date corresponds to the earliest observations available for high-yields.

The returns for the high-yields, credits, and commodities are retrieved from Bloomberg indices, specified in Appendix 14, while equity returns are sourced from the Fama French online library[1] alongside the risk-free rate. Bond returns are the implied returns from 10-year zero coupon yields published by the Federal Reserve[2]:

$$r_t^{Bond} = \frac{e^{-(9+\frac{11}{12})*yield_t}}{e^{-10*yield_{t-1}}} - 1,$$

where the numerator corresponds to an approximation of the price at time t of a zero-coupon bond which had 10 years to maturity the previous month, while the denominator corresponds to the price of a zero-coupon bond with 10 years to maturity implied by the yield at time $t-1$.

	AR (%)	**SD (%)**	**SR**
Equity	11.57	15.63	0.5312
High-Yields	8.33	8.40	0.6031
Credits	7.18	5.81	0.6731
Bonds	7.84	10.16	0.4505
Commodities	6.09	20.56	0.1372
August 1983 - July 2003			
Equity	12.50	15.95	0.4509
High-Yields	9.61	7.58	0.5674
Credits	9.89	5.41	0.8479
Bonds	11.48	11.07	0.5578
Commodities	10.00	17.39	0.2698
August 2003 - January 2023			
Equity	10.61	15.32	0.6161
High-Yields	7.03	9.17	0.6383
Credits	4.39	6.10	0.5275
Bonds	4.11	9.03	0.3255
Commodities	2.08	23.35	0.0387

Table 1: Descriptive statistics of monthly returns of 5 asset classes. Time period Aug 01 1983 – Jan 31 2022 (462 monthly returns).

We present the performance of the different assets in Table 1 and respective correlations in Table 2. When considering the SR alone, one can see that on one side credits rank the highest while commodities are remarkably under-performing, however when taking into account the high correlation that credits have with bonds (0.77) and high-yields (0.57), it is not necessarily

[1] $https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html$
[2] $https://www.federalreserve.gov/data/nominal-yield-curve.htm$

	Equity	High-Yields	Credits	Bonds	Commodities
Equity		0.640	0.343	0.032	0.229
HY			0.570	0.121	0.231
Credits				0.768	0.079
Bonds					-0.141
August 1983 - July 2003					
Equity		0.530	0.263	0.154	-0.038
HY			0.454	0.259	-0.123
Credits				0.928	-0.030
Bonds					-0.026
August 2003 - January 2023					
Equity		0.745	0.423	-0.130	0.446
HY			0.657	-0.026	0.452
Credits				0.606	0.142
Bonds					-0.270

Table 2: Correlations between Asset Classes. Time period: August 1983 - January 2023 (462 monthly returns).

the case that credits will be the dominating class given possible risk-diversification gains from combining uncorrelated or negatively correlated assets, especially if short positions are not allowed.

Moreover, this research aims to look into how these parameters can change dependent on macroeconomic or market indicators. When computing the same statistics on different subsets of the data, one can see evidence of time variation. For instance, in the second half of the sample, there is a noticeable decrease in the average returns across all assets, specifically for credits, bonds, and commodities which are lower in comparison to the first half of the sample, accompanied by a decrease in the SR, while for equity and high-yields the latter increases.

Table 2 illustrates a change in dynamics of the correlations. For instance, equities, high-yields, and credits become increasingly correlated and less correlated with bonds.

The used macroeconomic and market indicators suggested are motivated by the previous studies and described in Appendix 15.

For market valuation metrics, we use the S&P500 Earnings and Dividend yields, which are indicators of the current (or even forward-looking) valuation of equities, with higher values corresponding to relatively cheaper equities. The Dividend yield is of particular interest since Fama and French (1988) argue it has predictive power over the equity premium, as well in the multi-asset context according to Campbell, Chan and Viceira (2003). Moreover, we also consider Shiller's Cyclically Adjusted Price Earnings Ratio (CAPE), which corresponds to the Price-Earnings ratio using the average of the earnings in the last 10 years and adjusted for inflation (Campbell & Shiller, 1998). Moreover, we also consider the monthly volatility of equities, by averaging the daily squared returns and transforming them into yearly values.

Macroeconomic variables based on US data retrieved from FRED are also included in this research. For example, GDP growth and inflation by Kritzman et al. (2012) identify macro-economic regimes and briefly show that commodities thrive when inflation is up, equity benefit with economic growth, and bonds benefit from deflation. Industrial production (year-on-year) is included as an alternative to GDP in defining economic growth since it is available on a monthly frequency (as opposed to quarterly), and does not have considerable time lags in its announcements.

The term spread as the difference between the 10-year yield published by the Federal Reserve and the 1-Month US treasury bill shows the steepness of the yield curve. This variable has been recorded to be a predictor of future economic activity (Harvey (1993) and Estrella and Hardouvelis (1991)), where a negative value (inverted curve) is associated with the expectation of a recession, which may lead investment away from risky assets.

The credit spread which is the difference between the average of the 20-year BAA corporate yields and the average of the 20 and 30-year treasury yields[3]. This measure is a reflection of the default risk of corporate bonds, with a higher spread translating into a higher risk of default. This variable is also argued to indicate business conditions by Fama and French (1989) with high spreads during recessions, while also being a predictor for bond returns.

Another measure included in this research characterizes the liquidity in the market. Although usually computed using the TED spread, as the difference between the 3-month Libor and the 3-month Treasury Bill, we follow the approach of Franz (2013) in order to assess a longer monthly time series. We use the time series from FRED composed of the difference

[3]The inclusion of the 30-Year yields is justified by the lack of data on FRED for the 20-Year yields between 1987 and 1993

between the Federal Funds Rate and the 3-Month treasury bill, which is mentioned to be highly correlated with the TED spread (0.974). Using the Federal Funds Rates instead of the Libor still acts as an indicator for liquidity in the market since it refers to the overnight charge for banks to borrow amongst each other. We should expect higher values will correspond to higher costs for banks to finance, thus reducing the liquidity in the market.

Finally, we include the use of the Chicago FED National Financial Conditions subindices as in Simonian and Wu (2019b) where the authors use the credit, leverage, and risk subindices individually, and industrial production and inflation, to define economic states. The Credit subindex aims to measure lending conditions in the overall economy, the Leverage subindex is composed of equity and debt measures, while the Risk subindex captures the volatility and funding risk in the financial sector, where positive values of these indicators represent tighter than average conditions. Additionally we include the non-financial leverage subindex, which encompasses the leverage of households and non-financial businesses. Positive values reflect tighter than average conditions within the scope of that subindex, for instance, the credit subindex maximum of 2.66 (Table 3) occurred on December 2008.

Features	**Min**	**Max**	**Median**	**Average**	**Std. Dev.**
US Earnings yield (%)	0.81	10.80	4.97	5.11	1.72
US Dividend Yield (%)	1.11	4.85	2.01	2.30	0.84
Equity Volatility (%)	4.71	91.52	12.47	15.21	9.58
Liquidity Spread	-0.64	1.38	0.13	0.28	0.36
Credit Subindex	-0.72	2.65	-0.11	0.01	0.43
Leverage Subindex	-1.73	3.73	-0.11	0.00	0.79
Risk Subindex	-1.15	2.52	-0.53	-0.35	0.53
Non Financial Leverage Subindex	-2.01	2.65	-0.05	-0.02	1.07
Credit Spread (%)	1.15	5.44	1.78	1.85	0.56
Inflation (%)	-1.96	8.93	2.70	2.82	1.60
IP YoY (%)	-17.26	16.18	2.60	2.04	4.27
Yields 10Y (%)	0.55	13.57	4.86	5.25	2.77
Term Spread (%)	-1.10	5.00	1.95	1.98	1.29
CAPE	8.87	44.20	24.72	24.26	7.83

Table 3: Descriptive Statistics of the Features - 14 Macroeconomic variables. Time period: August 1983 - January 2023 (462 monthly returns).

4.2 In-sample performance

In order to analyze the mechanics of the proposed AAF, we start by training the model using the whole data set and making in-sample predictions. This also acts as a sanity check since the model should outperform a parsimonious model with constant weights resulting from solving (2) over the entire sample (SR strategy) or EW portfolio (EW strategy), due to its flexibility in choosing different allocations for each month.

There are a set of hyper-parameters to be tuned before constructing a random forest: the number of trees, the maximum depth of each tree, the minimum leaf size, the number of variables considered at each split. Since each leaf will correspond to an estimation of the covariance matrix and of the expected return, we set the minimum size of a leaf to 36. This gives us 3 years of data to estimate the weights. Moreover, given the reduced size of our data, we follow the suggestion from Breiman (2001) for reduced data sets and use 200 trees in the forest, and use a maximum depth of 3. As for the subset of explanatory variables used at each split, a common practice is to consider the equivalent to 1/3 or the square root of the total number of variables, which in our case is 4.

	Geometric Return (%)	**Std. Dev. (%)**	**SR**	**Annual TO (%)**	**Max DD (%)**	**Break-Even (%)**
AAF	10.99	6.54	1.18	94	15.92	3.75
SR	8.99	7.67	0.75	20	19.40	N.A.
EW	8.20	7.67	0.64	26	30.63	N.A.

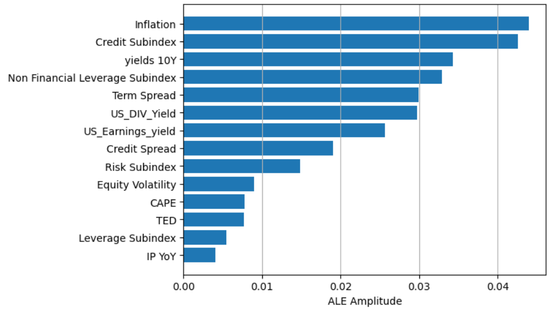
Table 4: Annualized in-sample performance metrics for portolios. Time period: August 1983 - January 2023. The column 'Break-Even' shows the cost per monthly turnover that sets the net performance equal to that of the $\frac{1}{N}$ strategy. 'N.A.' is used when referring to the EW strategy or a strategy with lower turnover and higher performance than that of the EW portfolio.

