Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis - Quantitative Finance

Dynamic Economic Tracking Portfolios: Leveraging Random and Local Linear Forests for Macroeconomic Forecasts

Maurits van Altvorst (557923)

Supervisor:	Dr. A. Tetereva
Second assessor:	M. Mueller
Date:	1st July 2024

Abstract

This study advances the construction of economic tracking portfolios (ETPs) by integrating random forests (RFs) and local linear forests (LLFs) to capture complex, non-linear relationships between asset returns and macroeconomic factors. We develop RF and LLF-based ETPs to track inflation, consumption growth, and industrial production growth over short (1-month) and long-term (1-year) horizons. Our findings demonstrate that machine learning-based approaches consistently outperform linear ETPs. For the 1-year horizon, LLF ETPs achieve statistically significant improvements in tracking inflation and consumption growth, with Mincer-Zarnowitz R^2 values increasing from 4.4% to 6.9% for inflation and from 3.1% to 7.4% for consumption growth. Shapley value analysis reveals that relationships between assets and macroeconomic factors vary with economic conditions, challenging static factor decompositions. Kernel principal component analysis of LLF kernels identifies distinct economic regimes, offering a new approach to understanding changing economic relationships. While machine learning-based ETPs show improved performance, they exhibit higher turnover and sensitivity to uninformative assets, presenting practical implementation challenges. These advancements provide improved tools for economic forecasting, risk management, and deeper insights into the dynamic relationships between financial markets and macroeconomic conditions.

1 Introduction

The complex relationships between financial markets and macroeconomic variables have long been a subject of study in economics. Understanding how asset returns reflect and predict changes in key economic indicators is crucial for investors, policymakers, and economists alike. The relationships between markets and the economy forms the foundation of many economic theories and models, from the Efficient Market Hypothesis to modern macro-finance models.

Economic tracking portfolios (ETPs), first introduced by Breeden et al. (1989) and later expanded by Lamont (2001), have shown to be powerful tools in capturing these complex market-economy dynamics. These portfolios are designed to track specific macroeconomic variables, offering insights into market expectations about future economic conditions. They provide a means to extract real-time economic forecasts from financial markets, potentially offering more timely information than traditional economic factors which are often released with a lag.

Timely and accurate economic forecasts play a critical role in our modern economy. Central banks rely on macroeconomic indicators to guide monetary policy decisions. Governments need accurate economic projections to design fiscal policies and budget plans. Investors and businesses require reliable economic forecasts to make informed decisions about capital allocation and long-term strategies. In this context, ETPs serve as a bridge between fast-moving financial markets and broader macroeconomic conditions.

However, the current framework for constructing ETPs, as introduced by Lamont (2001), relies on linear regression models to link asset returns with future values of macroeconomic indicators. While this approach has proven effective, it operates under the assumption of static, linear relationships between financial markets and economic variables. This assumption may be overly simplistic given the complex nature of modern economies.

The global economy is characterised by non-linear interaction effects, feedback loops, and regime shifts. Economic relationships that hold during periods of economic stability might change or reverse during crises or other economic shifts. For example, the relationship between inflation and unemployment, as described by the Phillips curve, has shown significant instability over time (Gordon, 2011). Similarly, the relationships between financial markets and macroeconomic conditions can also potentially change depending on the economic environment.

The limitations of linear models in capturing these effects can lead to suboptimal portfolio construction and less accurate economic forecasts. This, in turn, could lead to inefficient resource allocation or misguided monetary or fiscal policy decisions, potentially leading to economic instability.

Given these challenges, we propose to enhance the construction of ETPs with the use of machine learning techniques, specifically random forests (RFs) and local linear forests (LLFs). These methods are capable of capturing non-linear relationships and interactions between variables, potentially enhancing the performance of ETPs.

This leads to our central research question:

Can we improve the forecasting accuracy of economic tracking portfolios introduced by Lamont (2001) by utilising random forest techniques, and what new insights can these enhanced portfolios provide about the relationships between financial markets and macroeconomic variables?

By integrating these machine learning techniques into the construction of ETPs, we aim to develop more accurate tools for economic forecasting and risk management. Improved ETPs could help central banks by providing better real-time estimates of inflation expectations, leading to more effective inflation targeting. For investors and asset managers, enhanced consumption growth or industrial production tracking portfolios could lead to more informed investment decisions and more efficient capital allocation. Moreover, more accurate ETPs could be used to hedge existing trading strategies against macroeconomic risks, improving financial stability.

From a broader economic perspective, businesses could use improved ETPs to make more accurate demand forecasts, helping with production, inventory, and investment decisions. Policymakers could leverage better forecasts for better fiscal policies, potentially leading to more effective countercyclical measures and smoother business cycles.

This paper makes several key contributions to the existing literature. First, it introduces two new approaches for constructing ETPs that can capture non-linear relationships and interaction effects between macroeconomic variables and asset returns. Second, by employing Shapley values, we provide insights into how different economic variables contribute to tracking portfolio allocations across different economic regimes. This analysis illustrates the changing relationships between economic variables and asset returns. Thirdly, we use kernel principal component analysis (KPCA) on the LLF kernel to identify distinct economic regimes based on the similarity of historical economic environments. This technique provides a data-driven method for identifying economic regimes, which could be valuable for economic research, policy analysis, or even trading strategies. These advancements not only enhance the accuracy of economic forecasting tools, but also deepen our understanding of the changing relationships between financial markets and the broader economy.

Moreover, by comparing the performance of RF and LLF-based ETPs with traditional linear models across different time horizons and economic variables, we offer an evaluation of the benefits and limitations of these techniques in economic forecasting and portfolio construction. This provides insights into when and where these machine learning-based techniques offer the most significant improvements over traditional methods.

The rest of this paper is organised as follows. We begin with an overview of the existing literature, highlighting the relevance of machine learning-based tracking portfolios within the broader context of previous research. Next, we describe the data used in this study, followed by a detailed explanation of the methodology. This includes the construction of tracking portfolios using linear methods as described by Lamont (2001) and its extensions based on RFs and LLFs. We describe Shapley values and kernel PCA and how they can help with the interpretation of RF-based ETPs. We then present the results, focusing on the performance improvements and on the insights gained through Shapley values and KPCA. Finally, we conclude by discussing the implications of our findings and identifying potential areas for future research.

2 Related Literature

This study intersects several areas of economic and financial literature, including ETPs, applications of machine learning in finance, and macroeconomic forecasting. We review these areas to give better context and to highlight the potential contributions of this research.

2.1 Economic Tracking Portfolios and Asset-Macroeconomic Factor Relationships

The relationship between financial markets and macroeconomic variables has long been a subject of study in economics. Early work by Fama and Schwert (1977) explored the relationship between stock returns and inflation, laying the groundwork for more sophisticated approaches to linking financial markets and macroeconomic variables.

