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

January 2023

Abstract

We propose a novel Asset Allocation Forest (AAF) model that combines the well-established machine learning tool with the conventional portfolio optimization method. The determination of locally optimal portfolio weights, which dynamically respond to market conditions, effectively captures market regimes, structural breaks and smooth transitions in a data-driven manner. We illustrate the proposed model using a multi-asset portfolio consisting of equities, bonds, credit, high yield and commodities. The AAF consistently outperforms established benchmarks, including the Hidden Markov Model (HMM), even when trading costs are taken into account. It also opens the door to valuable economic insights. By constructing accumulated local effects (ALE) plots, we find evidence of flight-to-safety, suggesting a strategic shift from riskier assets to less volatile bonds during periods of increased market turbulence. Furthermore, our model shows a pronounced preference for bonds in inflationary periods, demonstrating its adaptability to different economic conditions.

Keywords: interpretable machine learning, portfolio allocation, random forests

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

In recent decades, significant attention of researchers and practitioners has been directed towards determining the optimal allocation of portfolio weights. Despite considerable efforts, a consensus remains elusive regarding the superior performance of any specific approach over the equally weighted portfolio strategy. Notably, while many financial problems possess clearly defined economic objectives, the process of estimating portfolio weights often lacks integration with such objectives. Consequently, the neglect of economic considerations in parameter estimation for portfolio weights may result in substantial losses.

It is widely acknowledged that the volatility, covariances, and expected returns of assets are intricately linked to prevailing market and economic conditions. Diverse assets exhibit varied responses to distinct economic environments, as demonstrated by studies such as [Ilmanen et al., 2014] and [Yang et al., 2009]. In addressing this complexity, we propose a novel Asset Allocation Forest (AAF) approach. This model conceptualizes optimal portfolio weights as nonparametric functions of market conditions, effectively capturing market regimes, structural shifts, and smooth transitions in a data-driven manner. Specifically, the AAF employs a modified version of the random forest method (RF) as the basis for this nonparametric function.

The proposed model is applied to a portfolio encompassing a range of asset classes, including equities, bonds, credit instruments, high yield assets, and commodities. Results demonstrate that the AAF framework adeptly captures the nuanced dynamics and responses of these asset classes, particularly within the context of US markets. Notably, the incorporation of macroeconomic and market variables enables the delineation of distinct economic environments, thereby enhancing the model's efficacy in portfolio management.

The AAF model addresses two predominant challenges for investors. The first challenge arises from the non-linear dependence of asset returns on market and macroeconomic conditions (see e.g. [Flannery and Protopapadakis, 2002], [Boucher and Tokpavi, 2019], [Simonian and Wu, 2019a]) for portfolio construction. The traditional approach to addressing the challenge of non-linearity in asset returns and macroeconomic variables has been to use rolling window estimation or HMMs to model regimes. While HMMs provide a mechanism for navigating through different market regimes, they are not without limitations. In particular, HMMs operate under the assumption of state independence and often face parameter estimation challenges that can limit their applicability and predictive accuracy in certain contexts.

The rolling window approach is also not flexible enough to adapt quickly to changing market conditions. The standard choice of window size corresponds to a time interval of several years. Obviously, market conditions cannot be assumed to be constant over such a long period. As a result, rolling window portfolio optimization is not able to capture rapidly changing market conditions, especially switching regimes. As a result, the AAF uses a tree-based model that dynamically adjusts portfolio weights in response to market conditions, with an optimization objective tied to maximizing the Sharpe ratio. This approach facilitates market-specific optimization, ensuring a tailored response to localized economic scenarios. Our forest approach makes it easy to flexibly model evolving portfolio weights in a data-driven manner and capture the interactions of different regimes with respect to different market conditions. This is in contrast to latent regime-switching models, where only a few regimes are feasible and the number of regimes is difficult to determine.

The second challenge is to mitigate high weight fluctuations, which by their nature induce high trading costs. To overcome this, a turnover penalty is seamlessly integrated into the optimization rule. In addition, the inclusion of bagging in the AAF serves to stabilise the portfolio weights, thereby sensibly managing the associated trading costs.

The proposed adaptation of the random forest method will include portfolio optimization under the classical mean-variance framework of Markowitz [1952]. The benchmarks defined to evaluate the performance of the model are the naive $\frac{1}{N}$ (equal weighting) strategy and the classical SR optimiszation where the parameters are estimated directly from the historical data (global SR optimization) or through an HMM. The analysis is performed both in-sample and out-of-sample, and statistical tests are performed to verify the significance of the results. Transaction costs are also taken into account by measuring the turnover of the strategies, calculating gross and net returns, and incorporating penalties for large allocation changes over time into the model.

The research finds that the AAF leads to higher performance in terms of SR, even when transaction costs are taken into account, and with some degree of statistical significance. Over a 20-year rolling period, the AAF approach demonstrates superior performance compared to other strategies. The out-of-sample net SR for AAF is 0.72, surpassing those of $1/N$ (0.49), and Hidden Markov Model (HMM) strategies (0.66). One advantage found over classical SR maximization is the timely adjustment to rapidly changing market conditions. This can be seen in the weighting of commodities in 2022, where the classical SR completely neglects the asset class based on past performance, while the AAF is able to increase exposure in a period when commodities were the best performing asset class.

We also address the issue of machine learning methods being a 'black box' and show that interpretation is still possible, inferring economic states from the behaviour of different asset classes with respect to the input variables through the use of ALE plots Apley and Zhu [2020]. This method allows the visualisation of the impact of a single variable and assess how the explanatory variables affect the output of the model.

The remainder of this paper is organized as follows: Section 2 presents a literature review, Section 3 introduces our AAF model. Section 4 presents the data and Section 5 demonstrates our empirical results and Section 6 discusses the main findings of the paper and concludes.

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 and market conditions, 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 and 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 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 market conditions 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, defined 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 and 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 et al. [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 and 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 et al. [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 $r_t = (r_{t,1}, \ldots, r_{t,N})^\top$ denote their excess returns with mean μ_t and covariance Σ_t. Then a portfolio with allocation weights $w_t = (w_{t,1}, \ldots, w_{t,N})^\top$ has returns with expectation $\mu_{t,P}(w_t) \stackrel{\text{def}}{=} w_t^\top \mu_t$ and variance $\sigma^2_{t,P}(w_t) \stackrel{\text{def}}{=} w_t^\top \Sigma_t w_t$. 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) $SR_t \stackrel{\text{def}}{=} \frac{w_t^\top \mu_t - rf_t}{\sqrt{w_t^\top \Sigma_t w_t}}$ with rf_t being the riskfree rate. Thus, at each time t, the investor solves the following optimization problem:

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

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:

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

where λ is a penalty for changes in portfolio weights (trading costs) and $\sigma_{t,EW}$ is the volatility of the equally weighted portfolio, $i = 1, \ldots, N$. 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 maximizing 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 maximize 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 et al. [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 et al. [2009a], but contemporary transaction costs are closer to five bps, as noted by De Nard et al. [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 and market, 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 market condition features Carrizosa et al. [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 or market condition 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 and market condition 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 or 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. It is important to note that equation (2) represents merely a singular instance of leaf-specific optimization challenges. Depending on the specific context of the problem under consideration, it may be replaced by a global minimum portfolio problem or any other portfolio optimization approach. 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.