Looking at Table 4, we confirm that the AAF clearly outperforms in-sample, with an annualized return of almost 2 percentage points above that of SR strategy, while still maintaining a significantly lower volatility. In order to assess if the higher turnover hinders the superior return of the model, we compute the net returns of each month by subtracting that month's turnover multiplied by a constant δ which acts as the implied trading cost. The objective is to find the δ such that a strategy has the same cumulative net return as that of the EW portfolio, and assess the magnitude of such value. The value is given in the last column of Table 4, where one can see that trading costs would have to be 384 bps in order for the model to produce the same net result as the EW portfolio. This is a clearly excessive value, thus the AAF still outperforms the two benchmark models after accounting for trading costs. Moreover, as expected from the matching volatility constraint when optimizing, the SR and EW strategies have the same volatility.

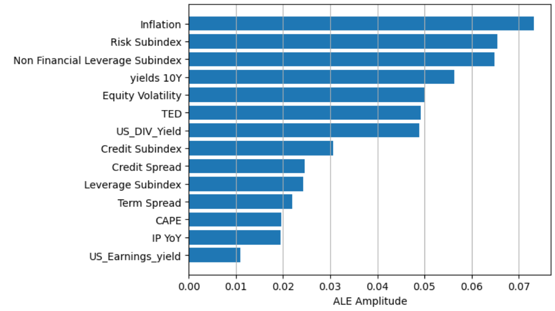
4.3 Out-of-sample performance

An initial estimation dataset consists of 20 years of monthly data, while subsequent estimations accumulate the most recent data points, thus using an expanding window. Furthermore, following Gu et al. (2020), the models and parameters are reestimated every twelve months. We also start with $\lambda = 0$, so we do not take into account any penalty for rebalancing the portfolio. First, we compare the performance of the EW strategy with that of the SR strategy. Figure 3 shows that the performances are close until 2015. After this period, the SR strategy clearly outperforms. This is a period characterised by the negative performance of commodities, during which the SR strategy adapts and has a low weighting in this asset class, giving it an advantage over the simple EW approach. At the beginning of 2022, however, the situation is reversed, with all asset classes except commodities performing negatively. Due to the backward-looking nature of the SR approach, this leads to a declining performance, which is eventually in line with the EW strategy. Both strategies have similar cumulative returns at the end of the period.

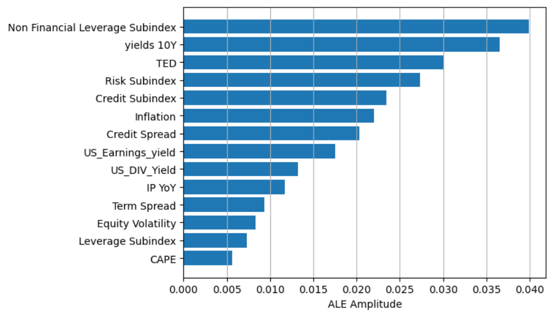
Analysing Table 5, we can see that the annualized returns of all the strategies are similar. However, when the risk is taken into account, it is clear that the AAF method fulfils its main purpose by delivering a higher SR due to the lower volatility. Figure 4a illustrates the dynamics of the weights of each asset class over time. It shows that credit has gone from being the dominant asset class to a marginal position alongside commodities. In contrast, bonds and high-yields become more dominant over time. This is consistent with the characteristics mentioned in Section 4.1, where credits show a decline in performance and an overall increase



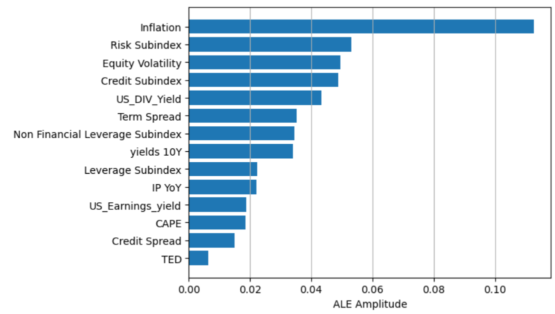
(a) Equity



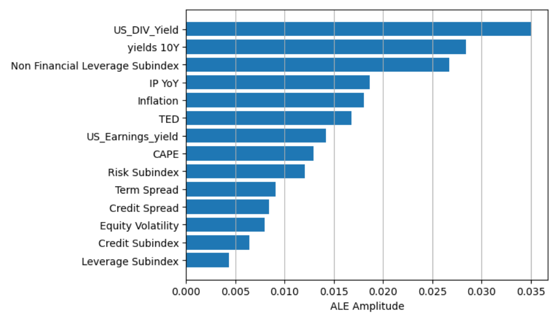
(b) High-Yields



(c) Credits



(d) Bonds



(e) Commodities

Figure 2: Amplitude of ALE values Difference between the highest and lowest ALE values per feature for each asset class. Higher values should indicate a higher variable importance.

	Geometric Return (%)	**Std. Dev. (%)**	**SR**	**Annual TO (%)**	**Max DD (%)**	**Break-Even (%)**
AAF	6.22	7.38	0.68	82	23.12	1.42
SR	5.56	7.56	0.58	38	28.88	1.39
EW	5.40	8.51	0.50	27	30.63	N.A.
HMM	4.94	7.31	0.52	192	20.81	N.A.

Table 5: Annualized out-of-sample performance metrics using an expanding window. Time period: August 2003 - January 2023. The 'Break-Even' column shows 'N.A.' when referring to the EW strategy or a strategy with a lower cumulative gross return than the EW strategy.

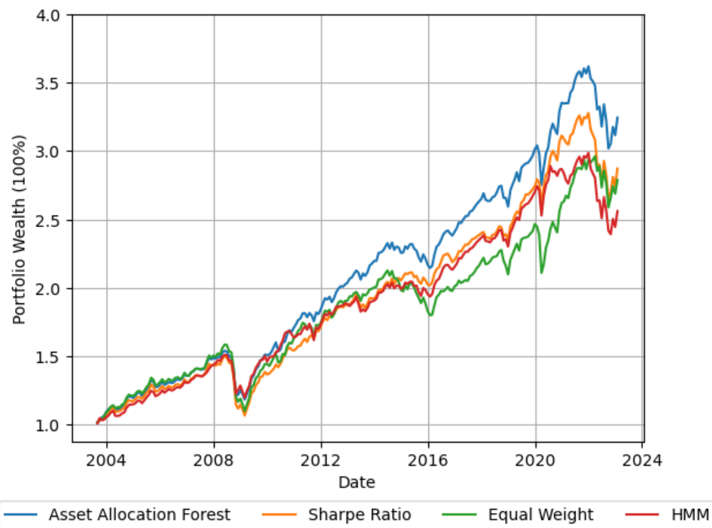


Figure 3: Out-of-sample cumulative portfolio wealth with expanding window. Time Period: August 2003 - January 2023 (229 monthly returns).

in correlations with the other asset classes. The weight of high-yields increases due to the higher risk-adjusted performance. The increase in bonds is due to the reduction in the portfolios' volatility as a result of decreasing (even negative) correlations with equities and high-yields. We can also see the stability of the weights in the later estimates, due to the use of an expanding window, which makes the estimation of the expected returns and the variance-covariance matrix more stable, as a larger sample makes the estimation of the parameters less sensitive to new observations. On the other hand, it results in a relatively low turnover (0.38), slightly higher than that of the EW strategy (0.27). This translates into low transaction costs and also makes it less adaptable to new data, such as the outperformance of commodities in 2022.

The next benchmark considered is the HMM with two states. The performance of this model is underwhelming, as it does not outperform the EW benchmark, but also has the highest turnover. Looking at Figure 4b, we see significant changes in asset allocation. These portfolios appear to be more volatile in the latter part of the sample, with a constant switch between bonds and credits, leading to higher turnover. A possible motivation is the loss of persistence of the states when estimating a HMM with increasing sample size. If the states are less persistent, the probability of being in each state in the future will be more diluted, leading to more pronounced changes in the parameters with increasing step.

(a) SR **(b)** Hidden Markov Model **(c)** AAF

Figure 4: Evolution of portfolio weights over time for the following strategies: (a) SR strategy, (b) HMM strategy, and (c) AAF between August 2003 and January 2023 when considering an expanding window. Asset classes: Equity, High-Yields, Credits, Bonds, Commodities

.

Finally, we turn to the proposed methodology of this research, the AAF strategy. Unlike the previous models, the estimation of portfolio weights is conditional on a number of macroeconomic and market indicators. The out-of-sample portfolio wealth of the AAF, shown in Figure 3, shows that the cumulative return is higher compared to the other strategies. Table 5 also shows that it not only outperforms all the other strategies, but also has a lower standard deviation. This results in the highest SR compared to all other models.

On the other hand, it also has a much higher turnover than the EW and SR strategies. This can be seen in the portfolio composition over time in Figure 4c, where it is noticeable that the weights are constantly changing. This can be explained by the estimation of the forest every twelve months. Although we use an expanding window approach, the AAF portfolios do not stabilise as the size of the training set increases. This is due to its ability to incorporate new information related to macroeconomic conditions. This is the main advantage of the AAF strategy. Although this may lead to higher turnover, we also observe an advantage in the presence of commodities towards the end. While in the case of the global SR strategy, the poor performance over most of the sample alienates the asset class, the AAF distinguishes recent performance from historical performance and includes commodities when they are rising.

Nevertheless, there are some patterns in both cases, namely a decreasing trend in the weights of credits and commodities, and a relatively stable weights of equities around 20%. The AAF strategy has a sharp increase in high-yields in 2009, making it the main asset class until the switch to bonds in 2021. In the other strategies, bonds become the dominant asset class throughout the forecast.

Overall, the performance statistics in Table 5 show that the proposed method has the highest cumulative return and also significantly reduces volatility compared to the EW strategy. This results in the highest SR among the strategies (0.69). However, the constant change in weights results in a higher turnover (82 annualy), which leads us to assess whether the strategy still outperforms after taking trading costs into account. Similarly to the previous section, we calculate the break-even point. Looking at the last column of the Table 5, we can see that both the AAF and the SR show 142 bps and 139 bps respectively for this metric, which are considered excessive trading costs for the period under review. Thus, both approaches are viable when considering trading costs. It is also important to note that

Figure 5: Rolling window (annualized) volatility of the EW portfolio. Time period: July 1995 and January 2023.

the standard deviations of the strategies differ from those of the EW strategy, even though we have the volatility matching constraint. As mentioned in the previous section, this is to be expected in the AAF case due to the averaging of 200 portfolios, but the SR strategy still has a volatility almost one percentage point lower than the EW strategy. If we calculate the 12-year rolling volatility of the EW strategy, as shown in Figure 5, we see an increase starting with the global financial crisis. As we use an expanding window, the volatility to be adjusted will converge to the shift rather than adjusting immediately, resulting in lower volatility overall.