Economic tracking portfolios have played a crucial role in understanding and managing macroeconomic risk, bridging the gap between financial markets and the economic environment. The work by Breeden et al. (1989) initiated the development by introducing macroeconomic tracking portfolios (called maximum correlation portfolios in his paper) to follow consumption patterns with the purpose of testing the Consumption Capital Asset Pricing Model (CCAPM). These portfolios were intended to track specifically consumption.

Expanding on this concept, Lamont (2001) proposed a framework for constructing tracking portfolios aimed at economic variables beyond consumption. Lamont's key insight was that asset returns contain information about future discount rates and cash flows, making them predictive of different economic indicators. By focusing on unexpected returns rather than total returns, Lamont's approach allowed for a broader application of tracking portfolios across various macroeconomic factors, including inflation, consumption growth, and industrial production growth.

Lamont's methodology involved linear regression to extract coefficients that link asset returns to future values of macroeconomic factors. This linear approach, while effective, has limitations in capturing complex, non-linear relationships often present in economic and financial data. Our research aims to address these limitations by incorporating machine learning techniques, specifically random forests, to enhance the tracking performance of these portfolios.

Subsequent research has built upon Lamont's work, expanding the application of ETPs. For instance, Vassalou (2003) used tracking portfolios to investigate the relationship between stock returns and GDP growth, finding that news related to future GDP growth is a significant risk factor in explaining the cross-section of equity returns. This shows the value of ETPs in revealing economic relationships, yet they largely follow the linear framework introduced by Lamont.

Recent research has further clarified relationships between specific asset classes and macroeconomic factors. Notably, Lohre et al. (2020) found that credits have mostly exposure to economic growth, while commodities predominantly have exposure to inflation. This aligns with our findings to some extent, but our research reveals more nuanced patterns.

2.2 Machine Learning in Financial Applications

The widespread presence of non-linear relationships in economics and finance has been well-documented in the literature, highlighting the limitations of linear models in capturing complex economic relationships. Koop and Potter (1999) demonstrated the widespread presence of non-linear relationships in macroeconomic time series, providing motivation for more flexible modelling approaches.

Machine learning techniques have become increasingly prevalent in finance, offering new ways to capture complex, non-linear relationships in financial data. Random forests (RFs), introduced by Breiman (2001), have proven particularly effective in handling non-linear relationships and interactions between covariates. Based on the concepts of Classification and Regression Trees (CART) (Breiman et al., 1984), bagging (Breiman, 1996), and randomised trees (Amit & Geman, 1997), RFs combine multiple decision trees to improve predictive accuracy while mitigating overfitting.

Within the area of asset pricing, Gu et al. (2020) conducted a study comparing various machine learning methods, including RFs, in predicting cross-sectional stock returns. Their findings demonstrate the effectiveness of tree-based methods in capturing non-linear and interaction effects between covariates, outperforming traditional linear factor models.

The application of RFs in risk management has also shown promise. Zhu et al. (2019) employed RFs to estimate default probabilities for credit risk, achieving more accurate predictions compared to a more traditional logistic regression approach.

In the field of portfolio optimisation, Ban et al. (2018) explored how machine learning techniques, including RFs, can enhance portfolio performance by taking advantage of interactions and non-linear relationships among covariates. Their results show that RF-based portfolios achieve higher Sharpe ratios compared to conventional mean-variance optimised portfolios.

The application of machine learning to macroeconomic forecasting, particularly inflation, has also shown promising results. Medeiros et al. (2021) demonstrated the effectiveness of RFs in inflation forecasting, consistently outperforming more traditional benchmarks. This superior performance is attributed to the model's variable selection mechanism and its ability to capture non-linear relationships between macroeconomic covariates and inflation. Building on this, our tracking portfolio-based approach incorporates RFs with ETPs, offering the advantage of quicker responses to exogenous shocks. Since asset returns update more frequently than macroeconomic data, our model can potentially provide more timely forecasts.

Finally, Asset-Pricing trees (AP-trees) provide a flexible approach to explaining cross-sectional returns in high-dimensional scenarios by utilising decision trees to group similar stocks (Bryzgalova et al., 2020). They create optimal portfolio splits to span the stochastic discount factor. AP-trees have shown high out-of-sample Sharpe ratios compared to more traditional methods. Extending this research, Cong et al. (2023) present a new approach to asset pricing with panel trees under global split criteria. They propose a method that incorporates the stochastic discount factor to effectively capture cross-sectional variations in asset returns. This approach uses decision trees to optimise splits based on a global criterion, enhancing the explanatory power of the model and providing a better understanding of the factors influencing asset prices.

2.3 Integration of Random Forests and Local Linear Forests with Economic Tracking Portfolios

Building on previous literature, this study proposes a new approach by integrating RFs with ETPs. While Bettencourt et al. (2024) introduced the Asset Allocation Forest for maximising Sharpe ratios, this study adapts this framework specifically for minimising tracking error in ETPs.

Our RF-based ETPs employ decision trees that partition historical data points based on macroeconomic indicators, optimising for minimal tracking error rather than traditional RF impurity measures like those used in CART (Breiman et al., 1984). This approach can capture complex, non-linear relationships between asset returns and macroeconomic factors that linear models may miss. Each leaf in our RF ETP contains an optimised portfolio conditional on the corresponding macroeconomic conditions, with the final portfolio weights determined by averaging across multiple trees.

To further enhance performance, this study employs local linear forests (LLFs), an extension to RFs designed to capture both global non-linear structures as well as local linear relationships. LLFs, as introduced by Friedberg et al. (2020), use the structure of the underlying RF to generate an adaptive kernel for local linear regression, combining the strengths of both RFs and local linear regression.

LLFs build upon generalised random forests (GRFs), as introduced by Athey et al. (2019), who developed a unified framework for estimating different types of statistical models. GRFs are designed to handle nonparametric estimation of conditional average treatment effects, quantile regression, instrumental variable regression, and other models that account for heterogeneity in data. A central concept in GRFs is 'honesty' (Wager & Athey, 2018), which involves splitting the training data into disjoint subsets for tree construction and leaf estimation. This approach helps avoid overfitting and is crucial for achieving theoretical consistency results. Although the literature shows considerable interest, no consistency results are available for random forests that use fully grown trees without the application of honesty. GRFs also employ the adaptive kernel interpretation of RFs, as advocated before by Meinshausen and Ridgeway (2006) and Hothorn et al. (2004).

To employ LLFs in the context of economic tracking portfolios, this study adapts the LLF framework to fit the specific needs of portfolio construction and optimisation. Instead of using traditional local linear regression, we implement a customised local linear regression approach that allows us to extract optimal portfolio weights from the coefficients. Our methodology also employs a different splitting rule optimised for minimising tracking error rather than the measures of impurity typically used in classification and regression tasks.

By integrating LLFs as opposed to RFs, we can leverage the local linear relationships between macroeconomic variables to construct portfolios that minimise tracking error. The use of honesty in the data splitting process helps with avoiding overfitting, critical in financial contexts where data is often noisy.