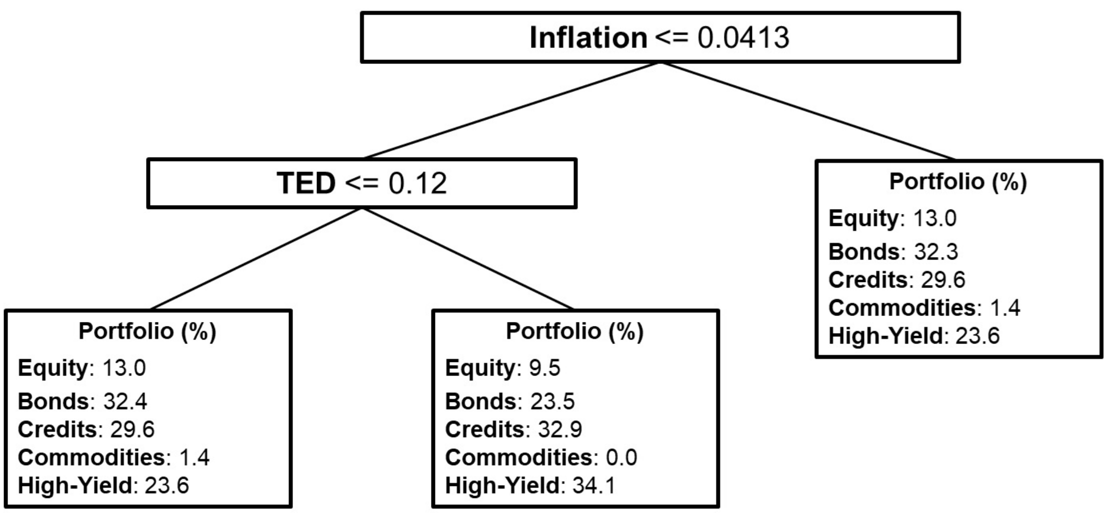


Figure 1: Asset allocation tree and the corresponding portfolio weights using CAPE and credit subindex as the splitting variables. Time period is August 1983–January 2023.

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 re-estimating 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.

There is a set of hyper-parameters to be tuned before constructing a random forest: the number of trees, the maximum depth of each tree or 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. 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.

4 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 A, 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

[1] $https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html$

[2] $https://www.federalreserve.gov/data/nominal-yield-curve.htm$

$t - 1$.

	AR (%)	SD (%)	SR
August 1983 - July 2023			
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 (annualized average of returns (AR), annualized standard deviation of returns (SD), annualized and Sharpe ratio (SR)) of monthly returns of 5 asset classes. Time period is August 1983–January 2023 (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 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

	Equity	High-yields	Credits	Bonds	Commodities
			August 1983 - January 2023		
Equity		0.640	0.343	0.032	0.229
High-yields			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
High-yields			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
High-yields			0.657	−0.026	0.452
Credits				0.606	0.142
Bonds					−0.270

Table 2: Correlations among asset classes. Time period is August 1983–January 2023 (462 monthly returns).

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 A.

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 et al. [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 and Shiller [1998]. Additionally, the CAPE ratio's significance in forecasting asset returns is supported by Borensztein and Ye [2020], Lettau and Ludvigson [2001], and Forsyth and

Mongrut [2022]. 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 term spread's influence on real economic activity and its role as a predictive tool for asset returns is further emphasized by Estrella and Hardouvelis [1991], Campbell and Cochrane [1999], and Justiniano et al. [2014].

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, reflects 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.Gilchrist and Zakrajšek [2012], Nordén and Weber [2009], and Okimoto and Takaoka [2022] have also highlighted the predictive power of credit spreads for economic activity and stock 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 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

Features	Min	Max	Med	Avg	SD
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 (minimum (Min), maximum (Max), median (Med), average (Avg), standard deviation (SD)) of the 14 macroeconomic and market condition variables. Time period is August 1983–January 2023.

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.

Table 4 summarizes the key literature motivating usage of chosen variables related to the business cycle and help forecast asset returns.

Variable	Key Literature
US earnings yield (%)	Fama and French [1992], Wu and Wang [2000], Abraham and Harrington [2018]
US dividend yield (%)	Fama and French [1988], Campbell et al. [2003]
Equity volatility (%)	Schwert [1989], Carlston [2018]
Liquidity spread	Acharya and Pedersen [2005], Franz [2013]
Credit subindex	Simonian and Wu [2019b], Gilchrist and Zakrajšek [2012]
Leverage subindex	Simonian and Wu [2019b], Cooper and Priestley [2012]
Risk subindex	Simonian and Wu [2019b], Tang and Whitelaw [2011]
Non-financial leverage subindex	Simonian and Wu [2019b], Chen et al. [2007]
Credit spread (%)	Gilchrist and Zakrajšek [2012], Nordén and Weber [2009], Okimoto and Takaoka [2022]
Inflation (%)	Kunwar et al. [2023], Li et al. [2017], Devaguptapu and Dash [2021]
IP YoY (%)	Kritzman et al. [2012], Fama [1990]
Yields 10Y	Lettau and Ludvigson [2001], Campbell [1987]
Term spread	Estrella and Hardouvelis [1991], Harvey [1993], Campbell and Cochrane [1999]
CAPE	Campbell and Shiller [1998], Borensztein and Ye [2020], Lettau and Ludvigson [2001], Forsyth and Mongrut [2022]

Table 4: Summary of literature on macroeconomic and market condition variables linked to asset returns.

5 Empirical results

5.1 Out-of-sample performance

We 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 similar until 2015. After this period, the SR strategy clearly outperforms the EW strategy. 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.

Strategy	TO, %	SD, %	Gross Return CW	SR	Max DD, %	BE, %	Net Return CW	SR	Max DD, %	BE, %
AAF	81.51	7.38	3.24	0.68	23.12	1.42	3.14	0.66	23.26	1.22
SR	37.74	7.56	2.87	0.58	28.88	1.39	2.83	0.57	29.00	1.19
EW	26.94	8.51	2.79	0.50	30.63	N/A	2.79	0.50	30.63	N/A
HMM	192.06	7.31	2.56	0.52	20.81	N/A	2.38	0.46	21.09	N/A

Table 5: Out-of-sample performance metrics: cumulative wealth (CW), annualized standard deviation (SD), annualized Sharpe ratio (SR), annualized turnover (TO), maximum drawdown (Max DD), and break-even (BE) for portfolios constructed using AAF, SR, EW, and HMM strategies. Net returns incorporate trading costs $c = 0.002$. Time period is August 2003 - January 2023 using expanding window and $\lambda = 0$.

Analysing Table 5, we can see that the cumulative wealth of the AAF strategy is slightly higher. Moreover, when the risk is taken into account, it is clear that the AAF method fulfills its main purpose by delivering a higher SR due to the lower volatility and higher returns. Figure 2a 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, where credits show a decline in performance and an overall increase 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.

(a) SR **(b)** HMM **(c)** AAF

Figure 2: Evolution of portfolio weights over time for the following strategies: (a) SR strategy, (b) HMM strategy, and (c) AAF strategy. Time period is August 2003–January 2023 using an expanding window and $\lambda = 0$. Asset classes: equity, high-yields, credits, bonds, commodities.

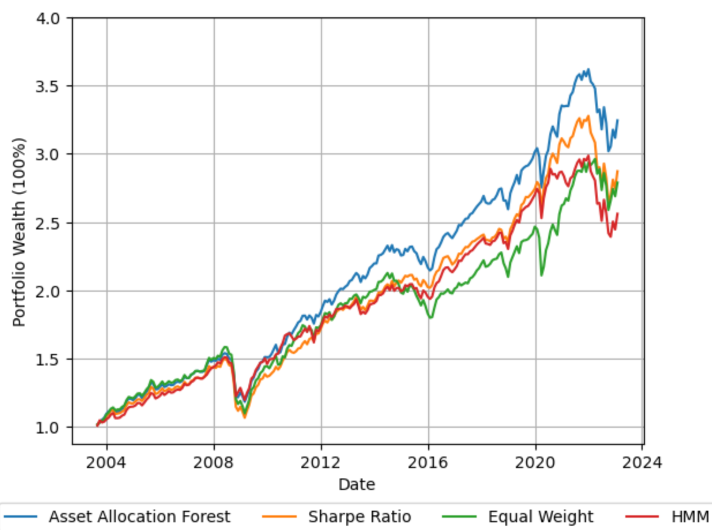


Figure 3: Out-of-sample cumulative portfolio wealth for portfolios constructed using AAF strategy, SR strategy, EW strategy, and HMM strategy. Time period is August 2003–January 2023 using an expanding window and $\lambda = 0$ (229 monthly returns).