Furthermore, in order to consolidate the results of the AAF, we want to assess whether the SR is statistically different using the test proposed by Ledoit and Wolf (2008). Table 6 shows that the performance of the AAF is indeed significantly better.

	SR (0.58)	**EW (0.50)**	**HMM (0.52)**
AAF (0.68)	1.06	**2.32****	1.30
SR (0.58)	-	0.64	0.47
EW (0.50)	-	-	-0.21

Table 6: Test statistics of the difference between the SR of row and column strategies. In brackets the SR of the respective strategy. Superscript * indicates a 10% significance level, ** indicates a 5% significance level, and *** indicates a 1% significance level.

4.4 Robustness checks

To assess the robustness of the model, we repeat the out-of-sample exercise using a rolling window instead of an expanding one. The window sizes are 12, 15 and 20 years and the same out-of-sample period is considered as in the previous section: from August 2003 to January 2023.

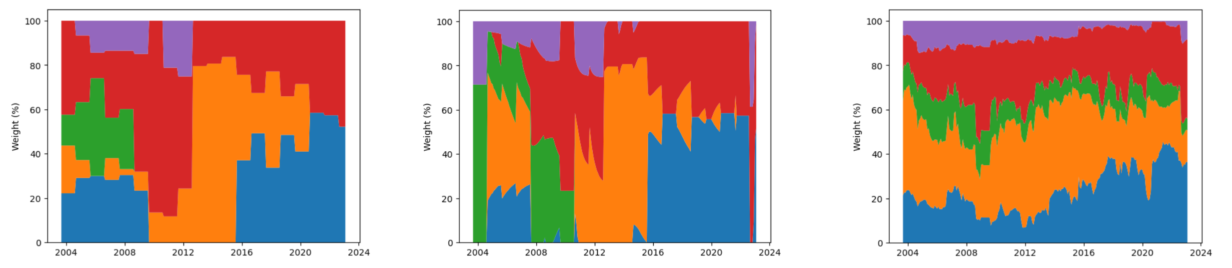
	Geometric Return (%)	**Std. Dev. (%)**	**Sharpe Ratio**	**Annual Turnover (%)**	**Max DD (%)**	**Break-Even (%)**
			20-Year Rolling Window			
AAF	6.71	7.40	0.75	86	21.19	2.08
SR	6.39	8.08	0.65	60	26.59	2.80
EW	5.40	8.51	0.50	27	30.63	N.A.
HMM	6.41	7.50	0.70	126	22.43	0.96
			15-Year Rolling Window			
AAF	6.39	7.41	0.70	88	20.09	1.52
SR	5.58	8.49	0.52	55	27.14	0.62
EW	5.40	8.51	0.50	27	30.63	N.A.
HMM	6.22	8.12	0.62	144	19.99	0.67
			12-Year Rolling Window			
AAF	6.76	7.55	0.74	80	19.77	2.42
SR	6.05	8.45	0.58	64	23.19	1.67
EW	5.40	8.51	0.50	27	30.63	N.A.
HMM	6.90	8.05	0.71	107	21.19	1.76

Table 7: Annualized out-of-sample performance measures using different rolling window sizes. Period: August 2003 - January 2023. The Break-Even column shows 'N.A.' when referring to the EW strategy or a strategy with a lower cumulative gross return than the EW strategy.

Looking at Table 7, we can see that all the strategies perform better than their expanding counterparts, regardless of the size of the window. On the other hand, the AAF manages to maintain a similar standard deviation, while the HMM and SR strategies show a significant increase in volatility. Although not as pronounced, AAF still outperforms the other strategies, with the exception of the 12-year rolling window, where HMM has a higher cumulative return. However, AAF still has a higher SR with lower turnover and has the lowest drawdown.

The improvement in cumulative performance can be explained by the increase in volatility in the later part of the sample, as mentioned in the previous section and illustrated in Figure

5. This increase has an impact on the constraint to match the volatility of an EW strategy, pushing the strategies towards higher weights in riskier asset classes with higher expected returns.



(a) SR **(b)** Hidden Markov Model **(c)** AAF

Figure 6: Portfolio evolution of the strategies: (a) SR, (b) HMM and (c) AAF between August 2003 and January 2023, considering a 12-year rolling window. Asset classes: Equity, High-Yields, Credits, Bonds, Commodities

.

Looking at the evolution of the weights of the SR strategy, we can see that the allocation does not show the same stability as in the case of the widening window. The sharper changes in the portfolio are reflected in the increase in turnover from 0.38 to values above 0.50. However, the increase in turnover is offset by the increase in return as the break-even point increases relative to the EW strategy, except for the 15-year window case. Figure 6a also shows how the poor performance of equities following the global financial crisis led to the absence of this asset class from the portfolio for a long period. However, this is now being reversed due to the subsequent improved performance of equities. Combined with the higher volatility target, the allocation to equities exceeds the expanding window counterpart at the end of the sample (around 40%).

Looking at the other benchmark model, HMM, we see similar patterns in Figure 6b, namely the decline in equities after 2008 followed by the sharp increase in the weight of this asset class. Furthermore, HMM is the only model that consistently reduces turnover when switching to a rolling window. This is the result of omitting the early observations when credit performed better, which leads the HMM to drop the asset class completely. Whereas in the case of the expanding window, there was always a probability of a state in which this asset class was adding value, leading to a constant switch in the calculation of the unconditional expected moments, now credits are ignored, leading to a lower turnover.

On the other hand, the AAF with a 12-year rolling window still manages to maintain the smoothness of the weight evolution (Figure 6c), with a slight increase in turnover. Similarly to the SR case, the increase in turnover is compensated by the higher performance, as the break-even value is higher for all rolling window sizes. The increased equity weighting in the latter test periods is also noticeable and is driven by the improved equity performance and the higher overall volatility mentioned above.

Overall, the proposed AAF manages to consistently deliver a higher SR than its counterparts, while also outperforming a naive EW strategy after taking into account realistic trading costs. Similarly to the previous section, we perform the statistical test on the significance of the SR, with the results shown in Table 8. Although the significance levels vary between 1% and 10%, the AAF strategy is still statistically different from the EW strategy. It also outperforms the global SR strategy for the 12- and 15-year rolling windows.

The same conclusion can be drawn for the SR strategy and the HMM approach. Although the HMM has a higher cumulative performance for a 12-year rolling window, this outperformance is not certain once trading costs are taken into account, given the higher turnover for the EW strategy. It should also be noted that the use of a reduced data set may lead to an unstable estimation of the HMM parameters.

Another aspect to highlight is the improvement in performance when using fewer observations, which should be counterintuitive for machine learning methods that aim to learn from larger data sets. On the one hand, this indicates the potential to explore alternative explanatory variables, while on the other hand it is the result of the limitations of the decisions on SR maximisation, where lower volatility can be prioritised over returns. The latter opens up the possibility of adapting the model by considering different decision rules in the construction portfolio, depending on the established goal.

20-Year Rolling Window			
	SR (0.65)	**EW (0.50)**	**HMM (0.70)**
AAF (0.75)	1.48	**2.97*****	0.45
SR (0.65)	-	1.15	-0.76
EW (0.50)	-	-	-1.33
15-Year Rolling Window			
	SR (0.52)	**EW (0.50)**	**HMM (0.62)**
AAF (0.70)	**2.33****	**1.92***	0.77
SR (0.52)	-	0.15	-1.06
EW (0.50)	-	-	-0.74
12-Year Rolling Window			
	SR (0.58)	**EW (0.50)**	**HMM (0.71)**
AAF (0.74)	**1.81***	**2.04****	0.34
SR (0.58)	-	0.48	**-1.73***
EW (0.50)	-	-	-1.21

Table 8: Test statistics on the difference between SR of row and column strategies. In brackets is the SR of each strategy. Superscript * indicates a 10% significance level, ** indicates a 5% significance level, and *** indicates a 1% significance level.

4.5 Turnover penalty

In this section, we aim to test the effect of including the penalty term parameter λ in (2). As the introduction of this condition further increases the computational burden of the estimation, we apply the penalty for the 20-year rolling window case rather than an expanding one. We consider three possible values for λ: [0, 0.002, 0.005], where 0 is the same scenario as in the previous section, 20 bps is a moderate assumption for trading costs, and 50 bps is a conservative assumption also made by DeMiguel, Garlappi and Uppal (2009b).

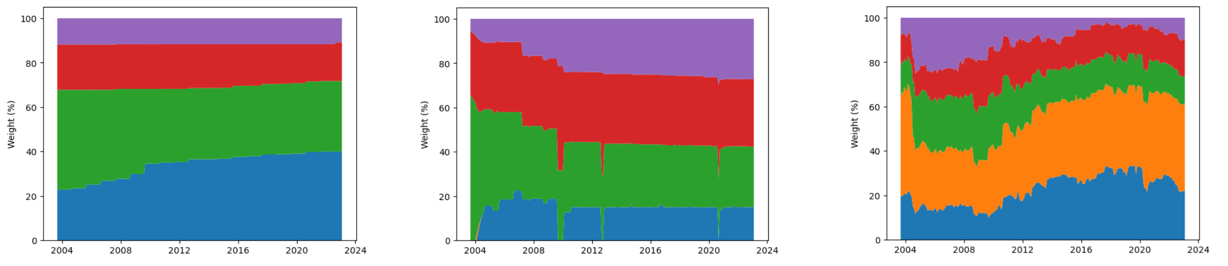
Looking at the results in Table 9, the turnover decreases consistently with λ, as expected since we are now penalizing larger changes in allocation, namely from 0 bps to 50 bps. In the case of the global SR strategy, when using λ of 50 bps, the turnover is even lower than in the EW case, which is the result of an almost constant portfolio allocation as shown in Figure 7a.