2.4 Non-linear Relationships and Regime Dependence in Asset-Macroeconomic Factor Interactions

Our research provides new insights into the regime-dependent nature of relationships between asset returns and macroeconomic variables. This builds upon earlier work by Vassalou (2003), who used tracking portfolios to investigate the relationship between stock returns and GDP growth, finding that stocks track expectations of GDP growth. Our findings, through the use of Shapley values, reveal more complex dynamics.

For instance, we find that the relationship between commodities and inflation is not static, as suggested by Lohre et al. (2020). Instead, it varies with the economic environment. Our results show that as consumption growth increases, the importance of commodities relative to credits in tracking inflation rises, with an interaction effect with the average S&P 500 earnings yield.

Furthermore, our analysis of the LLF kernel using kernel PCA reveals distinct economic regimes, providing a data-driven method for regime identification. This contributes to the literature on economic regime shifts, such as the work by Hamilton (2016) on regime-switching models based on the macroeconomic environment. Our approach offers a new perspective by constructing regimes based on changes in the relationships between asset returns and macroeconomic variables.

2.5 Interpretability and Economic Insights from Random Forest Models

A common criticism of machine learning models in economics is their "black box" nature. By employing Shapley values, we interpret the behaviour of the RF and LLF ETPs to provide economic insights into the drivers of portfolio allocation decisions. This aligns with recent efforts to make machine learning methods more economically meaningful (Mullainathan & Spiess, 2017).

Our Shapley value analysis reveals, for instance, that year-over-year inflation has the most substantial effect on the leverage of high-yields, equities, and credits in consumption growth tracking portfolios. This finding provides a nuanced view of how inflation affects the relationships between asset returns and consumption growth, contributing to the literature on the relationship between inflation and consumption (Bekaert & Engstrom, 2010).

3 Data

Our study examines ETPs comprising five distinct asset classes: equities, high-yield instruments, credits, bonds, and commodities. We analyse monthly returns from August 1983 to January 2023, aligning with the dataset used by Bettencourt et al. (2024) in their study on optimising asset allocation forests for maximising Sharpe ratios. This timeframe is determined by the availability of reliable high-yield return data, which begins in August 1983. Data is sourced from various financial indices and databases, as described in Appendix A. This section also contains descriptive statistics for the macroeconomic covariates. 10-year zero-coupon yields are converted to monthly returns, as described by Bettencourt et al. (2024)

$$R_t^{Bond} = \frac{\exp\left(-(9 + 11/12) \cdot \text{yield}_t\right)}{\exp\left(-10 \cdot \text{yield}_{t-1}\right)}.$$

For robustness checks, we also consider crude oil and gold as additional assets, though these are excluded from the main analysis as their returns are already represented within the broader commodity category.

We focus on tracking three key macroeconomic factors: inflation, consumption growth, and industrial production growth. All three factors are sourced from the FRED database (Federal Reserve Bank of St. Louis, 2024) and have been shown to significantly influence asset prices. Industrial production growth serves as a proxy for overall economic growth due to its monthly frequency and timely release.

Table 5 in Appendix A shows that there is a significant correlation between the considered asset returns and future values of the macroeconomic factors. These correlations remain largely unchanged after adjusting for the risk-free rate, which is of relevance when constructing zero-cost portfolios. We notice that the correlation differs considerably when +1 month or +1 year macroeconomic factors are considered. This is an indication that changes in short-term and long-term expectations affect asset returns differently.

Figure 1 presents the correlation matrix of the assets alongside the risk-free rate. The figure reveals significant positive correlations, especially between bonds and credits, equities and high-yields, and oil and commodities. These correlations are important to consider during portfolio construction to avoid excessive exposure to a single factor.

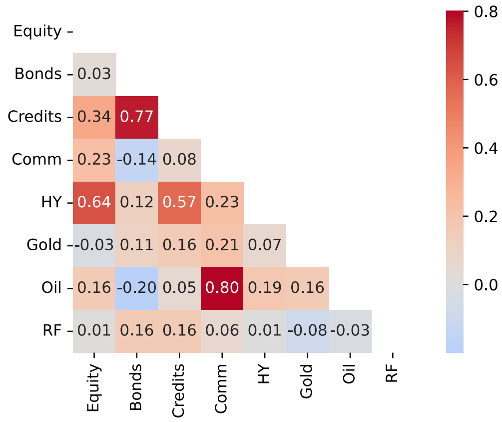


Figure 1: Correlation matrix including returns of equity, bonds, credits, commodities (comm), high-yields (HY), gold, oil, and the risk-free rate (RF) from August 1983 to January 2023.

For ETPs, the lagged control variables should be informative for forecasting asset returns. We use several control variables which are known to be historically significant in predicting asset returns and economic activity:

1. **The year-over-year change in term spread**: The term spread is the difference between 10-year and 1-month US Treasury bill yield, which is shown to be predictive of future economic activity. A positive term spread, where the 10-year yield is higher than the 1-month yield, is associated with an inverted yield curve and is predictive of an incoming recession (Estrella & Hardouvelis, 1991; Harvey, 1993).

2. **The year-over-year change in credit spread**: The credit spread is defined as the average difference between 20-year BAA corporate and 20-year US Treasury yields. A higher value is associated with corporate bonds trading at a premium and indicates higher default probabilities. Fama and French (1989) and Keim and Stambaugh (1986) show that this variable is predictive of future recessions and bond yields.

3. **The average S&P 500 dividend yield**: The average yield of the S&P 500 index is shown to be predictive of the future equity premium by Fama and French (1988) and Campbell and Shiller (1988).

4. **Macroeconomic factors**: We use the year-over-year industrial production growth, consumption growth and inflation macroeconomic factors as defined above as control variables. Stock and Watson (1989) and Fama (1981) illustrate the importance of these macroeconomic variables for predicting returns.

5. **The year-over-year change in average S&P 500 earnings yield**: The earnings yield is defined as the inverse of the price/earnings ratio. A higher earnings yield is predictive of higher future returns (Basu, 1983).

We assess the stationarity of all splitting variables using the Augmented Dickey-Fuller test. The results, presented in Appendix D, confirm that all considered variables are stationary, which is important for the validity of the subsequent analyses and model estimations.

4 Methodology

4.1 Linear Economic Tracking Portfolios

Lis and Porqueddu (2018) revealed that inflation exhibits a seasonal pattern, even when accounting for food and energy price fluctuations. Attempting to create an inflation-tracking portfolio that accurately captures these seasonality and trend components is unrealistic, as it would reveal arbitrage opportunities by trading against expected macro factors. This would be inconsistent with the rational expectations hypothesis in economics (Muth, 1961), which states that economic agents use all available information to form expectations about future economic conditions. Instead, we will focus on the changes in future expected inflation, as described by Lamont (2001), which we refer to as 'news'. By definition, news has a zero mean and can be naturally tracked in a portfolio. A news-tracking portfolio, hereafter referred to as an economic tracking portfolio, does not attempt to capture expected components of underlying macro factors but remains useful for other purposes such as hedging unexpected changes in future macroeconomic factors.