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 2b, 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.

Finally, we turn to the proposed methodology of this research, the AAF strategy. Unlike the previous models, the portfolio weights are conditioned on macroeconomic and market condition 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 in terms of returns, 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 2c, 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 and market 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 (0.68). 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.

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 which acts as the implied trading cost. The objective is to find a constant such that a strategy has the same cumulative net return as that of the EW portfolio, and assess the magnitude of such value. 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 the standard

Figure 4: Rolling window (annualized) volatility of the EW portfolio. Time period is July 1995–January 2023.

deviations of the strategies differ from those of the EW strategy, even though we have the volatility matching constraint. 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 4, 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 significantly different using the test proposed by Ledoit and Wolf [2008]. Table 6 shows that the performance of the AAF is indeed significantly better.

Gross Return (SR)				Net Return (SR)			
Strategy	SR	EW	HMM	Strategy	SR	EW	HMM
AAF (0.68)	1.06	2.32*	1.30	AAF (0.66)	0.93	2.14*	1.53
SR (0.58)	-	0.64	0.47	SR (0.57)	-	0.61	0.80
EW (0.50)	-	-	-0.21	EW (0.50)	-	-	0.07
HMM (0.52)	-	-	-	HMM (0.46)	-	-	-

Table 6: Test statistics of the difference between the annualized SR using the test proposed by Ledoit and Wolf [2008] for both gross and net returns with an expanding window and $\lambda = 0$ with significance indicated by stars. Superscript * indicates a 10% significance level, ** indicates a 5% significance level, and *** indicates a 1% significance level.

5.2 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.

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 an 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 wealth after controlling for transaction costs. 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 4. 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.

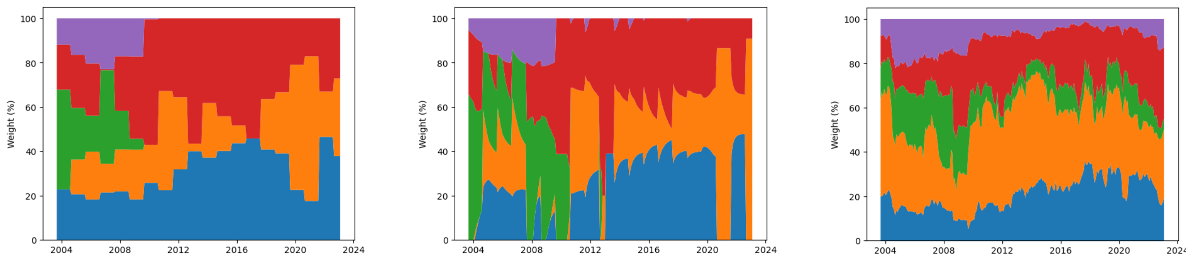
Looking at the evolution of the weights of the SR strategy using a 20-year rolling window shown in Figure 5, we can see that the allocation does not show the same stability as in the case of the expanding window. The sharper changes in the portfolio are reflected in the increase in turnover from 0.38 to 0.64. However, the increase in turnover is offset by the

	TO, %	SD, %	Gross Return CW	SR	Max DD, %	BE, %	Net Return CW	SR	Max DD, %	BE, %
Rolling Window: 12 Years										
AAF	80	7.55	3.58	0.74	19.77	2.42	3.47	0.72	19.92	2.22
SR	64	8.45	3.14	0.58	23.19	1.67	3.07	0.56	23.36	1.47
EW	27	8.51	2.79	0.50	30.63	N.A.	2.79	0.50	30.63	N.A.
HMM	107	8.05	3.68	0.71	21.19	1.76	3.53	0.68	21.52	1.56
Rolling Window: 15 Years										
AAF	88	7.41	3.34	0.70	20.09	1.52	3.23	0.68	20.22	1.32
SR	55	8.49	2.88	0.52	27.14	0.62	2.82	0.51	27.24	0.42
EW	27	8.51	2.79	0.50	30.63	N.A.	2.79	0.50	30.63	N.A.
HMM	144	8.12	3.25	0.62	19.99	0.67	3.07	0.58	20.17	0.47
Rolling Window: 20 Years										
AAF	86	7.40	3.55	0.75	21.39	2.08	3.43	0.72	21.39	1.88
SR	60	8.08	3.35	0.65	26.59	2.80	3.27	0.63	26.71	2.60
EW	27	8.51	2.79	0.50	30.63	N.A.	2.79	0.50	30.63	N.A.
HMM	126	7.51	3.36	0.70	22.43	0.96	3.20	0.66	22.63	0.76

Table 7: Out-of-sample performance metrics (annualized standard deviation (SD), annualized Sharpe ratio (SR), annualized turnover (TO), maximum drawdown (Max DD), cumulative wealth (CW), and break-even (BE)) for portfolios constructed using AAF strategy, SR strategy, EW strategy, and HMM strategy. Time period is August 2003 - January 2023 using a 12-year rolling window, a 15-year rolling window, a 20-year rolling window, and $\lambda = 0$.

increase in return as the break-even point increases relative to the EW strategy, except for the 15-year window case. Figure 5a 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 5b, 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 the credit performed better, which led 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.



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

Figure 5: Evolution of portfolio weights over time for the following strategies: (a) SR strategy, (b) HMM strategy, and (c) AAF strategy. Time period is August 2003–January 2023 using a 20-year rolling window and $\lambda = 0$. Asset classes: equity, high-yields, credits, bonds, commodities.

On the other hand, the AAF with a 20-year rolling window still manages to maintain the smoothness of the weight evolution (Figure 5c), 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 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.

	Gross Return (SR)			Net Return (SR)			
Strategy	SR	EW	HMM	Strategy	SR	EW	HMM
			12-year rolling window				
AAF (0.74)	1.81*	2.04**	0.34	AAF (0.72)	1.75*	1.91*	0.40
SR (0.58)	-	0.48	-1.73*	SR (0.56)	-	0.43	-1.56
EW (0.50)	-	-	-1.21	EW (0.50)	-	-	-1.08
HMM (0.71)	-	-	-	HMM (0.68)	-	-	-
			15-year rolling window				
AAF (0.70)	2.33**	1.92*	0.77	AAF (0.68)	2.19**	1.76*	0.87
SR (0.52)	-	0.15	-1.06	SR (0.51)	-	0.11	-0.82
EW (0.50)	-	-	-0.74	EW (0.50)	-	-	-0.57
HMM (0.62)	-	-	-	HMM (0.58)	-	-	-
			20-year rolling window				
AAF (0.75)	1.48	2.97***	0.45	AAF (0.72)	1.36	2.78***	0.56
SR (0.65)	-	1.15	-0.76	SR (0.63)	-	1.09	-0.50
EW (0.50)	-	-	-1.33	EW (0.50)	-	-	-1.14
HMM (0.70)	-	-	-	HMM (0.66)	-	-	-

Table 8: Test statistics of the difference between the annualized SR of row and column strategies (AAF strategy, SR strategy, EW strategy, and HMM strategy) using the test proposed by Ledoit and Wolf [2008] for gross and net returns. Time period is August 2003 - January 2023 using a 12-year rolling window, a 15-year rolling window, a 20-year rolling window, and $\lambda = 0$. In parentheses, we report the annualized SR of the respective strategy. Superscript * indicates a 10% significance level, ** indicates a 5% significance level, and *** indicates a 1% significance level.