Regarding the AAF, it is clear that the constraint on portfolio estimation has a negative

	Geometric Return (%)	Std. Dev.(%)	Sharpe Ratio	Annual Turnover	Max DD (%)	Break-Even (%)
			$\lambda = 0$ **bps**			
AAF	6.71	7.40	0.75	86	21.19	2.08
SR	6.39	8.08	0.65	60	26.59	2.80
EW	5.40	8.51	0.50	27	30.63	N.A.
HMM	6.41	7.50	0.70	126	22.43	0.96
			$\lambda = 20$ **bps**			
AAF	6.49	7.67	0.69	64	23.77	2.82
SR	5.81	8.51	0.54	27	27.82	92.86
EW	5.40	8.51	0.50	27	30.63	N.A.
HMM	5.49	8.23	0.52	92	24.76	0.13
			$\lambda = 50$ **bps**			
AAF	6.44	8.13	0.65	42	27.15	6.30
SR	6.21	8.28	0.61	26	25.81	N.A.
EW	5.40	8.51	0.50	27	30.63	N.A.
HMM	4.05	7.65	0.38	41	24.38	N.A.

Table 9: Out-of-sample performance metrics with penalty parameter on turnover (λ) for the period between August 2003 and January 2023, using a 20-year rolling window, with an adjustment to the turnover penalty parameter (λ). The 'Break-Even' column shows 'N.A.' when referring to the EW strategy, or to a strategy with a gross cumulative return lower than the EW strategy or a strategy with a higher cumulative gross return than the EW strategy but with lower turnover.

impact on the SR, with both a reduction in return performance and an increase in volatility, which is consistent with the fact that we are restricting the options of the model to derive optimal portfolios. Looking at the evolution of the allocation in the 50 bps case in Figure 7c, it is clearly conditioned by the initial estimation, with credits maintaining a more stable proportion instead of being taken over by bonds and high yields (Figure 4c). However, in addition to the reduction in performance, the AAF seems to be compensated by the reduction in turnover, taking into account the increase in the break-even point, particularly in the case of 50 bps, where the transaction costs per percentage point would have to be more than 600 bps to achieve the same return as the EW strategy.



(a) SR **(b)** Hidden Markov Model **(c)** AAF

Figure 7: Portfolio evolution with turnover penalty. Portfolio evolution of the strategies: (a) SR, (b) HMM and (c) AAF between August 2003 and January 2023, considering a 12-year rolling window and $\lambda = 0.005$. Asset classes: Equity, High-Yields, Credits, Bonds, Commodities.

4.6 Relation of Portfolio Weights to Macroeconomic Conditions

One of the drawbacks of using machine learning methods is the lack of transparency. This is evident when using a random forest or any ensemble method, as opposed to a single tree that can be represented graphically. This is not possible when the predictions of a significant number of such trees are combined simultaneously. Therefore, the final issue to be explored is to try to open the 'black box' and understand how the model's predictions are affected by the macroeconomic and market data used as inputs. To do this, we use the Accumulated Local Effects (ALE) plots proposed by Apley and Zhu (2020). We refer to Appendix B.2 for more technical details.

Similarly to the in-sample performance section, we estimate the AAF using the full dataset so that we can focus on the full-sample relationship between the features and the resulting allocations, rather than estimating multiple models using different datasets.

The interpretation of these plots should not focus on the ALE values themselves, as these are by definition shifted to sum to zero, but on their variation with respect to the particular feature. Moreover, although the entire dataset is used to estimate the ALE values, both the 20 lowest and the 20 highest values (totalling about 8% of the observations) with respect to the characteristic are selected for the plots in order to obtain a clearer graphical representation.[4] The characteristics are divided into 20 buckets, so each bucket contains an

[4] In the case of equity volatility, the values range from 4% to 91.5%. If we take the 20 highest values, the upper limit of the interval drops dramatically to 31%, giving a much clearer picture.

average of 24 observations, so all the buckets are represented in the plot even after removing the referenced observations.

It is important to note that the output of the model always corresponds to a portfolio. Therefore, allocating a higher weight to a particular asset class does not imply a positive performance in that specific context, only that it is preferable to the other asset classes considered.

First, we assess the sensitivity of the different asset classes to the different features by calculating the difference between the highest and lowest ALE values. In this context, larger differences imply larger changes in weights, and hence greater sensitivity to the feature. Figure 2 shows that inflation has the greatest impact on equities, high-yields and bonds, while credit is most affected by the non-financial leverage sub-index and commodities by dividend yield. Furthermore, if we look at the values on the horizontal axis, we see that bonds and high-yields have the highest values: 0.11 and 0.07 respectively. This result is consistent with the constant switching between these two asset classes in the forecasting exercises.

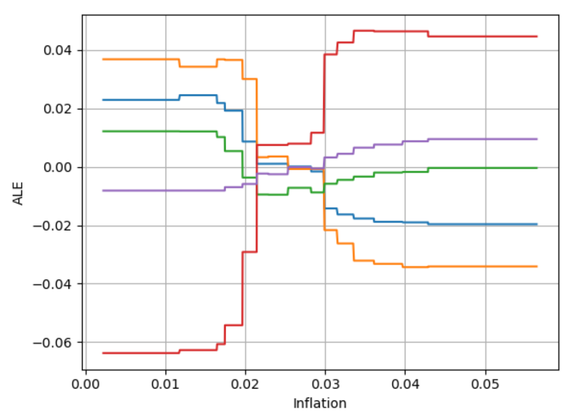


Figure 8: Accumulated Local Effects - Inflation. Accumulated Local Effects plot of Inflation with respect to the asset classes. Asset classes: Equity, High-Yields, Credits, Bonds, Commodities.

Looking at Figure 8, we see that the overall pattern is not linear. It is clear that both commodities and bonds receive larger weights as inflation increases. In the case of commodities, the increase is less pronounced and roughly linear between 1.8% and 4%. Bonds, on the other hand, show a sharp shift between 2% and 3% inflation. On the other hand, equities and high-yields seem to be less favoured as inflation rises, even showing the same sharp changes as bonds, but in the opposite direction and less pronounced. While the positive impact of inflation on commodities is consistent with Ilmanen et al. (2014), who claims that commodities have inflation-hedging properties, the same research shows that government bonds are hit the hardest during high inflation. Moreover, when assigning portfolio tilts in the case of a regime of higher inflation, Kritzman et al. (2012) suggests a reduction in exposure to US 10-year bonds, contrary to our results. It is also noteworthy that our model identifies a threshold around the 2% mark, which is considered to be the general inflation target for central banks.

Assets	Below 2% (145)	Between {2% ; 3%} (141)	Above 3% (188)
Equity	1.15	0.61	0.05
Bonds	-0.04	0.52	0.75
Credits	1.03	0.52	0.56
Commodities	0.17	0.26	0.03
High-Yields	1.36	0.65	0.12

Table 10: Annualized SR of asset classes according to lagged inflation values. The number of observations in each interval is given in brackets.

To get an insight into the performance of asset classes for different inflation values, we calculate the annualized SR under the three different intervals identified by the ALE plot. Looking at Table 10, we can see that the SR of bonds increases at higher inflation intervals, while that of equities and high-yield bonds decreases.

Figure 9 shows the ALE plot with respect to equity volatility, with equities, credit and commodities showing no clear pattern or pronounced changes along the feature. However, high-yields and bonds seem to react symmetrically, with high-yields being replaced by bonds at higher levels of volatility and the reverse at lower levels. The general expectation of increased turbulence in equity markets is a flight to quality. In the case of Kritzman et al. (2012), the regimes identified as market turbulence show a negative risk premium in both equities minus bonds and high-yields minus treasury, consistent with our relationship of a shift from high-yields to bonds as equity volatility increases. Similar to the case of inflation, where the relationship between the feature and the model output does not seem to be linear,

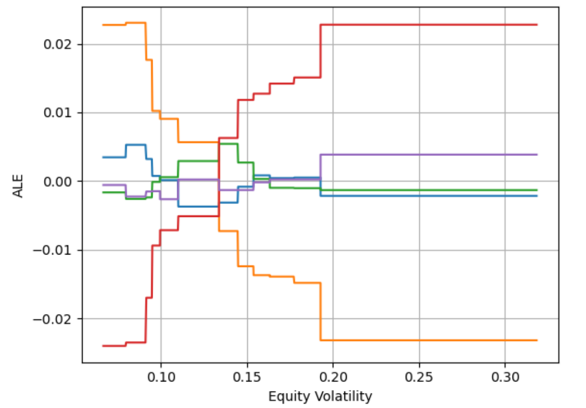


Figure 9: Accumulated Local Effects (ALE) plot of equity volatility across asset classes. Asset classes: Equity, High-Yields, Credits, Bonds, Commodities.

it is noticeable that there are abrupt increases (decreases) for bonds (high-yields) when the following (approximate) thresholds are crossed: 9.5%, 13% and 19%.

Equity Volatility below 9.5% (120)				
Equity	**Bonds**	**Credits**	**Commodities**	**High-Yields**
0.89	0.74	1.12	0.21	1.69
Equity Volatility above 19%(98)				
Equity	**Bonds**	**Credits**	**Commodities**	**High-Yields**
0.24	0.61	0.68	-0.03	0.43

Table 11: Annualized SR of the asset classes during the periods defined by the stock volatility of the previous months. The number of observations in each interval is given in brackets.

Taking the two extreme ranges implied by the above thresholds, below 9.5% and above 19% volatility, we can see in Table 11 that bonds behave as a safe haven in the most turbulent conditions. On the one hand, the drop in SR when volatility is high is less pronounced for bonds, falling from 0.74 to 0.61, while for riskier assets such as high yield bonds the drop is severe, from 1.69 to 0.43. Moreover, bonds become more desirable from a risk perspective in periods of higher volatility. Table 12 shows that while correlations between asset classes generally increase in such periods, the opposite is true for bonds. The correlation between bonds and equities falls from 0.18 to -0.01 or, in the case of high-yields, from 0.47 to -0.03, justifying the increase in the allocation to bonds.