In the following equations, t represents the current timestamp, while s will be used to denote historical timestamps in later sections. The economic tracking portfolio proposed by Lamont (2001) works as follows. News is defined as

$$\Delta\mathbb{E}\left[y_{t+h} \mid \mathcal{F}_t\right] = \mathbb{E}\left[y_{t+h} \mid \mathcal{F}_t\right] - \mathbb{E}\left[y_{t+h} \mid \mathcal{F}_{t-1}\right], \tag{1}$$

where $y_{t+h} \in \mathbb{R}$ is a macroeconomic variable such as consumption or inflation at time $t+h$ and $\mathcal{F}_t$ is the information set known at the end of timestamp t. The tracking portfolio returns are given by $r_t = w'R_t$, where r_t denotes tracking portfolio returns from the end of timestamp $t-1$ to the end of timestamp t, $R_t \in \mathbb{R}^K$ denotes the vector of asset returns from the end of timestamp $t-1$ to the end of timestamp t, and $w \in \mathbb{R}^K$ denotes the desired weights of the tracking portfolio. h corresponds to the desired forecasting horizon for the macroeconomic variable. For instance, when forecasting one-year ahead inflation, we would let $h = 12$ and let y_{t+12} denote the year-over-year inflation from timestamp t to $t+12$.

$R_t \in \mathbb{R}^K$ consists of zero-cost portfolio returns. Zero-cost portfolios involve taking long positions in some assets and short positions in others, such that the overall cost is zero. In our case, we subtract the risk-free rate to turn asset returns into zero-cost portfolios. This approach is advantageous because it allows us to avoid imposing any leverage constraints on the portfolio, thereby eliminating the need to adjust the scale of the macroeconomic variables to match those of returns. Since the tracking portfolio is constructed as a linear combination of these zero-cost portfolios, the resulting portfolio also maintains a zero-cost structure.

To construct the tracking portfolio and determine w, Lamont (2001) considers unexpected returns $\tilde{R}_t = R_t - \mathbb{E}[R_t \mid \mathcal{F}_{t-1}]$ as opposed to total returns and choose w to maximise the correlation between $\tilde{r}_t = w'\tilde{R}_t$ and news $\Delta\mathbb{E}[y_{t+h} \mid \mathcal{F}_t]$. This leads to the regression equation

$$\Delta\mathbb{E}[y_{t+h} \mid \mathcal{F}_t] = w'\tilde{R}_t + \eta_t, \tag{2}$$

where $\eta_t \in \mathbb{R}$ corresponds to the component of news that cannot be explained by unexpected returns. We expect w to be non-trivial for any macro variable y_t that is correlated with future cash flows and discount rates, as unexpected returns are expected to reflect this information.

The term $\mathbb{E}[y_{t+h} \mid \mathcal{F}_t]$ cannot be directly observed, but through algebraic manipulation, we can derive the following relationship:

$$\begin{aligned} y_{t+h} &= \mathbb{E}[y_{t+h} \mid \mathcal{F}_t] + e_{t,t+h} \\ &= \mathbb{E}[y_{t+h} \mid \mathcal{F}_{t-1}] + \Delta\mathbb{E}[y_{t+h} \mid \mathcal{F}_t] + e_{t,t+h}. \end{aligned} \tag{3}$$

Additionally, we control expected returns for control variables:

$$\mathbb{E}[R_t \mid \mathcal{F}_{t-1}] = dZ_{t-1}, \tag{4}$$

with $Z_{t-1} \in \mathbb{R}^J$ representing control variables known at the end of timestamp $t-1$. We assume that expected returns are linear functions of these control variables. Any violation of this assumption could lead to misspecification. However, Lamont (2001) argue that returns are largely unpredictable over short time horizons, suggesting that this model is relatively robust to misspecification (Campbell, 1991).

We control for expected returns using the set of predictor variables Z_{t-1}:

$$\mathbb{E}[y_{t+h} \mid \mathcal{F}_{t-1}] = fZ_{t-1} + \mu_{t-1}. \tag{5}$$

This approach is consistent with the literature on return predictability (Campbell & Shiller, 1988; Fama & French, 1989) and helps to isolate the news component of returns. Through algebraic manipulation of (1) - (5), the final estimation equation is obtained:

$$\begin{aligned} &\Delta\mathbb{E}[y_{t+h} \mid \mathcal{F}_t] = w'\tilde{R}_t + \eta_t \\ &\implies y_{t+h} - \mathbb{E}[y_{t+h} \mid \mathcal{F}_{t-1}] - e_{t,t+h} = w'(R_t - \mathbb{E}[R_t \mid \mathcal{F}_{t-1}]) + \eta_t \\ &\implies y_{t+h} = w'R_t - w'dZ_{t-1} + \mathbb{E}[y_{t+h} \mid \mathcal{F}_{t-1}] + e_{t,t+h} + \eta_t \\ &\implies y_{t+h} = w'R_t - w'dZ_{t-1} + fZ_{t-1} + \mu_{t-1} + e_{t,t+h} + \eta_t \\ &\implies y_{t+h} = w'R_t + cZ_{t-1} + \epsilon_{t,t+h}, \end{aligned} \tag{6}$$

where $c = f - w'd$ and $\epsilon_{t,t+h} = \mu_{t-1} + e_{t,t+h} + \eta_t$. Equation 6 can consistently be estimated using least-squares because the three components of $\epsilon_{t,t+h}$ are independent with R_t and Z_{t-1}. The resulting estimate can be used to find the corresponding weights for the economic tracking portfolio, w.

4.2 Random Forest-based Economic Tracking Portfolios

To capture non-linear relationships and interactions between macroeconomic variables, we extend the linear ETP framework using RFs. This approach is motivated by the growing evidence of non-linearities and regime-switching behaviour in macroeconomic relationships (Teräsvirta, 1994; Hamilton, 2016).

We adapt the asset allocation forest (AAF) framework introduced by Bettencourt et al. (2024) to minimise tracking error instead of maximising the Sharpe ratio. The random forest underpinning an RF ETP consists of a collection of B decision trees. For each tree T_b in the forest, we determine the leaf $L_b(Z_{t-1})$ based on the vector of macroeconomic variables $Z_{t-1} \in \mathbb{R}^J$ known at the end of timestamp $t-1$. Within each leaf, we compute a corresponding target portfolio $\hat{w}_b(Z_{t-1})$ that minimises tracking error:

$$\hat{w}_b(Z_{t-1}) = \underset{w \in \mathbb{R}^K}{\operatorname{argmin}} \quad \left[\lambda \|w\|_2^2 + \frac{1}{L_b(Z_{t-1})} \sum_{s \in L_b(Z_{t-1})} \left(w' R_s - y_{s+h} \right)^2 \right],$$

where $R_s \in \mathbb{R}^K$ denotes the vector of asset returns at historical timestamp s and λ is a penalty on the L_2-norm of w which acts as a form of regularisation. This parameter is found via 5-fold cross-validation within each individual leaf. This optimisation problem is equivalent to Ridge regression, which has a closed-form analytical solution.