5.3 Turnover penalty

In this section, we aim to test the effect of including the penalty term parameter λ in (2). In is important to mention that the penalty parameter is applied not only to AAF strategy but also to SR strategy and HMM strategy. As the introduction of this condition further increases the computational burden of the estimation, we apply the penalty for the expanding window case and the 20-year rolling window. The results are reported in Table 9 and Table 10 correspondingly. 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 et al. [2009b].

Looking at the results in Table 9 and Table 10, the turnover decreases consistently with λ, as expected since we are now penalizing larger changes in allocation, namely from 0 bps

	TO, %	SD, %	Gross return CW	SR	Max DD, %	BE, %	Net return CW	SR	Max DD, %	BE, %
					$\lambda = 0$ bps					
AAF	81.51	7.38	3.24	0.68	23.12	1.42	3.14	0.66	23.26	1.22
SR	37.74	7.56	2.87	0.58	28.88	1.39	2.83	0.57	29.00	1.19
EW	26.94	8.51	2.79	0.50	30.63	N/A	2.76	0.50	30.69	N/A
HMM	192.06	7.31	2.56	0.52	20.81	N/A	2.38	0.46	21.09	N/A
					$\lambda = 20$ bps					
AAF	59.27	7.79	3.20	0.64	25.36	2.16	3.12	0.62	25.47	1.96
SR	24.74	7.76	2.91	0.57	24.24	N/A	2.88	0.57	24.30	N/A
EW	26.94	8.51	2.79	0.50	30.63	N/A	2.76	0.50	30.69	N/A
HMM	74.91	6.83	2.78	0.62	18.86	N/A	2.70	0.59	19.01	N/A
					$\lambda = 50$ bps					
AAF	39.66	8.24	3.15	0.59	29.02	4.87	3.10	0.58	29.10	4.67
SR	24.98	7.74	3.04	0.61	24.11	N/A	3.01	0.60	24.17	N/A
EW	26.94	8.51	2.79	0.50	30.63	N/A	2.76	0.50	30.69	N/A
HMM	46.12	6.92	2.81	0.62	19.67	0.22	2.76	0.60	19.79	0.02

Table 9: Out-of-sample performance metrics (cumulative wealth (CW), annualized standard deviation (SD), annualized Sharpe ratio (SR), annualized turnover (TO), maximum drawdown (Max DD), and break-even (BE)) for portfolios constructed using AAF, SR, EW, and HMM strategies. Net returns incorporate trading costs $c = 0.002$. Time period is August 2003 - January 2023 using an expanding window and $\lambda = \{0, 0.002, 0.005\}$.

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 6a and Figure 7a.

Regarding the AAF, it is clear that the constraint on portfolio estimation has a negative 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 6c, it is clearly caused by the initial estimation, with credits maintaining a more stable proportion instead of being taken over by bonds and high yields. 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.

Table 11 shows the results of the statistical test on the significance of the SR for gross and

	TO, %	SD, %	Gross return CW	SR	Max DD, %	BE, %	Net return CW	SR	Max DD, %	BE, %
$\lambda = 0$ bps										
AAF	86.26	7.40	3.55	0.75	21.19	2.08	3.43	0.72	21.39	1.88
SR	60.17	8.08	3.35	0.65	26.59	2.80	3.27	0.63	26.71	2.60
EW	26.94	8.51	2.79	0.50	30.63	N/A	2.76	0.49	30.63	N/A
HMM	126.48	7.50	3.36	0.70	22.43	0.96	3.20	0.66	22.63	0.76
$\lambda = 20$ bps										
AAF	63.41	7.67	3.42	0.70	23.73	2.88	3.34	0.68	23.86	2.68
SR	27.08	8.50	3.01	0.54	27.82	93.85	2.97	0.54	27.89	93.65
EW	26.94	8.51	2.79	0.50	30.63	N/A	2.76	0.49	30.63	N/A
HMM	92.40	8.23	2.84	0.52	24.76	0.13	2.74	0.50	24.88	N/A
$\lambda = 50$ bps										
AAF	42.45	8.13	3.38	0.65	27.15	6.30	3.32	0.64	27.25	6.10
SR	25.98	8.28	3.24	0.61	25.81	N/A	3.21	0.60	25.87	N/A
EW	26.94	8.51	2.79	0.50	30.63	N/A	2.76	0.49	30.63	N/A
HMM	41.37	7.65	2.17	0.38	24.38	N/A	2.13	0.37	24.46	N/A

Table 10: Out-of-sample performance metrics (cumulative wealth (CW), annualized standard deviation (SD), annualized Sharpe ratio (SR), annualized turnover (TO), maximum drawdown (Max DD), and break-even (BE)) for portfolios constructed using AAF, SR, EW, and HMM strategies. Net returns incorporate trading costs $c = 0.002$. Time period is August 2003 - January 2023 using a 20 years rolling window and $\lambda = \{0, 0.002, 0.005\}$.

net returns. Although the significance levels vary between 5% and 10%, the AAF strategy is still statistically significantly different from the EW strategy for $\lambda = 0$ and $\lambda = 20$bps.

Gross Return (SR)				Net Return (SR)			
Strategy	SR	EW	HMM	Strategy	SR	EW	HMM
$\lambda = 0$ bps							
AAF (0.68)	1.06	2.32*	1.30	AAF (0.66)	0.93	2.14*	1.53
SR (0.58)	-	0.64	0.47	SR (0.57)	-	0.61	0.80
EW (0.50)	-	-	-0.21	EW (0.50)	-	-	0.07
HMM (0.52)	-	-	-	HMM (0.46)	-	-	-
$\lambda = 20$ bps							
AAF (0.64)	1.00	2.09*	0.02	AAF (0.62)	0.86	1.97*	0.08
SR (0.57)	-	0.86	−0.58	SR (0.57)	-	0.87	−0.42
EW (0.50)	-	-	−0.86	EW (0.50)	-	-	−0.76
HMM (0.62)	-	-	-	HMM (0.59)	-	-	-
$\lambda = 50$ bps							
AAF (0.59)	-0.30	1.47	−0.37	AAF (0.58)	-0.35	1.42	−0.34
SR (0.61)	-	1.51	−0.22	SR (0.60)	-	1.51	−0.16
EW (0.50)	-	-	−0.86	EW (0.50)	-	-	−0.82
HMM (0.62)	-	-	-	HMM (0.60)	-	-	-

Table 11: Test statistics of the difference between the annualized SR using the test proposed by Ledoit and Wolf [2008] for both gross and net returns with an expanding window and $\lambda = \{0, 0.002, 0.005\}$, with significance indicated by stars. Superscript * indicates a 10% significance level, ** indicates a 5% significance level, and *** indicates a 1% significance level.