A third example of a non-linear relationship captured by the model concerns the Chicago

Equity Volatility below 9.5% (120)				
	Bonds	**Credits**	**Commodities**	**High-Yields**
Equity	0.18	0.27	0.20	0.54
Bonds		0.96	-0.03	0.47
Credits			0.04	0.60
Commodities				0.21
Equity Volatility above 19% (98)				
	Bonds	**Credits**	**Commodities**	**High-Yields**
Equity	-0.01	0.51	0.31	0.73
Bonds		0.52	-0.06	-0.03
Credits			0.31	0.68
Commodities				0.32

Table 12: Correlations between asset classes during periods delimited by the stock volatility of the previous month. The number in brackets indicates the number of observations.

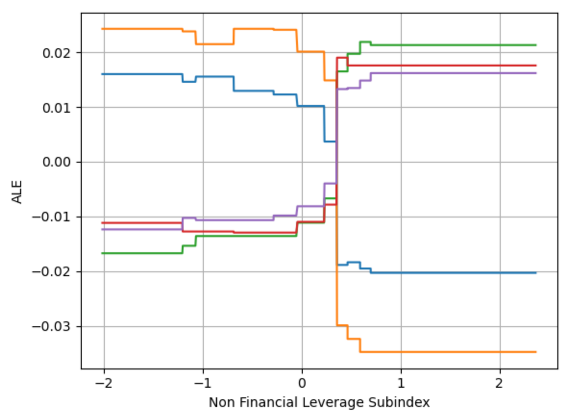


Figure 10: Accumulated Local Effects plot of Non-Financial Leverage subindex in respect to the asset classes. Asset classes: Equity, High-Yields, Credits, Bonds, Commodities

non-financial leverage subindex. As explained by the Chicago FED[5], this subindex is an indicator of the leverage of households and commercial firms, so higher positive values may indicate the risk of future deleveraging and a consequent economic downturn, and thus a switch to less risky assets. In Figure 10 we see a clear switch from risky assets, equities and high-yields, to less risky assets, bonds and credit, and even commodities at a certain threshold of around 0.36. This exemplifies the "safe haven" characteristic of US bonds described by Gulko (2002), where investors move from stocks to bonds after crashes. Given that higher values of this metric may indicate the early stages of a recession, this finding is consistent

[5]https://www.chicagofed.org/research/data/nfci/about

with the research in Gorton and Rouwenhorst (2006), which finds that this stage of the business cycle benefits commodities and hurts equities.

To complement this analysis, we show the different in-sample performances of the asset classes conditional on the threshold. Table 13 shows a clear decrease in the performance of the riskier assets, equities and high-yields, while credits are only slightly affected and bonds show an increase in SR. The remaining sub-indices: risk, leverage and credit also show a similar relationship, with higher values corresponding to increasingly tight financial conditions leading to a shift towards less risky assets. The ALE plots for these sub-indices and the remaining characteristics can be found in the Appendix D.5.

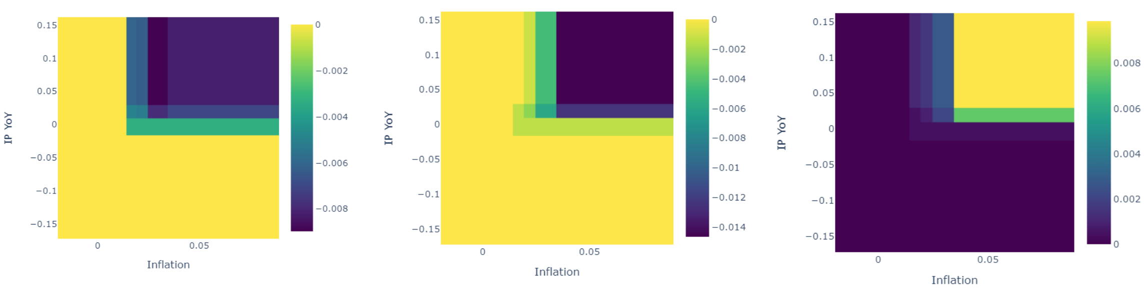
NFL bellow 0.36 (284)				
Equity	**Bonds**	**Credits**	**Commodities**	**High-Yields**
0.92	0.38	0.70	-0.01	1.03
NFL above 0.36 (190)				
Equity	**Bonds**	**Credits**	**Commodities**	**High-Yields**
0.06	0.54	0.64	0.33	0.19

Table 13: Annualized SR of the asset classes during the periods delimited by the previous months' non-financial leverage sub-index. The number of observations in each interval is given in brackets.

While the previous part of the section focused on individual features, it is also instructive to assess how the interaction between features affects the output of the model. This is relevant as we are using tree-based models which make use of such interactions. It is also consistent with the premise of identifying states defined by two variables. Therefore, we illustrate the use of ALE plots to assess the interaction between two different characteristics. The values are plotted in heat maps where, similarly to the previous scenario, instead of looking at the values themselves, we look at how they vary according to the characteristics of interest. As we are looking at two variables at the same time, each plot corresponds to a single asset class. We will look at the plot areas with the greatest difference between the minimum and maximum values, so that the interaction effect is most pronounced.

Following the example in Ilmanen et al. (2014), where growth and inflation are considered simultaneously, we also plot the interactions between the characteristics: inflation and industrial production. We choose to show the ALE plots without removing the first-order effects in order to have a complete overview of the impact on asset classes, rather than limiting ourselves to the second-order effect. The ALE plots in Figure 11, identify the upper right quadrant (higher inflation and industrial production) where the interaction is stronger. Similar to Ilmanen et al. (2014), equities and bonds seem to be negatively affected by this scenario, while commodities seem to benefit.

Overall, the ALE plots provide insights into the non-linear relationships captured by the model and, to some extent, the interaction effects. However, it should be remembered that the model aims to maximise the SR for a given set of asset classes.



(a) Equity **(b)** Bonds **(c)** Commodities

Figure 11: Accumulated Local Effects for interactions between inflation and annual industrial production growth (IP YoY) for equities, bonds and commodities.

5 Conclusion

This study introduces AAF to model optimal weights as a function of market conditions without the need for statistical assumptions, by adapting the algorithm of the machine learning method random forests. This extension of the method is applied in the context of a multi-asset portfolio composed of equities, bonds, credit, high-yields and commodities, while attempting to capture their dynamics under different states defined by the proposed indicators.

Our results show that this approach leads to a superior out-of-sample SR (0.68) compared to the SR (0.58), EW (0.50) and HMM (0.52) benchmarks. In order to take transaction costs into account, the break-even point is calculated with respect to the EW portfolio, resulting in an excess value of 142 bps, implying that the strategy is profitable after transaction costs. However, it is worth noting that the model still delivers a significantly higher turnover (0.82) compared to the EW case (0.27). To address this, we also test the inclusion of a regularisation term to penalize portfolio changes and are able to significantly reduce the turnover to 0.42, while still maintaining a higher SR (0.65). Furthermore, the robustness of the results is further verified by using different sizes of rolling windows when estimating the model. The model provides a higher SR, which is then further tested to be statistically significant.

In addition to the performance of the model, we have also assessed how the input variables affect the result and have drawn conclusions about the dynamics of the different asset classes through the use of ALE plots. From this analysis, some relationships were to be expected, such as higher volatility in equities leading to a flight to safety by shifting from risky assets to less risky bonds. In addition, although inflationary periods should have a negative impact on bonds with the expected rise in interest rates, our model shows a preference for bonds in this scenario. It is important to note that the model uses a one-month lag for the explanatory variables with respect to returns, so the initial inflationary shock has already occurred.

Finally, while the results are promising, it should be noted that they are conditional on the limited universe of asset classes considered. Further research with additional assets such as gold, emerging market equities, Treasury Inflation-Protected Securities (TIPS) or currencies may yield different ratios for the same characteristics. In addition, the portfolio allocation uses historical averages as expected returns and performs SR maximization to find optimal portfolio allocations, which may not be in line with a potential investor's views or preferences. As mentioned in the research, SR maximisation may prioritise the reduction of volatility, while at the same time reducing the expected return. Nevertheless, this research is also a basis for these further extensions, not only with the proposed method, but also in the interpretation of the 'black box' inherent in machine learning.

References

Apley, D. W. & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B*, *82.4*, 1059—1086.

Athey, S. & Imbens, G. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *22*(1), 685—725.

Baum, L., Petrie, T., Soules, G. W. & Weiss, N. (1970). A maximization techinique occuring in the statistical analysis of probabilistic functions of markov chains. *The Annuals of Mathematical Statistics*, *41*(1), 164—171.

Boucher, C. & Tokpavi, S. (2019). Stocks and bonds: Flight-to-safety for ever? *Journal of International Money and Finance*, *95*, 27–43.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123—140.

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

Campbell, J. Y., Chan, Y. L. & Viceira, L. M. (2003). A multivariate model of strategic asset allocation. *Journal of Financial Economics*, *67*, 41—80.

Campbell, J. Y. & Shiller, R. J. (1998). Valuation ratios and the long-run stock market outlook. *The Journal of Portfolio Management*, *24*(2), 11—26.

Carrizosa, E., Molero-Río, C. & Romero Morales, D. (2021). Mathematical optimization in classification and regression trees. *Top*, *29*(1), 5–33.

Chen, N.-F., Roll, R. & Ross, S. A. (1986). Economic forces and the stock market. *The Journal of Business*, *59*(3), 383—403.

Chopra, V. K. & Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, *19*(2), 6—11.

Coulombe, P. G. (2021). *The macroeconomy as a random forest.*

Dacco, R. & Satchell, S. (1999). Why do regime-switching models forecast so badly? *Journal of Forecasting*, *18*, 1—16.

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

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

DeMiguel, V. & Nogales, F. J. (2009). Portfolio selection with robust estimation. *Operations Research*, *57*(3), 560–577.

De Nard, G., Ledoit, O. & Wolf, M. (2021). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics*, *19*(2), 236–257.

de Prado, M. L. (2022). Machine learning for econometricians: The readme manual. *The Journal of Financial Data Science*, *4*(3), 10—30.

Detzel, A., Novy-Marx, R. & Velikov, M. (2023). Model comparison with transaction costs. *The Journal of Finance*.

Estrella, A. & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. *The Journal of Finance*, *46*(2), 555—576.

Fama, E. F. & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economic*, *22*, 3—25.