Our splitting criterion differs from the traditional CART methods (Breiman et al., 1984), focusing on minimising the sum of tracking errors across child nodes instead of impurity:

$$\begin{aligned}(j^*, \tau^*) = \underset{\substack{j \in \{1,\dots,J\},\\ \tau \in \mathbb{R}}}{\operatorname{argmin}} \Bigg[& \min_{w_{\text{left}} \in \mathbb{R}^K} \left(\sum_{s: Z_{t-1,j} \leq \tau} \left[(w'_{\text{left}} R_s - y_{s+h})^2 + \lambda_{\text{left}} \left\| w_{\text{left}} \right\|_2^2 \right] \right) \\ & + \min_{w_{\text{right}} \in \mathbb{R}^K} \left(\sum_{s: Z_{t-1,j} > \tau} \left[(w'_{\text{right}} R_s - y_{s+h})^2 + \lambda_{\text{right}} \left\| w_{\text{right}} \right\|_2^2 \right] \right) \Bigg],\end{aligned} \tag{7}$$

where j^* corresponds to the optimal splitting feature, τ^* corresponds to the optimal splitting threshold, and $\lambda_{\text{left}}, \lambda_{\text{right}} \in \mathbb{R}_{\geq 0}$ are L_2 regularisation parameters found using 5-fold cross-validation within the leaves. This approach takes the regularisation effect of the Ridge regressions in the leaves into account while constructing the splits and ensures that the resulting splits focus on minimising tracking errors, regardless of the current macroeconomic environment.

Similar to ordinary random forests, we perform recursive subpartitioning of the data where we iteratively split the data into increasingly smaller partitions or leaves until we reach leaf sizes that are smaller than a predefined threshold. This process allows the trees to capture complex interactions and non-linear relationships in the data. We also make use of bagging, as introduced by Breiman (1996), which involves generating multiple subsets of the training data making use of random sampling with replacement. This enhances model performance by reducing variance. Each tree in the forest is trained on a different sample, ensuring that the trees are diverse which improves the model's overall robustness and out-of-sample performance.

To obtain the overall posterior portfolio weights from the RF ETP, we take the average of the predicted portfolios from all underlying trees:

$$\hat{w}(Z_{t-1}) = \frac{1}{B} \sum_{b=1}^{B} \hat{w}_b(Z_{t-1}). \tag{8}$$

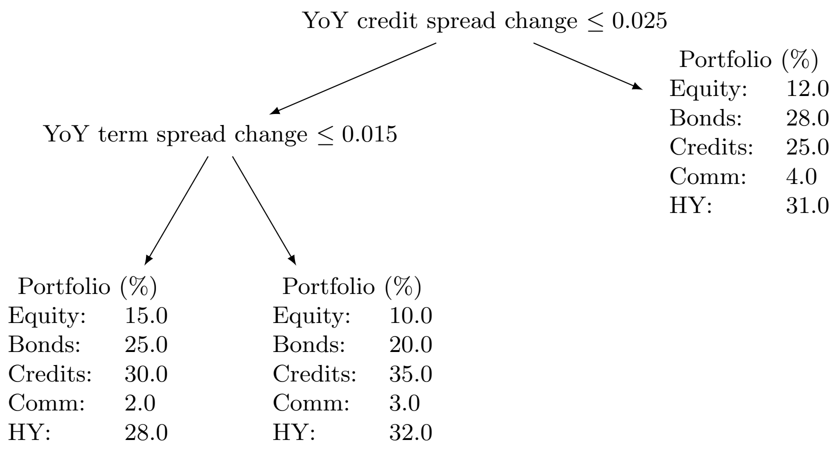


Figure 2: Asset allocation tree based on year-over-year (YoY) changes in credit spread and term spread, illustrating how different economic conditions lead to distinct portfolio allocations.

Figure 2 illustrates an example of an asset allocation tree using credit spread and term spread as macroeconomic indicators. This tree demonstrates how different economic conditions lead to different portfolio allocations.

4.3 Local Linear Forest-based Economic Tracking Portfolio

To further enhance the ability to capture both global non-linear structures and local linear relationships, we employ local linear forests (LLFs) as introduced by Friedberg et al. (2020). This approach combines the strengths of random forests in handling interaction effects with the precision of local linear regression. This is particularly relevant given the potential for both global non-linearities and local linear trends in macroeconomic relationships (Teräsvirta & Anderson, 1992).

LLFs build on the adaptive kernel method interpretation of RFs, differing from the traditional ensemble perspective of random forests. This reinterpretation allows for a more nuanced prediction model that uses the forest structure to weigh observations in a local linear regression. To illustrate this concept, we consider a simplified univariate prediction scenario.

Specifically, consider a dataset of size S with observations $(Z_1, Y_1), \ldots, (Z_S, Y_S)$, where the goal is to predict $\mu(z) = \mathbb{E}[Y \mid Z = z]$. The RF prediction is expressed as

$$\hat{\mu}(z) = \frac{1}{B} \sum_{b=1}^{B} \sum_{s=1}^{S} Y_s \cdot \frac{\mathbb{1}\{Z_s \in L_b(z)\}}{|L_b(z)|} = \sum_{s=1}^{S} Y_s \cdot \frac{1}{B} \sum_{b=1}^{B} \frac{\mathbb{1}\{Z_s \in L_b(z)\}}{|L_b(z)|} = \sum_{s=1}^{S} \alpha_s(z) Y_s. \tag{9}$$

Here, $\alpha_s(z) \in \mathbb{R}_{\geq 0}$ represents the weights implicitly generated by the random forest for each observation s, thus framing the random forest as a kernel method (Athey et al., 2019; Meinshausen & Ridgeway, 2006; Hothorn et al., 2004). The LLF extends this concept by not simply averaging the weighted observations, but by fitting weighted ridge regressions for each prediction point using the kernel weights $\alpha_s(z)$.