5.4 Relation of portfolio weights to market 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 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

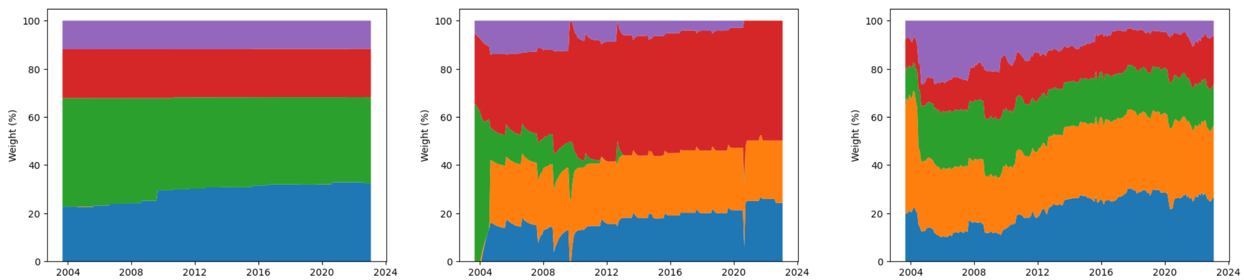
Strategy	Gross Return (SR) SR	EW	HMM	Strategy	Net Return (SR) SR	EW	HMM
			$\lambda = 0$ bps				
AAF (0.75)	1.48	2.97**	0.45	AAF (0.72)	1.36	2.78**	0.56
SR (0.65)	-	1.15	−0.76	SR (0.63)	-	1.08	−0.50
EW (0.50)	-	-	−1.33	EW (0.50)	-	-	−1.14
HMM (0.70)	-	-	-	HMM (0.66)	-	-	-
			$\lambda = 20$ bps				
AAF (0.70)	2.11*	2.47*	1.64	AAF (0.68)	1.98*	2.35*	1.72
SR (0.54)	-	0.41	0.25	SR (0.54)	-	0.41	0.53
EW (0.50)	-	-	−0.21	EW (0.50)	-	-	−0.08
HMM (0.52)	-	-	-	HMM (0.50)	-	-	-
			$\lambda = 50$ bps				
AAF (0.65)	0.57	2.31*	2.09*	AAF (0.64)	0.51	2.24*	2.09*
SR (0.61)	-	1.42	2.35*	SR (0.60)	-	1.42	2.40*
EW (0.50)	-	-	1.23	EW (0.50)	-	-	1.28
HMM (0.38)	-	-	-	HMM (0.37)	-	-	-

Table 12: Test statistics of the difference between the annualized SR using the test proposed by Ledoit and Wolf [2008] for both gross and net returns. Time period is August 2003 - January 2023 using a 20-year rolling window and $\lambda = \{0, 0.002, 0.005\}$. In parentheses, we report the annualized SR of the respective strategy. Superscript * indicates a 10% significance level, ** indicates a 5% significance level, and *** indicates a 1% significance level.

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 (about 8% of the observations) with respect to the characteristic are selected for the plots in order to obtain a clearer graphical representation.[3] The characteristics are divided into 20 buckets, so each bucket contains an average of 24 observations, so all the buckets are represented in the plot even after removing the referenced observations.

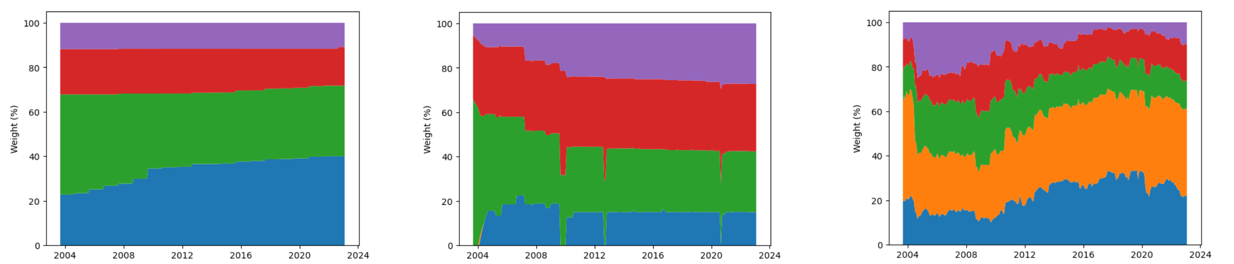
It is important to note that the output of the model always corresponds to a portfolio. Therefore, allocating a higher weight to a particular asset class does not imply a positive performance in that specific context, only that it is preferable to the other asset classes considered.

[3]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.



(a) SR **(b)** HMM **(c)** AAF

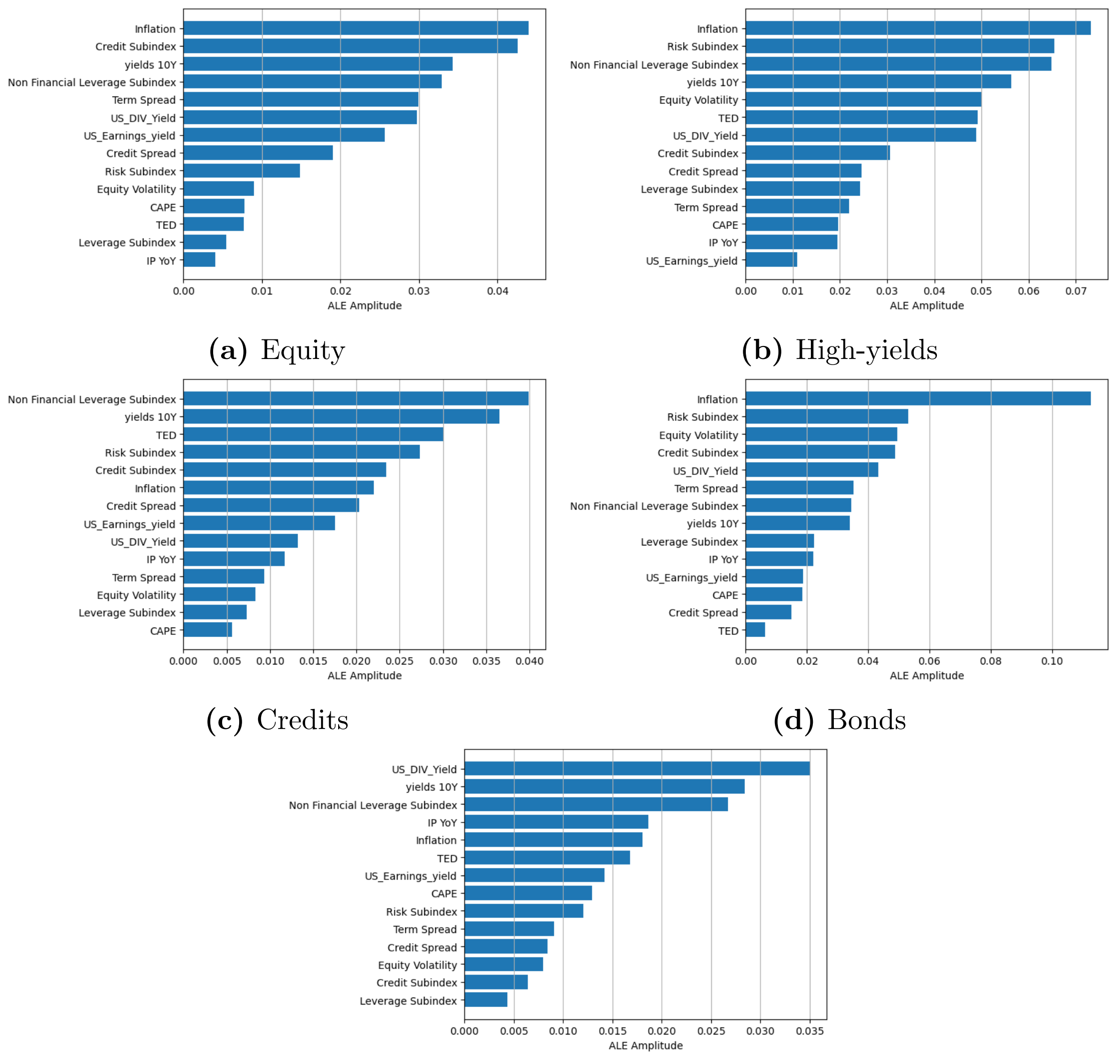
Figure 6: Evolution of portfolio weights over time for the following strategies: (a) SR strategy, (b) HMM strategy, and (c) AAF strategy. Time period is August 2003–January 2023 using an expanding window and $\lambda = 0.005$. Asset classes: Equity, high-yields, credits, bonds, commodities.