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

Fan, J., Wang, W. & Zhong, Y. (2019). Robust covariance estimation for approximate factor models. *Journal of Econometrics*, *208*(1), 5–22.

Fan, J., Zhang, J. & Yu, K. (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, *107*(498), 592–606.

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

Franz, R. (2013). *Macro-based parametric asset allocation.*

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of statistics*, 1189–1232.

Gorton, G. & Rouwenhorst, K. G. (2006). Facts and fantasies about commodity futures. *Financial Analysts Journal*, *62*(2), 47—68.

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

Guidolin, M. & Timmermannc, A. (2007). Asset allocation under multivariate regime switching. *Journal of Economic Dynamics Control*, *31*, 3503—3544.

Gulko, L. (2002). Decoupling. *The Journal of Portfolio Management*, *28*(3), 59—66.

Harvey, C. R. (1993). Term structure forecasts economic growth. *Financial Analysts Journal*, *49*(3), 6—8.

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

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

Kritzman, M., Page, S. & Turkington, D. (2012). Regime shifts: Implications for dynamic strategies. *Financial Analysts Journal*, *68*(3), 22—39.

Ledoit, O. & Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, *15*(5), 850—859.

Ledoit, O. & Wolf, M. (2011). Robust performances hypothesis testing with the variance. *Wilmott*, *55*, 86—89.

Loh, W.-Y. (2011). Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, *1*(1), 14–23.

Markowitz, H. M. (1952). Portfolio selection. *The Journal of Finance*, *7*, 77—91.

Samitas, A. & Armenatzoglou, A. (2014). Regression tree model versus markov regime switching: a comparison for electricity spot price modelling and forecasting. *Operational*

Research, *14*, 319–340.

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

Simonian, J. & Wu, C. (2019a). Minsky vs. machine: New foundations for quant-macro investing. *The Journal of Financial Data Science*.

Simonian, J. & Wu, C. (2019b). Minsky vs. machine: New foundations for quant-macro investing. *The Journal of Financial Data Science*, *1*(2), 94—110.

Uysal, A. S. & Mulvey, J. M. (2021). A machine learning approach in regime-switching risk parity portfolios. *Journal of Financial Data Science*, *3*(2), 87—108.

Wu, B., DiCiurcio, K. J., Yeo, B. & Wang, Q. (2022). Forecasting us equity and bond correlation—a machine learning approach. *The Journal of Financial Data Science*, *4*(1), 76—86.

Yang, J., Zhou, Y. & Wang, Z. (2009). The stock–bond correlation and macroeconomic conditions: One and a half centuries of evidence. *Journal of Banking Finance*, *33*(4), 670—680.

A Appendix: Data

Asset Class	Source	Ticker
Equity	Fama French library	'Mkt-RF'+'RF'
Bonds	10-year maturity US ZCB yields	SVENY10
Credits	ICE BofA US Corporate TR	BAMLCC0A0CMTRIV (FRED)
High Yields	Bloomberg US Corporate High Yield Total Return Index	LF98TRUU (Bloomberg)
Commodities	S&P GSCI Commodity TR	S&P GSCI Commodity TR
Risk-Free	Fama French library	RF

Table 14: Asset Classes sources

Feature	Source	Ticker
US Earnings Yield	Quandl	MULTPL/SP500_EARNINGS_YIELD_MONTH
US Dividend Yield	Quandl	MULTPL/SP500_DIV_YIELD_MONTH
Industrial Production	FRED	INDPRO
CPI-Urban (Inflation)	FRED	CPIAUCSL
CAPE	econ.yale.ed	CAPE
20-Year US Yields	FRED	GS20
30-Year US Yields	FRED	GS30
BAA 20-Year yields	FRED	BAA
Liquidity spread	FRED	TB3SMFFM
Credit Subindex	FRED	NFCICREDIT
Leverage Subindex	FRED	NFCILEVERAGE
Risk Subindex	FRED	NFCIRISK
Non-Financial Leverage Subindex	FRED	NFCINONFINLEVERAGE
Daily Equity Returns (Equity Volatility)	Fama French Library	NA

Table 15: Macroeconomic and Market Features sources

B Appendix: Methodology

B.1 Hidden Markov Model

An HMM assumes that a variable $r_t = (r_{t,1}, \dots, r_{t,N})^\top$ follows a distribution whose parameters depend on the regime at time t ($S_t = i$). In our case, y_t corresponds to the monthly returns of the asset classes and will be assumed to follow a multivariate normal distribution, such that:

$$r_t | S_t = i \sim N(\mu_i^{HMM}, \Sigma_i^{HMM}), i = \{1, 2, ..., k\}, \tag{4}$$

with this being the case for k states. Furthermore, the Markov component in the model is present in the dynamics of changing states, where the probability of going to a specific state the next period is solely dependent on the current state:

$$P(S_{t+1} = j|S_t = i) = P(S_{t+1} = j|S_t = i, S_{s<t}).$$

In order to estimate this model, two aspects have to be defined apriori: the number of states k in (4), and the assumed distribution (multivariate normal in this case). Thereafter, the Baum-Welch algorithm (Baum et al., 1970) is applied to get the following estimates: the parameters of the normal distribution (μ_i, Σ_i); the transition probability matrix Γ, where the element in row i and column j corresponds to $\gamma_{i,j} = P(S_{t+1} = j|S_t = i)$; initial probabilities $\pi = (\pi_1, ..., \pi_k)$, the probabilities of being in the different states at time $t = 0$, and the implied probabilities of being in a specific state in the subsequent periods: vector p_t where element i corresponds to $p_{t,i} = P(S_t = i)$.

The incorporation of HMM in asset allocation uses the formulation in (2), but replacing the estimated parameters μ and Σ by the expected counterparts in the next period according to the model:

$$\pi_{t+h} = (\Gamma^h)^T p_t$$

$$\hat{\mu}_{t+h} = \sum_{i=1}^{k} \pi_{t+h,i} \mu_i^{HMM}$$

$$\hat{\Sigma}_{t+h} = \sum_{i=1}^{k} \pi_{t+h,i} \Sigma_i^{HMM}, \tag{5}$$

where in equation (B.1) we compute the probabilities of being in the different states h months in the future, which are then used as weights in the computation of $\hat{\mu}_{t+h}$ and $\hat{\Sigma}_{t+h}$ as the weighted average of the mean and covariance matrix in the different states.

B.2 Accumulated Local Effects

The aim of of accumulated local effects (ALE) by Apley and Zhu (2020) is to measure the isolated effect of a single feature on the weights of a model within small intervals.

To achieve this, we divide the observations into k quantiles according to the respective feature (Z_j) with boundaries: $z_{0,j}, z_{1,j}, ..., z_{k,j}$, so that the observations in quantile i are: $Z_{i,j} : z_{i-1,j} < Z_j \leq z_{i,j}, \text{for } i = \{1, 2, ...k\}$. The ALE values are the average of the differences in the predictions when the feature j is replaced by the upper and lower bounds ($z_{i,j}$ and $z_{i-1,j}$) in the observations within the quantile:

$$\widehat{ALE}_j(x) = constant + \sum_{s=1}^{Z_j(x)} \frac{1}{n_{Z_{s,j}}} \sum_{i:x_i \in Z_{s,j}} [\hat{f}(x_{i,-j}, z_{i,j}) - \hat{f}(x_{i,-j}, z_{i-1,j})],$$

where $\widehat{f}(x)$ is the prediction of the model for the observation x, $(x_{i,-j}, z_{i,j})$ is the observation x_i with the value for the feature j replaced by $z_{i,j}$, $n_{Z_{i,j}}$ is the number of observations in the quantile $Z_{i,j}$, $Z_j(x)$ corresponds to the quantile of x with respect to feature j, and 'constant' is the value that sets the mean of all $\widehat{ALE}_j(x)$ to zero.

ALE plots are preferred to the more popular partial dependence plots for the following reasons: Partial dependence plots, as introduced by Friedman (2001), make predictions using the entire data set, changing only the variable of interest. However, this approach leads to inconsistent observations with uninformative predictions. Furthermore, when we average the predictions, we also average the effects of the remaining variables on the prediction. Therefore, by calculating the difference when only the variable of interest changes, we isolate its specific effect.

C Performance metrics

This section aims to specify metrics and methods used to compare the performances of the different methods.

The annual turnover of a strategy is a measure of how many positions in the portfolio are bought/sold throughout time. A strategy with high turnover is likely to imply high transaction costs, which may wipe out the potential gains. We apply the turnover definition applied by DeMiguel et al. (2009b):

$$\text{Annual Turnover} = \frac{12}{T-1} \sum_{t=1}^{T-1} \sum_{i=1}^{N} |w_{i,t+1} - w_{i,t}|,$$

where $w_{i,t+1}$ is the target weight estimated for asset i at the beginning of month $t+1$, while $w_{i,t}$ is the weight of asset i before rebalancing in month $t+1$. Since we are accounting for both the buys and sells, we are computing a double-sided turnover.

The maximum drawdown is the highest loss relative to the highest point (up until that moment) throughout the lifetime of the portfolio:

$$\text{Max Drawdown} = min_{0 \leq t_1 < t_2 \leq T} \frac{r^p_{t_2}}{r^p_{t_1}} - 1,$$

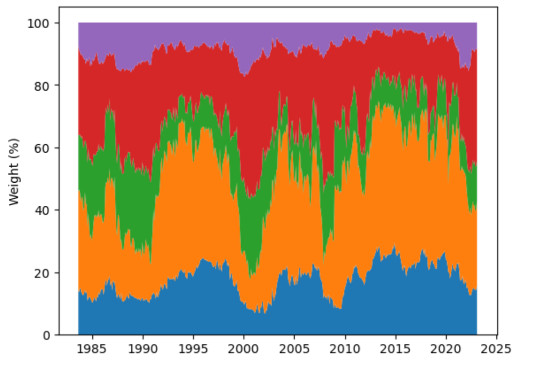
where r^p_t is the cumulative return of the respective strategy at month t.

In order to assess if different portfolio strategies do in fact differ from each other in a statistically significant way, we perform the test proposed by Ledoit and Wolf (2008). This test has the advantage of taking into account possible heteroskedasticity and autocorrelations in the first and second moments of returns. Moreover, we may also assess if the volatilities of the different strategies differ significantly using a similar approach proposed in Ledoit and Wolf (2011). The implementation of these tests was done using the code provided in Michael Wolf's library[6].