While this example illustrates the kernel interpretation for a univariate prediction, our LLF ETP extends this concept to predict entire portfolios. Specifically, instead of predicting a single value $Y \in \mathbb{R}$, we predict a vector of portfolio weights $w \in \mathbb{R}^K$. The LLF ETP combines the kernel interpretation with the regression equation by Lamont (2001). However, instead of taking the average portfolio weight as done with RF ETPs, we fit a weighted ridge regression on all historical data for each prediction. The weights for this regression are provided by $\alpha_s(Z_t)$, the kernel weights derived from the RF structure, where t represents the current timestamp and s represents the timestamp of a historical data point. The prediction formula for the LLF ETP is:

$$\begin{bmatrix} \hat{c}(Z_t) \\ \hat{w}(Z_t) \end{bmatrix} = \left(\begin{bmatrix} Z^{\text{train}\prime} \\ R^{\text{train}\prime} \end{bmatrix} A(Z_t) \begin{bmatrix} Z^{\text{train}} & R^{\text{train}} \end{bmatrix} + \lambda I \right)^{-1} \begin{bmatrix} Z^{\text{train}\prime} \\ R^{\text{train}\prime} \end{bmatrix} A(Z_t) Y^{\text{train}}, \qquad (10)$$

where $R^{\text{train}} \in \mathbb{R}^{S \times K}$ is the matrix of zero-cost asset returns for the S historical train observations, $Z^{\text{train}} \in \mathbb{R}^{S \times J}$ is the matrix of lagged covariates for all historical observations, and $Y^{\text{train}} \in \mathbb{R}^S$ are the corresponding future target factors y_{s+h} for all historical observations. $\lambda \in \mathbb{R}_{\geq 0}$ corresponds to the strength of the L_2-penalty and is found using 5-fold cross-validation for each separate predicted value. $A(Z_t) \in \mathbb{R}^{S \times S}$ is a diagonal matrix where $A(Z_t)_{ss} = \alpha_s(Z_t)\ \forall s \in \{1, \dots, S\}$ and $A(Z_t)_{su} = 0\ \forall s \neq u$. The weights $\alpha_s(Z_t)$ are derived from the RF and represent the importance of the historical observations for the current prediction in the weighted regression.

$\hat{c}(Z_t) \in \mathbb{R}^J$ and $\hat{w}(Z_t) \in \mathbb{R}^K$ correspond to the final obtained parameters in Lamont (2001)'s regression equation. However, unlike in Lamont's original regression equation, these parameters are dependent on current macroeconomic conditions via the adaptive LLF kernel weighting. The final portfolio is $\hat{w}(Z_t)$, while $\hat{c}(Z_t)$ is truncated and not used in the final portfolio construction. This truncation is consistent with Lamont's original approach, where the c coefficients are used to control for expected returns but are not part of the final tracking portfolio.

This approach allows for a flexible, data-driven method to capture both global non-linear structures and local linear trends, potentially improving the model's ability to adapt to different economic regimes.

4.4 Model Training, Hyperparameters, and Evaluation

To construct our ETPs, we employ an expanding approach, retraining the models annually. The initial training set spans 20 years, ensuring sufficient data for accurate predictions. Consequently, our first portfolio is constructed in August 2003, with the final portfolio construction occurring in December 2022.

For hyperparameter selection, we adopt a conservative approach to avoid potential look-ahead bias. Instead of performing cross-validation every training iteration, we use fixed hyperparameters based on recommendations from prior studies for datasets of similar size. These include parameters such as the number of trees, minimum observations per leaf, and maximum tree depth. A detailed list of hyperparameters and their values can be found in Appendix B.

To evaluate portfolio performance, we use the Mincer-Zarnowitz (MZ) regression (Mincer & Zarnowitz, 1969):

$$y_{t+h} = \alpha + \beta \cdot r_t + \epsilon, \tag{11}$$

where $y_{t+h} \in \mathbb{R}$ is the future target variable, $r_t \in \mathbb{R}$ are the portfolio returns, and α and β are estimated using OLS. This regression accounts for potential constant risk premia (α) and scales the returns to match the target variable, regardless of the average portfolio leverage (β). The primary metric for assessing tracking performance is the R^2 of this regression, which quantifies how well ETP returns predict the future macroeconomic variable.

We conduct robustness checks to investigate the sensitivity of our results to hyperparameter choices, which are detailed in Appendix C. These checks demonstrate that while there is some variation in performance with different hyperparameters, the RF and LLF models consistently outperform linear ETPs across a wide range of parameter values.

4.5 Portfolio Interpretation using Shapley Values

To interpret the complex relationships captured by RF and LLF ETPs, we employ Shapley values. Shapley values, originally derived from cooperative game theory (Shapley, 1953), are widely used in machine learning to interpret model predictions (Lundberg & Lee, 2017; Štrumbelj & Kononenko, 2014). In our context, Shapley values can help us understand how different macroeconomic indicators contribute to the allocation decisions within the tracking portfolios.

Let w_{t+1} be the portfolio weights of the portfolio constructed at the end of timestamp t. These weights are determined by the covariates $Z_t \in \mathbb{R}^J$. We define a portfolio metric $M(w_t)$ as a function of the weights. For example, we might be interested in the relative allocation between commodities and bonds,

$$M(w_t) = \frac{w_{t,\text{comm}} - w_{t,\text{bonds}}}{\|w_t\|_1}, \tag{12}$$

where $\|w_t\|_1 = \sum_{k=1}^{K} |w_{t,k}|$ is the L_1-norm of the weights, representing the total portfolio leverage.

The Shapley value for covariate $j \in \{1, \ldots, J\}$ is given by

$$\phi_j = \sum_{Q \subseteq N \setminus \{j\}} \frac{|Q|!(|N| - |Q| - 1)!}{|N|!} \left[v(Q \cup \{j\}) - v(Q) \right], \tag{13}$$

where $N = \{1, 2, \ldots, J\}$ is the set of all covariates, Q is a subset of N that does not include j, and $v(Q)$ is the value function, which represents the expected portfolio metric $M(w_t)$ when only the covariates in subset Q are known:

$$v(Q) = \mathbb{E}\left[M(w_t) \mid Z_{t-1,Q} \right]. \tag{14}$$

In practice, we approximate this expectation using a Monte Carlo approach, shuffling the values of covariates not in S and averaging the portfolio metrics (Lundberg & Lee, 2017).

This framework allows us to quantify the impact of each macroeconomic covariate on portfolio allocation decisions. For instance, we can investigate how the year-over-year inflation rate affects the relative allocation between commodities and bonds in our inflation-tracking portfolio.

Moreover, this approach allows us to extend the analysis of Lohre et al. (2020), who identified static relationships between asset classes and macroeconomic factors. By computing Shapley values across different time periods, we can investigate how these relationships evolve dynamically, providing insights into changing economic regimes.

4.6 Portfolio Robustness to the Choice of Underlying Assets using Shapley Values

The performance and stability of ETPs can be significantly influenced by the selection of underlying assets. While linear ETPs regress on the entire historical dataset to obtain optimal portfolio weights, RF and LLF-based ETPs utilise more localised approaches. RF ETPs perform Ridge regression within the leaves of trees, often relying on as few as 36 months of data, while LLF ETPs perform weighted Ridge regression where weights are unevenly distributed across historical data points. These approaches, while beneficial for capturing non-linear relationships, may lead to potential instability when dealing with a large number of assets or in the presence of uninformative assets.

To investigate the robustness of portfolios to the choice of underlying assets, we examine two key aspects: the impact of incorporating many tradable, informative assets, and the effect of adding an uninformative white noise (WN) asset. This analysis is important for understanding the practical implications of using machine learning-based ETPs in real-world scenarios where asset selection may not always be optimal.

We adopt a methodology inspired by the Shapley value approach used in machine learning interpretability (Lundberg & Lee, 2017). In our context, instead of estimating the average impact of a feature on a prediction, we estimate the average impact of including a specific number of assets on portfolio performance. This allows us to quantify how the number of assets affects the tracking ability of different ETP methods.