(a) SR **(b)** HMM **(c)** AAF

Figure 7: Evolution of portfolio weights over time for the following strategies: (a) SR strategy, (b) HMM strategy, and (c) AAF strategy. Time period is August 2003–January 2023 using a 20-year rolling window and $\lambda = 0.005$. Asset classes: Equity, high-yields, credits, bonds, commodities.

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 8 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.



(a) Equity **(b)** High-yields **(c)** Credits **(d)** Bonds **(e)** Commodities

Figure 8: Amplitude of ALE values for the 14 market condition variables across asset classes. Amplitude is the difference between the highest and lowest ALE values per feature for each asset class. Higher values should indicate a higher variable importance.

Looking at Figure 9, 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. Although the positive impact of inflation on commodities is consistent with Ilmanen et al. [2014], who claims that commod-

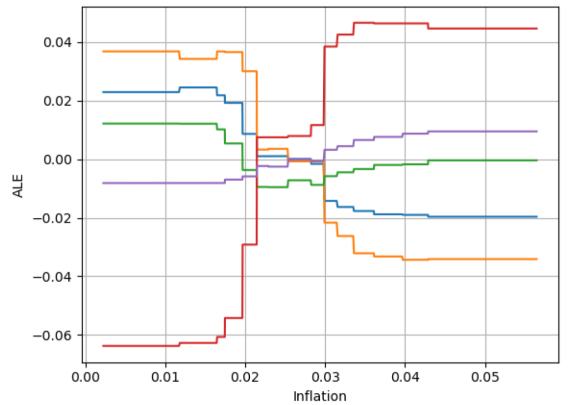


Figure 9: ALE plot for inflation across asset classes. Asset classes: equity, high-yields, credits, bonds, commodities. Time period is August 1983–January 2023.

ities have inflation-hedging properties, the same research shows that government bonds are hit the hardest during high inflation. Furthermore, 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 should also be noted that our model identifies a threshold around the 2% mark, which is considered the general inflation target for central banks.

Assets	Below 2% (145)	Between 2% and 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 13: Annualized SR for different asset classes conditioned on the previous months' inflation. The number of observations in each interval is given in parentheses.

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 13, we can see that the SR of bonds increases at higher inflation intervals, while that of equities and high-yield bonds decreases.

Figure 10 shows the ALE plot for equity volatility, across equities, credit, and commodities showing no clear pattern or pronounced changes along the feature. However, high-yields and

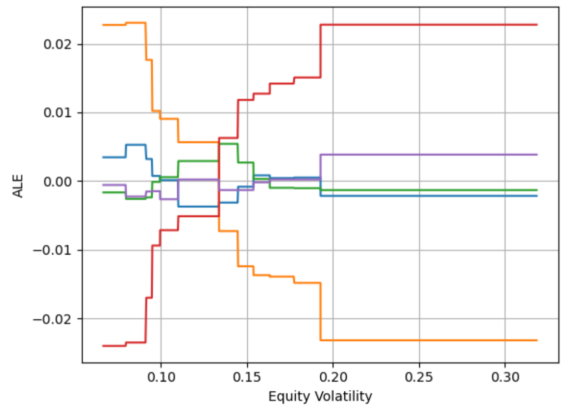


Figure 10: ALE plot for equity volatility across asset classes. Asset classes: equity, high-yields, credits, bonds, commodities. Time period is August 1983–January 2023.

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 the equity markets is a flight toward 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. Similarly to the case of inflation, where the relationship between the feature and the model output does not seem to be linear, 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 14: Annualized SR for different asset classes conditioned on the previous months' equaty volatility. The number of observations in each interval is given in parentheses.

Taking the two extreme ranges implied by the above thresholds, below 9.5% and above 19% volatility, we can see in Table 14 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

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

severe, from 1.69 to 0.43. Moreover, bonds become more desirable from a risk perspective in periods of higher volatility. Table 15 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.

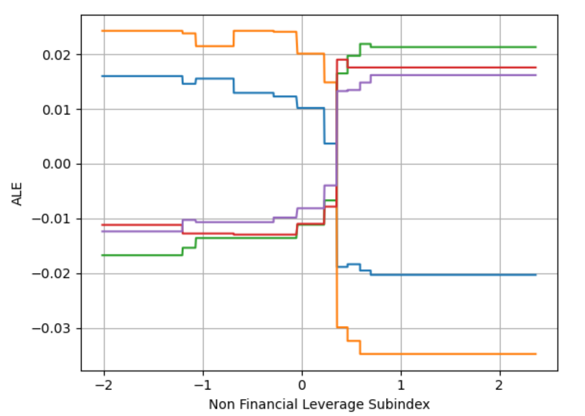


Figure 11: ALE plot for non-financial leverage subindex across asset classes. Asset classes: equity, high-yields, credits, bonds, commodities. Time period is August 1983–January 2023.

A third example of a non-linear relationship captured by the model concerns the Chicago non-financial leverage subindex. As explained by the Chicago FED[4], this subindex is an indicator of the leverage of households and commercial firms, so higher positive values may

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

indicate the risk of future deleveraging and a consequent economic downturn, and thus a switch to less risky assets. In Figure 11 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 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 16 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.1.

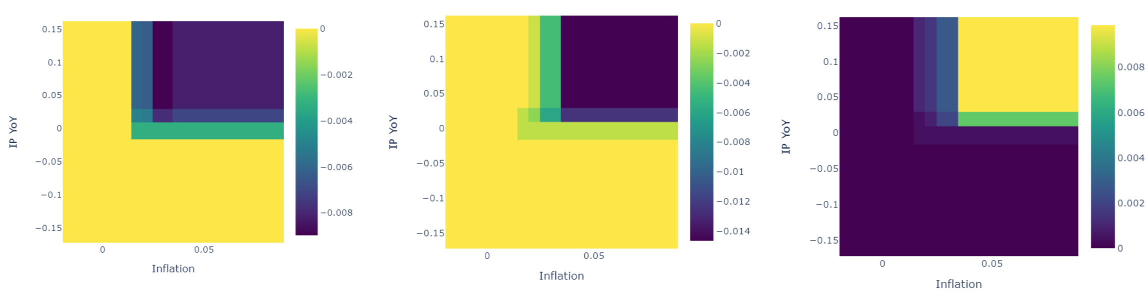
NFL bellow 0.36 (284)				
Equity	Bonds	Credits	Commodities	High-yields
0.92	0.38	0.70	-0.01	1.03
NFL above 0.36 (190)				
Equity	Bonds	Credits	Commodities	High-yields
0.06	0.54	0.64	0.33	0.19

Table 16: Annualized SR for different asset classes conditioned on the previous months' non-financial leverage sub-index. The number of observations in each interval is given in parentheses.

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 12, 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 maximize the SR for a given set of asset classes.



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

Figure 12: ALE plots for interactions between inflation and annual industrial production growth (IP YoY) for equities, bonds and commodities. Time period is August 1983–January 2023.

6 Conclusion

This study introduces the asset allocation forest procedure, which models optimal portfolio weights as a function of market conditions without relying on statistical assumptions. This innovative approach adapts the random forests algorithm from machine learning and applies it to a multi-asset portfolio comprising equities, bonds, credit, high-yields, and commodities. The proposed model captures the dynamics of optimal portfolio weights as nonparametric functions of market conditions.