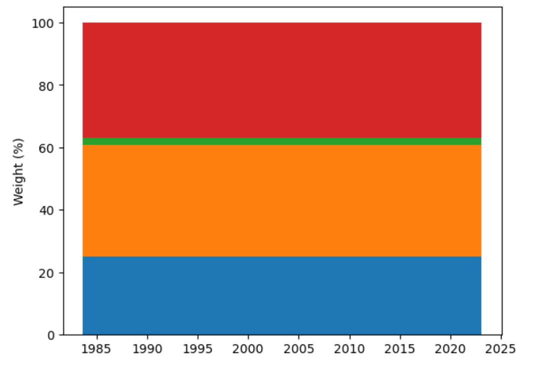
[6]https://www.econ.uzh.ch/en/people/faculty/wolf/publications.html

D Appendix: Empirical Results

D.1 In-sample Performance



(a) Asset Allocation Forest



(b) Sharpe Ratio

Figure 12: Portfolio Evolution of in-sample estimation. Time Period: Aug 1983 - Jan 2023. Asset classes: Equity, High-Yields, Credits, Bonds, Commodities.

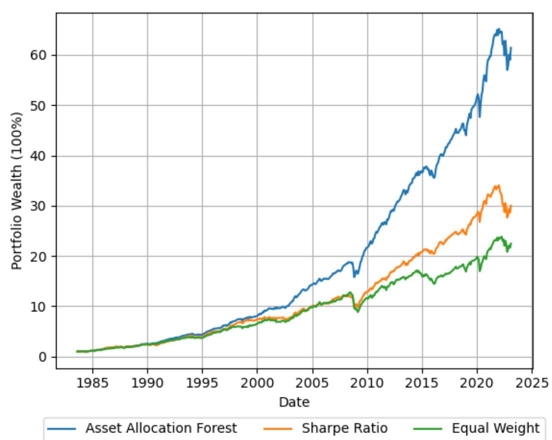
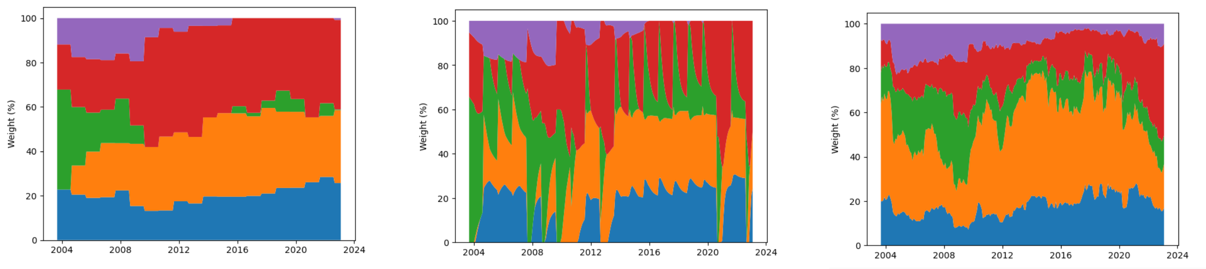


Figure 13: Portfolio Wealth with in-sample estimation.

D.2 Out-of-sample Performance



(a) Sharpe Ratio **(b)** HMM **(c)** Asset Allocation Forest

Figure 14: Portfolio Evolution of out-of-sample estimation with expanding window. Time period: August 2003 - January 2023. Asset classes: Equity, High-Yields, Credits, Bonds, Commodities.

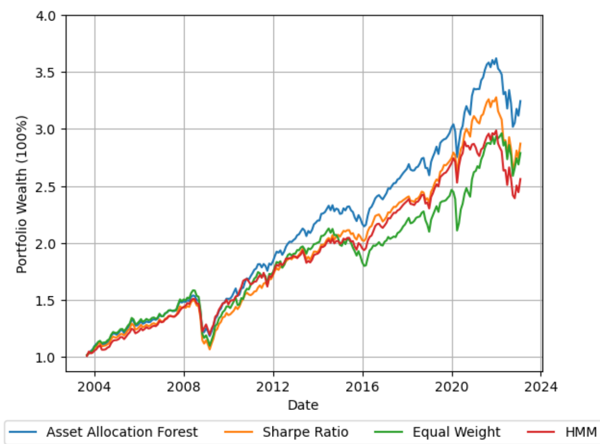
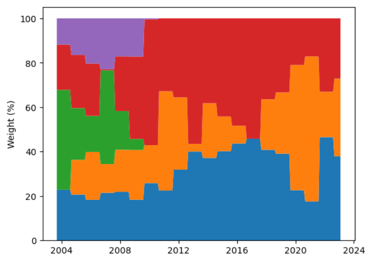
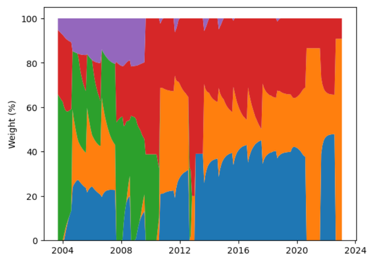


Figure 15: Portfolio Wealth with out-of-sample estimation with expanding window.

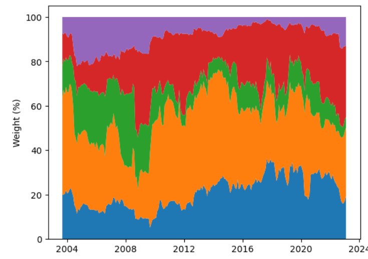
D.3 Out-of-sample Rolling Window



(a) Sharpe Ratio



(b) HMM



(c) Asset Allocation Forest

Figure 16: Portfolio Evolution of out-of-sample estimation with 20-Year rolling window. Time period: August 2003 - January 2023. Asset classes: Equity, High-Yields, Credits, Bonds, Commodities.

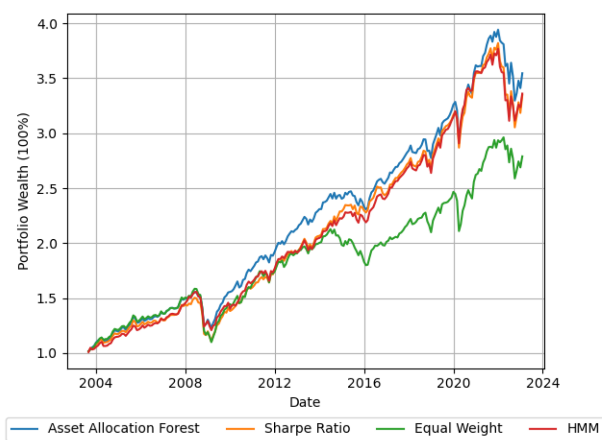
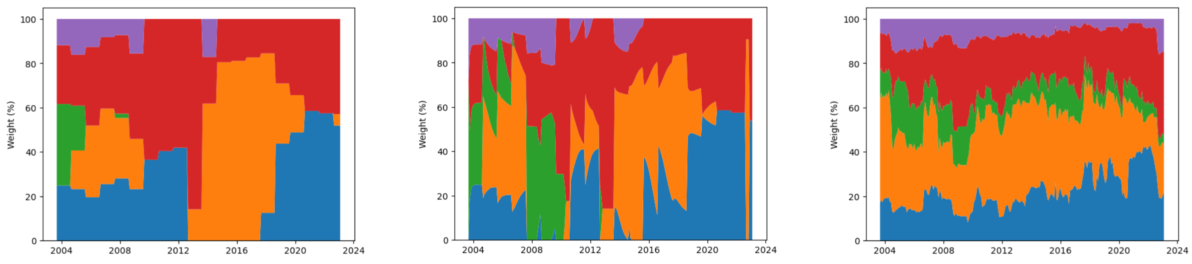


Figure 17: Portfolio Wealth with out-of-sample estimation with 20-Year rolling window.



(a) Sharpe Ratio **(b)** HMM **(c)** Asset Allocation Forest

Figure 18: Portfolio Evolution of out-of-sample estimation with 15-Year rolling window. Time period: August 2003 - January 2023. Asset classes: Equity, High-Yields, Credits, Bonds, Commodities.

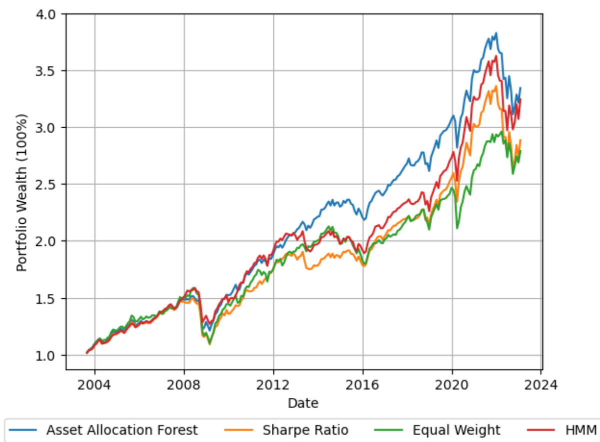
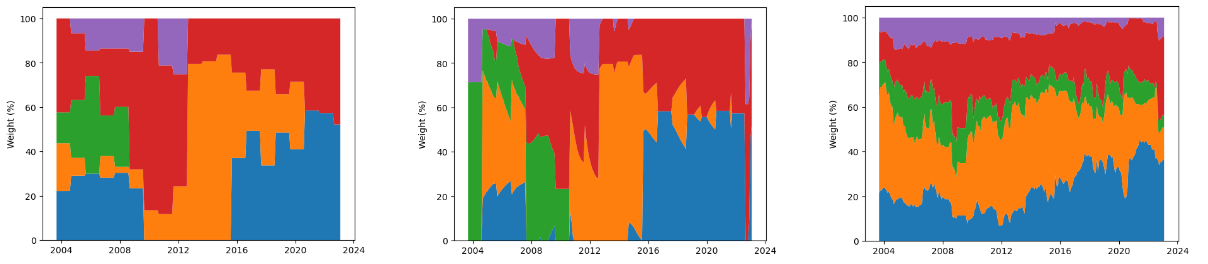


Figure 19: Portfolio Wealth with out-of-sample estimation with 15-Year rolling window.