The procedure is as follows:

1. **Determine Asset Subset**: For a given number of assets K (ranging from 3 to 7), we randomly select up to 20 different combinations of assets.

2. **Fit Portfolio**: For each chosen asset combination, we fit a portfolio using data up to 31 December 2009, allowing the portfolio to trade only the selected assets.

3. **Evaluate Portfolio**: We determine the out-of-sample MZ R^2 for each fitted portfolio using data from 1 January 2010 to 31 December 2019, then calculate the average MZ R^2 across all combinations for each K.

This approach allows us to assess how the performance of different ETP methods varies with the number of available assets, providing insights into their scalability.

To investigate the impact of potentially uninformative assets, we introduce a white noise (WN) asset into our analysis. The WN asset serves as a neutral asset, having zero correlation with all other assets and no information about future macroeconomic variables. This represents the real-world scenario where certain assets in a portfolio may not contribute meaningfully to tracking the target economic variable.

We repeat the above procedure, but this time including the WN asset in addition to the real assets. We perform this analysis three times with different realisations of the WN asset to reduce variance in our estimates. By comparing the average MZ R^2 of portfolios including and excluding the WN asset, we can assess the robustness of different ETP methods to the inclusion of noisy or irrelevant assets.

This analysis is particularly relevant in the context of economic tracking, where the relationship between asset returns and macroeconomic variables can be complex and time-varying. For instance, during periods of economic stress, certain assets may become less informative for tracking specific economic variables. A robust ETP should be able to maintain its performance even when such assets are included in the available set.

Furthermore, this approach allows us to investigate whether the superior performance of RF and LLF ETPs observed in our main results holds across different asset combinations. If these methods consistently outperform linear ETPs across various asset subsets, it would provide strong evidence for their robustness and practical utility.

4.7 LLF Kernel Interpretation using Kernel PCA

The LLF utilises a weighted ridge regression with weights extracted from the kernel implicitly generated by a random forest. To interpret this high-dimensional kernel and gain insights into the regimes captured by this model, we employ Kernel Principal Component Analysis (KPCA) (Schölkopf et al., 1997).

The essence of KPCA is to apply PCA within the high-dimensional feature space implicitly defined by the kernel function $k(z, z')$. In general, data points can almost surely be linearly separated if the feature space has sufficiently many dimensions. Let $\psi : \mathbb{R}^J \to \mathcal{H}$ be a mapping function that projects a data point Z_t into a Hilbert space. According to the Moore-Aronszajn theorem, for any symmetric positive definite function $k : \mathbb{R}^J \times \mathbb{R}^J \to \mathbb{R}$, there exists a unique Hilbert space $\mathcal{H}$ and a mapping function ψ such that $\forall z, z' \in \mathbb{R}^J$, it holds that

$$k(z, z') := \langle \psi(z), \psi(z') \rangle_{\mathcal{H}}.$$

This theorem implies that our LLF kernel can also be interpreted as mapping from the original feature space into a Hilbert space. This Hilbert space has infinitely many dimensions, making it straightforward to construct hyperplanes that divide the data points into distinct clusters, even if this was not possible before the feature transformation. This separability allows KPCA to effectively reduce the dimensionality while preserving the structure and relationships within the data. The mapping function ψ creates vectors that are linearly independent, eliminating the possibility of a covariance on which traditional PCA relies for eigendecomposition. KPCA circumvents the need to explicitly compute ψ by using the kernel trick.

KPCA is performed by solving the eigenvector equation

$$S\gamma v = (K - \mathbf{1_S}K - K\mathbf{1_S} + \mathbf{1_S}K\mathbf{1_S})v,$$

where S represents the number of data points, γ and v are the eigenvalue-eigenvector pairs, $K \in \mathbb{R}^{S \times S}$ represents the kernel matrix, and $\mathbf{1_S}$ denotes the square S-by-S matrix for which each element takes the value $1/S$. This centers the kernel matrix before performing the eigendecomposition which is necessary because, in general, points in the feature space $\mathcal{H}$ do not have a zero mean.

The economic interpretation of these principal components is important. Each obtained principal component represents a direction in the high-dimensional feature space along which the variation in the data is maximised. In the context of ETPs, these components can be interpreted as representing different economic regimes or factors that influence the relationship between asset returns and macroeconomic variables.

To visualise and interpret these economic regimes, we project the data onto the first two principal components. We can then plot the points in a two-dimensional space. This visualisation allows us to identify clusters of time points that the LLF considers similar in terms of their economic characteristics. For example, points clustered together might represent periods of economic recession, high inflation, or stable growth. The distance between points in this space can be interpreted as a measure of similarity as perceived by the LLF model.

5 Results

5.1 Tracking Performance

Table 1: ETP metrics for inflation, consumption, and growth portfolios. August 2003 - January 2023 with yearly retraining for RF and LLF ETP and monthly retraining for linear ETP. Highest Mincer-Zarnowitz R^2 per horizon and factor are in bold.

		1 month			1 year		
		Lin. ETP	RF ETP	LLF ETP	Lin. ETP	RF ETP	LLF ETP
Infl.	Avg. turnover	0.024	0.026	0.065	0.051	0.091	0.195
	Avg. leverage	0.046	0.013	0.031	0.073	0.027	0.055
	MZ R^2	0.40	**0.41**	0.34	0.044	0.059	**0.069**
	MZ intercept	0.0020	0.0021	0.0020	0.025	0.025	0.025
	MZ slope	1.15	2.86	1.20	1.66	3.98	1.90
Cons.	Avg. turnover	0.019	0.033	0.062	0.066	0.115	0.152
	Avg. leverage	0.024	0.011	0.019	0.106	0.026	0.054
	MZ R^2	0.020	**0.021**	0.017	0.031	0.048	**0.074**
	MZ intercept	0.0014	0.0014	0.0014	0.019	0.019	0.018
	MZ slope	2.09	4.69	1.44	2.26	8.87	4.79
Growth	Avg. turnover	0.030	0.071	0.093	0.038	0.33	0.264
	Avg. leverage	0.041	0.017	0.036	0.074	0.119	0.181
	MZ R^2	0.047	**0.13**	0.11	0.068	0.057	**0.071**
	MZ intercept	0.0009	0.0000	0.0000	0.0050	0.0045	0.0043
	MZ slope	3.86	6.01	2.48	8.89	3.87	2.44

Table 1 reveals that machine learning-based ETPs consistently outperform traditional linear ETPs across various macroeconomic factors and time horizons. For short-term forecasts (+1 month), RF ETPs demonstrate superior performance. Notably, the RF ETP achieves a MZ R^2 of 13% for tracking industrial production growth, compared to just 4.7% for the linear ETP.

This substantial improvement highlights the RF model's ability to capture complex, non-linear relationships between asset returns and macroeconomic variables.