For a 20-year rolling window, results demonstrate the superiority of the introduced asset allocation forest (AAF) approach with an out-of-sample gross SR of 0.75, outperforming the global SR optimization (SR) strategy (0.65), equal-weighted (EW) strategy (0.50), and hidden Markov model (HMM) strategy (0.70). These results hold after accounting for transaction costs and computing an out-of-sample SR for net returns. In this case, the corresponding SRs are 0.72 for the AAF strategy, 0.63 SR strategy, 0.49 for EW strategy, and 0.66 for HMM strategy. The break-even point for gross returns- relative to the EW portfolio yields an excess value of 208 basis points, indicating profitability after transaction costs. The break-even point for net returns is slightly lower, 188 basis points. Notably, the model exhibits higher turnover (0.86) compared to the EW case (0.27). This is addressed by introducing a regularization term to the change in the weights of two subsequent periods and reduces turnover to 0.42 while maintaining a higher SR (0.65). Robustness is confirmed using various rolling window sizes. Statistical tests verify the significance of the proposed approach's superior performance.

Beyond performance, the study assesses the impact of market conditions on results and reveals insights into asset class dynamics through ALE plots. Expected relationships, such as a flight to safety during higher equity volatility, are confirmed. Surprisingly, the model indicates a preference for bonds during inflationary periods, despite conventional expectations. It might be explained by the model's one-month lag for explanatory variables.

Finally, while the results are promising, it should be noted that they are conditional on the limited universe of asset classes considered. 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.

References

Abraham, R. and C. Harrington (2018): "Is Earnings Yield a Security Return Model Anomaly?" *Theoretical Economics Letters*, 8, 2116–2134.

Acharya, V. V. and L. H. Pedersen (2005): "Asset Pricing with Liquidity Risk," *Journal of Financial Economics*, 77, 375–410.

Apley, D. W. and J. Zhu (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. and G. Imbens (2019): "Machine Learning Methods That Economists Should Know About," *Annual Review of Economics*, 22, 685—725.

Baum, L., T. Petrie, G. W. Soules, and N. Weiss (1970): "A Maximization Techinique Occuring in the Statistical Analysis of Probabilistic Functions of Markov Chains," *The Annuals of Mathematical Statistics*, 41, 164—171.

Borensztein, E. and L. Ye (2020): "Corporate Debt Overhang and Investment in Emerging Economies: Firm-Level Evidence," *International Finance*, 24, 18–39.

Boucher, C. and S. Tokpavi (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, 123—140.

——— (2001): "Random Forests," *Machine Learning*, 45, 5—32.

Campbell, J. Y. (1987): "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18, 373–399.

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

Campbell, J. Y. and J. H. Cochrane (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205–251.

Campbell, J. Y. and R. J. Shiller (1998): "Valuation Ratios and the Long-Run Stock Market Outlook," *The Journal of Portfolio Management*, 24, 11—26.

Carlston, B. (2018): "Can Stock Market Liquidity and Volatility Predict Business Cycles?" *Studies in Economics and Finance*, 35, 81–96.

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

Chen, Roll, and Ross (1986): "Economic forces and the stock market," *Journal of business*, 383–403.

Chen, L., D. A. Lesmond, and J. Z. Wei (2007): "Corporate Yield Spreads and Bond Liquidity," *The Journal of Finance*, 62, 119–149.

Chopra, V. K. and W. T. Ziemba (1993): "The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice," *Journal of Portfolio Management*, 19, 6—11.

Cooper, I. and R. Priestley (2012): "The World Business Cycle and Expected Returns*," *Review of Finance*.

Coulombe, P. G. (2021): "The Macroeconomy as a Random Forest," .

Dacco, R. and S. Satchell (1999): "Why do Regime-switching Models Forecast so Badly?" *Journal of Forecasting*, 18, 1—16.

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

de Prado, M. L. (2022): "Machine Learning for Econometricians: The Readme Manual," *The Journal of Financial Data Science*, 4, 10—30.

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

——— (2009b): "Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?" *The Review of Financial Studies*, 22, 1915—1953.

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

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

Devaguptapu, A. and P. Dash (2021): "Global Commodity Prices and Inflation Expectations," *International Journal of Emerging Markets*, 18, 1053–1077.

Estrella, A. and G. A. Hardouvelis (1991): "The Term Structure as a Predictor of Real Economic Activity," *The Journal of Finance*, 46, 555—576.

Fama, E. F. (1990): "Stock Returns, Expected Returns, and Real Activity," *The Journal of Finance*, 45, 1089–1108.

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

——— (1989): "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economic*, 25, 23—49.

——— (1992): "The Cross-Section of Expected Stock Returns," *The Journal of Finance*, 47, 427–465.

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

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

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

Forsyth, J. and S. Mongrut (2022): "Does Duration of Competitive Advantage Drive Long-Term Returns in the Stock Market?" *Revista Contabilidade & Finanças*, 33, 329–342.

Franz, R. (2013): "Macro-Based Parametric Asset Allocation," .

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

Gilchrist, S. and E. Zakrajšek (2012): "Credit Spreads and Business Cycle Fluctuations," *American Economic Review*, 102, 1692–1720.

Gorton, G. and K. G. Rouwenhorst (2006): "Facts and Fantasies about Commodity Futures," *Financial Analysts Journal*, 62, 47—68.

Gu, S., B. Kelly, and D. Xiu (2020): "Empirical Asset Pricing via Machine Learning," *The Review of Financial Studies*, 33, 2223—2273.

Guidolin, M. and A. Timmermannc (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, 59—66.

Harvey, C. R. (1993): "Term Structure Forecasts Economic Growth," *Financial Analysts Journal*, 49, 6—8.

Ilmanen, A., T. Maloney, and A. Ross (2014): "Exploring Macroeconomic Sensitivities: How Investments Respond to Different Economic Environments," *Journal of Portfolio Management*, 40, 87—99.

Jagannathan, R. and T. Ma (2003): "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps," *The Journal of Finance*, 58, 1651—1683.

Justiniano, A., G. E. Primiceri, and A. Tambalotti (2014): "The Effects of the Saving and Banking Glut on the U.S. Economy," *Journal of International Economics*, 92, S52–S67.

Kritzman, M., S. Page, and D. Turkington (2012): "Regime Shifts: Implications for Dynamic Strategies," *Financial Analysts Journal*, 68, 22—39.

Kunwar, S., S. Timilsina, S. Nepal, P. Shrestha, A. Shrestha, and M. Gahatraj (2023): "Impact of Money Supply and Inflation on Economic Growth in Nepal," *The Lumbini Journal of Business and Economics*, 11, 381–396.

Ledoit, O. and M. Wolf (2008): "Robust performance hypothesis testing with the Sharpe ratio," *Journal of Empirical Finance*, 15, 850—859.

——— (2011): "Robust Performances Hypothesis Testing with the Variance," *Wilmott*, 55, 86—89.

Lettau, M. and S. Ludvigson (2001): "Consumption, Aggregate Wealth, and Expected Stock Returns," *The Journal of Finance*, 56, 815–849.

Li, J., J.-P. Chavas, X. Etienne, and C. Li (2017): "Commodity Price Bubbles and Macroeconomics: Evidence from the Chinese Agricultural Markets," *Agricultural Economics*, 48, 755–768.

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

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

Nordén, L. and M. Weber (2009): "The Co-movement of Credit Default Swap, Bond and Stock Markets: An Empirical Analysis," *European Financial Management*, 15, 529–562.

Okimoto, T. and S. Takaoka (2022): "The Credit Spread Curve Distribution and Economic Fluctuations in Japan," *Journal of International Money and Finance*, 122, 102582.