(a) Sharpe Ratio **(b)** HMM **(c)** Asset Allocation Forest

Figure 20: Portfolio Evolution of out-of-sample estimation with 12-Year rolling window. Time period: August 2003 - January 2023. Asset classes: Equity, High-Yields, Credits, Bonds, Commodities.

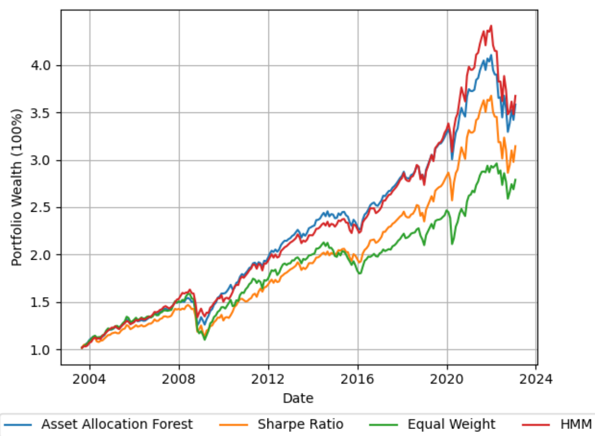
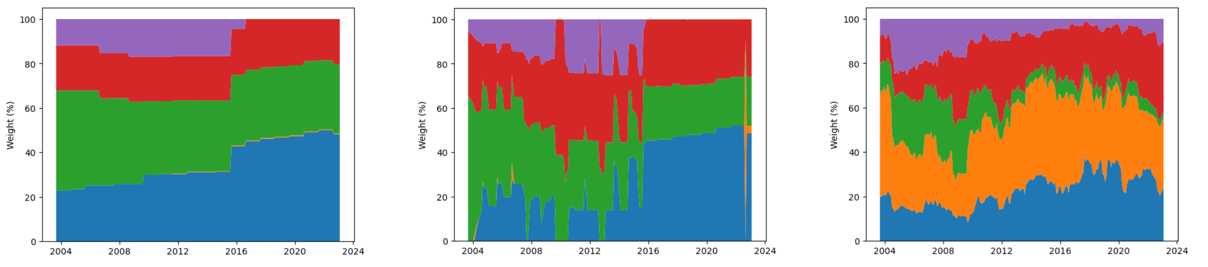


Figure 21: Portfolio Wealth with out-of-sample estimation with 12-Year rolling window.

D.4 Turnover Penalty



(a) Sharpe Ratio **(b)** HMM **(c)** Asset Allocation Forest

Figure 22: Portfolio Evolution when considering $\lambda = 0.002$ and a 20-year rolling window. Time period: August 2003 - January 2023. Asset classes: Equity, High-Yields, Credits, Bonds, Commodities.

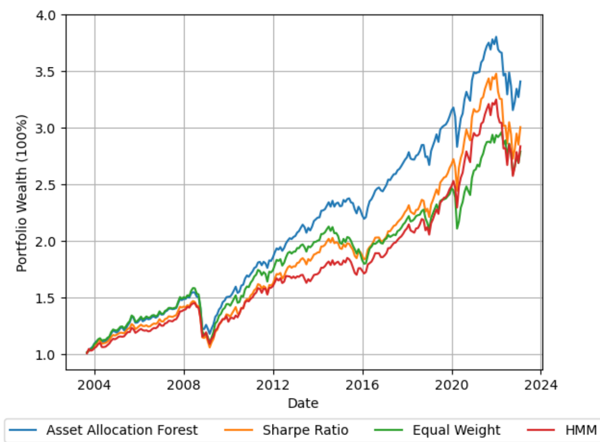
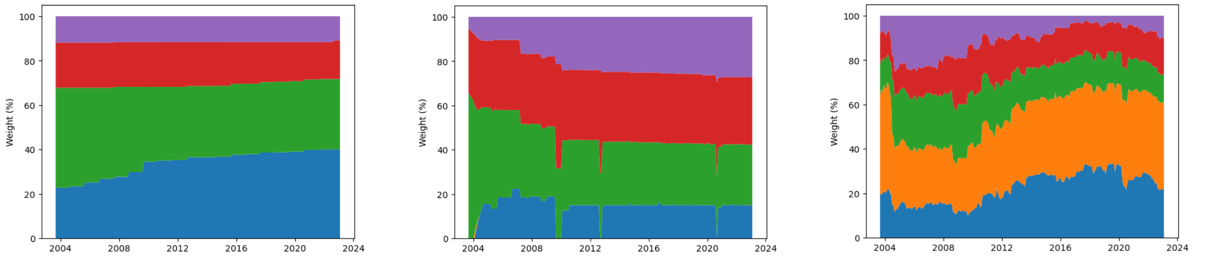


Figure 23: Portfolio Wealth when considering $\lambda = 0.002$ and a 20-year rolling window.



(a) Sharpe Ratio **(b)** HMM **(c)** Asset Allocation Forest

Figure 24: Portfolio Evolution when considering $\lambda = 0.005$ and a 20-year rolling window. Time period: August 2003 - January 2023. Asset classes: Equity, High-Yields, Credits, Bonds, Commodities.

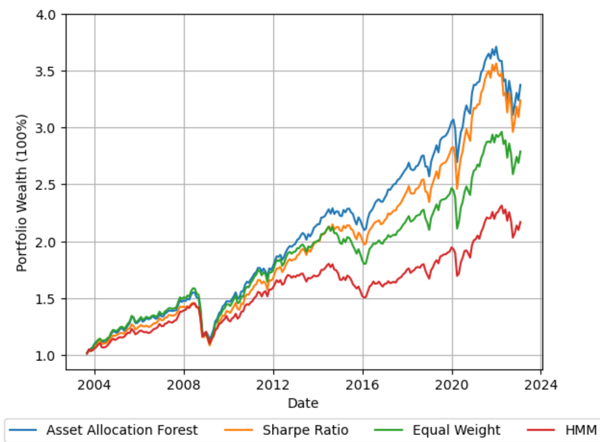


Figure 25: Portfolio Wealth when considering $\lambda = 0.005$ and a 20-year rolling window.

D.5 ALE plots

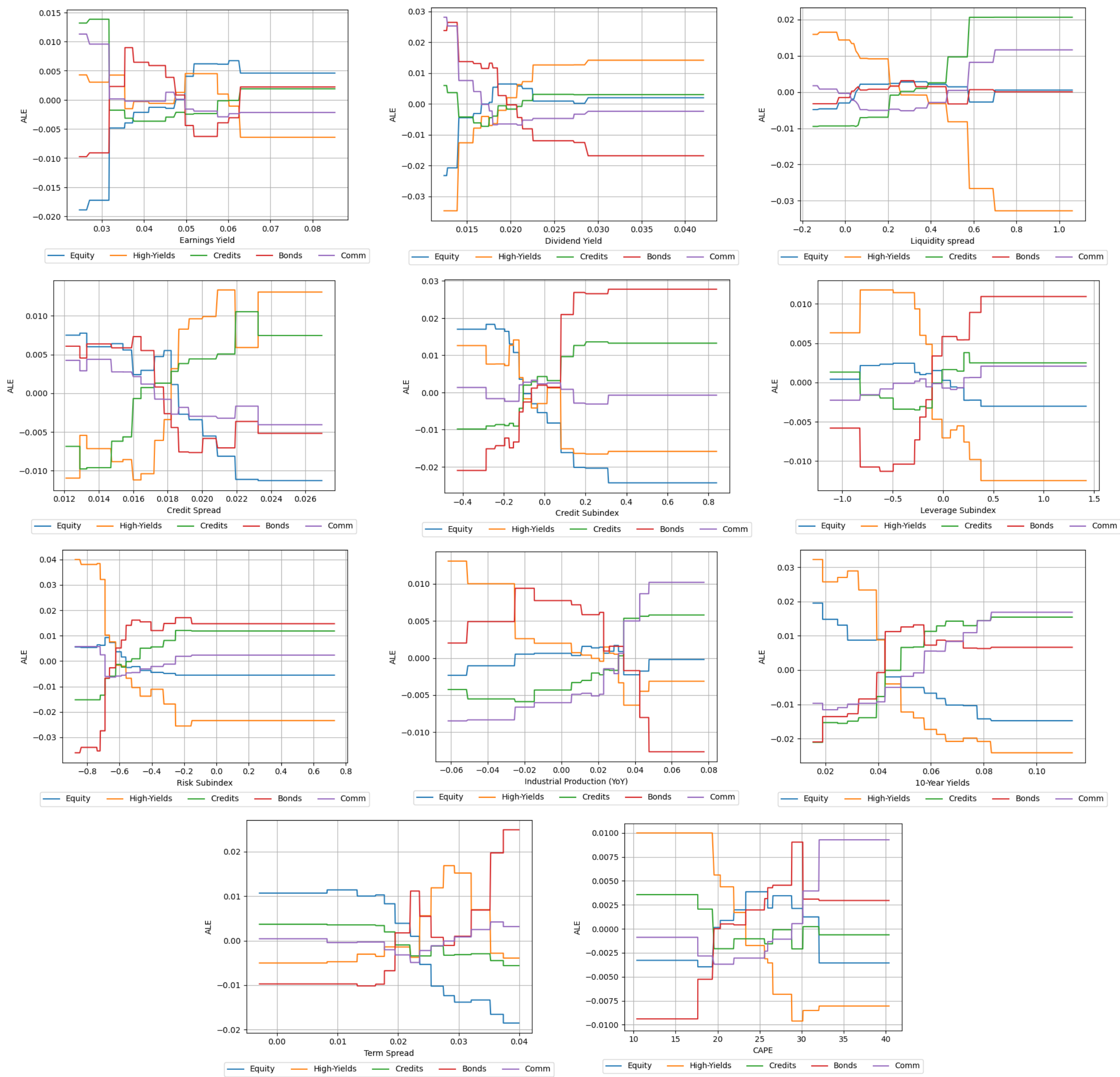


Figure 26: Accumulated Local Effects plot for the following features: Earnings Yield, Dividend Yield, Liquidity spread, Credit Spread, Credit Subindex, Leverage Subindex, Risk Subindex, Industrial Production, Yields 10Y, Term Spread, CAPE