Samitas, A. and A. Armenatzoglou (2014): "Regression tree model versus Markov regime switching: a comparison for electricity spot price modelling and forecasting," *Operational Research*, 14, 319–340.

Schwert, G. W. (1989): "Why Does Stock Market Volatility Change Over Time?" *The Journal of Finance*, 44, 1115–1153.

Sharpe, W. (1964): "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *The Journal of Finance*, 19, 425—442.

Simonian, J. and C. Wu (2019a): "Minsky vs. Machine: New Foundations for Quant-Macro Investing," *The Journal of Financial Data Science*.

——— (2019b): "Minsky vs. Machine: New Foundations for Quant-Macro Investing," *The Journal of Financial Data Science*, 1, 94—110.

Tang, Y. and R. F. Whitelaw (2011): "Time-Varying Sharpe Ratios and Market Timing," *Quarterly Journal of Finance*.

Uysal, A. S. and J. M. Mulvey (2021): "A Machine Learning Approach in Regime-Switching Risk Parity Portfolios," *Journal of Financial Data Science*, 3, 87—108.

Wu, DiCiurcio, Yeo, and Wang (2022): "Forecasting US Equity and Bond Correlation—A Machine Learning Approach," *The Journal of Financial Data Science*, 4, 76—86.

Wu, C. and X. Wang (2000): "The Predictive Ability of Dividend and Earnings Yields for Long-Term Stock Returns," *Financial Review*.

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

Advancing Markowitz: Asset Allocation Forest

– Supplementary Appendix –

A 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 1: 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 2: Macroeconomic and market features sources.

B Methodology

B.1 Hidden Markov model

An HMM assumes that a variable $r_t = (r_{t,1}, \ldots, r_{t,N})^\top$ follows a distribution whose parameters depend on the regime at time t ($S_t = i$). In our case, r_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 ¡ 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: $Z_{i,j} : z_{i-1,j} < Z_j \leq z_{i,j}$, 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.

- *Annual Turnover (TO)* 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]:

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

  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.

  We calculate the following four performance measures for both, gross and net, returns series. Gross returns are raw returns that do not take transactions costs into account and net returns with subtracted fees. When the portfolio is rebalanced at time $t+1$, it gives rise to a trade in each asset of magnitude $|w_{i,t+1} - w_{i,t^+}|$. Thus, to calculate the net return we reduce the return by the cost of such a trade over all assets, given by

  $$tc_{t+1} = c\sum_{i=1}^{N}|w_{i,t+1} - w_{i,t^+}|,$$

- *Maximum Drawdown (Max DD)*

  The *Max DD* is the highest loss relative to the highest point (up until that moment) throughout the lifetime of the portfolio:

  $$\text{Max DD} = \max_{0\leq t_1<t_2\leq T}\left(1 - \frac{r^p_{t_2}}{r^p_{t_1}}\right),$$

  where r^p_t denotes the cumulative return of the strategy at month t. t_1 andt_2 are indices representing time periods, with t_1 being the time at a peak and t_2 the time at a trough following t_1.

- *Cumulative wealth (CW)* CW generated by each benchmark strategy with initial investment $W_0 = 1$ is computed as follows for gross returns series:

$$W_{t+1} = W_t(1 + w_t^\top X_{t+1}).$$

CW for net returns series:

$$W_{t+1} = W_t(1 + w_t^\top X_{t+1})(1 - tc_{t+1}).$$

- *Sharpe ratio (SR)*

We compute SR of out-of-sample monthly returns series.

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[5].

D Empirical results

D.1 ALE plots

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

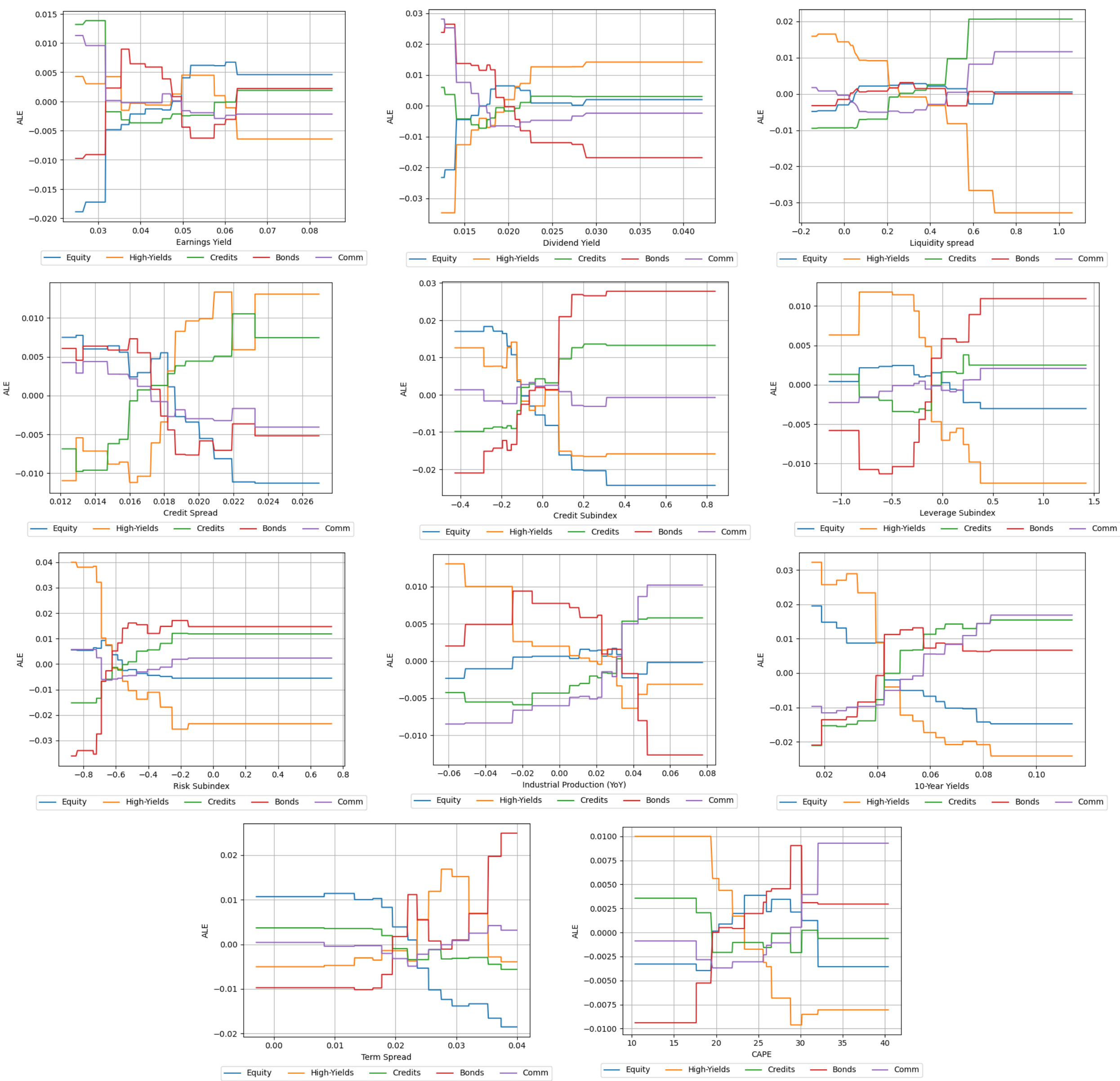


Figure D.4.1: ALE plots for earnings yield, dividend yield, liquidity spread, credit spread, credit subindex, leverage subindex, risk subindex, industrial production, yields 10Y, term spread, and CAPE across asset classes.