For longer-term forecasts (+1 year), LLF ETPs consistently achieve the highest MZ R^2 values. This could potentially be due to the LLF ETP's ability to model both global non-linear relationships and local linear trends, as described by Friedberg et al. (2020). However, further research would be needed to confirm whether local linear structures indeed become more pronounced or more important for economic tracking over longer time horizons.

Interestingly, the performance improvements vary across different macroeconomic factors. For inflation tracking, the enhancement is relatively modest, with the LLF ETP achieving an MZ R^2 of 6.9% for the 1-year horizon compared to 4.4% for the linear ETP. This relatively small improvement might be attributed to the well-documented challenges in forecasting inflation, as described by Stock and Watson (2007). In contrast, the improvement for consumption growth tracking is more substantial, with the LLF ETP reaching an MZ R^2 of 7.4% compared to the 3.1% for the linear ETP.

To assess the statistical significance of the performance improvements offered by the machine learning-based ETPs, we conduct one-sided Diebold-Mariano (DM) tests comparing the best-performing model to the linear ETP for each factor and horizon. We perform the test by taking the Mincer-Zarnowitz target factor predictions for the different portfolios and comparing them against the true values of the target factor. For the one-year horizon, the LLF ETP shows statistically significant improvements over the linear ETP for both inflation and consumption tracking. The one-sided DM test for inflation tracking yields a test statistic of -1.967 (p value $= 0.0125$), while for consumption tracking, the test statistic is -2.010 (p value $= 0.0113$). This indicates that the superior performance of the LLF ETP for long-term tracking and consumption is statistically significant at the 5% level. For IP growth tracking and all short-term (one-month) horizons, the improvements, while often present, are not statistically significant at the 5% level. This suggests that the enhancements of machine learning-based ETPs are most pronounced for long-term inflation and consumption tracking.

While the machine learning-based ETPs demonstrate superior tracking performance, they also exhibit higher turnover rates. This is also apparent from Figure 3, where the linear ETP exhibits more stable portfolio weights over time. Portfolio weights for the other factors and horizons can be found in Appendix G and show similar results. This increased turnover could potentially lead to higher transaction costs in practical applications, a trade-off that investors and portfolio managers would need to consider. The higher turnover rates are likely a result of the models' ability to adapt more quickly to changing economic conditions, as observed by Gu et al. (2020) in their study of machine learning methods in asset pricing.

It is worth noting that the performance of all ETPs, including linear ones, may be influenced by the specific economic conditions during our sample period. The sample period includes several significant economic events, such as the 2008 financial crisis and the Covid-19 pandemic, which could have impacted the relationship between asset returns and macroeconomic variables. Future research could explore the robustness of these results across different economic regimes.

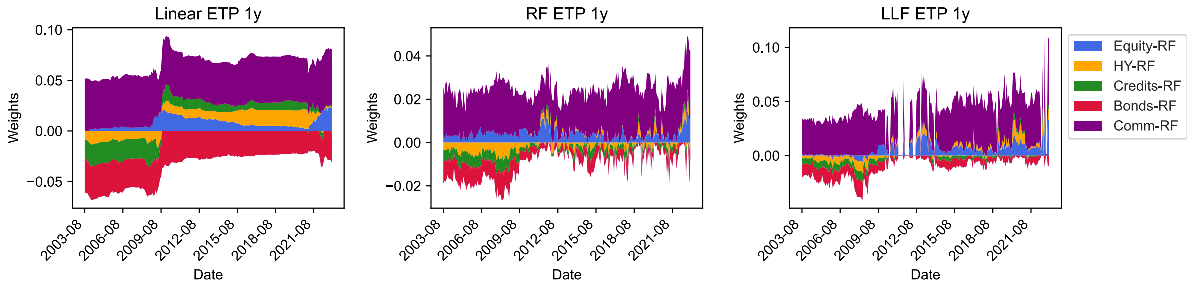


Figure 3: Weights of the inflation tracking portfolios with a 1-year horizon over time, the RFETP and LLETP are retrained yearly and the linear ETP is retrained monthly, August 1983 - January 2023.

5.2 Shapley Value Interpretation: One-Year Ahead Inflation LLF ETP

The LLF ETP demonstrates superior performance in predicting one-year inflation, achieving a MZ R^2 of 6.9% (Table 1). To gain deeper insights into this model's behaviour, we employ Shapley value analysis, focusing on the portfolio metric $M(w_t) = \frac{w_{t,\text{comm}} - w_{t,\text{credits}}}{\|w_t\|_1}$, which represents the relative allocation between commodities and credits, adjusted for overall portfolio leverage.

Lohre et al. (2020) previously showed that commodities are primarily exposed to inflation, while credits are mostly exposed to economic growth. However, our analysis reveals that these relationships are not static but dynamically evolve over time, reflecting changing economic conditions.

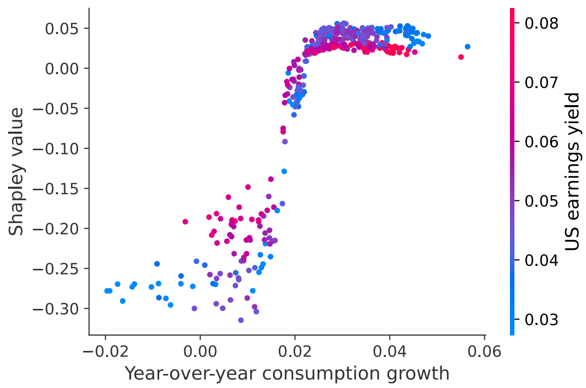


Figure 4: Shapley values with respect to $M(w_t) = \frac{w_{t,\text{comm}} - w_{t,\text{credits}}}{\|w_t\|_1}$ for the LLF inflation tracking portfolio, as a function of year-over-year consumption growth. Data spans August 1983 to January 2020, colour-coded by the S&P 500 average earnings yield.

Figure 4 illustrates the Shapley values from August 1983 to January 2020 as a function of year-over-year consumption growth. Notably, we deliberately exclude Covid-19 from this analysis due to the extreme negative year-over-year consumption growth of -11.6% in April 2020, which would make visualisation more difficult.

We observe a non-linear relationship between consumption growth and the relative importance of commodities versus credits in the inflation tracking portfolio. As consumption growth increases, the portfolio tends to allocate more towards commodities relative to credits, with this effect diminishing at around 2.5% year-over-year consumption growth. This relationship aligns with economic theory and empirical evidence on the link between economic growth and commodity prices. Hamilton (2009) describes that strong global economic growth can make oil prices more responsive to inflationary pressures. Consequently, during periods of high economic growth, commodities could serve as a more effective hedge against inflation, explaining the portfolio's greater commodity allocation.

Interestingly, we also observe an interaction effect with the average S&P 500 earnings yield. When the earnings yield is high, the effect of consumption growth on the commodity-credit allocation becomes less pronounced. Conversely, low earnings yields correspond to more extreme Shapley values, indicating a stronger influence of consumption growth on the portfolio allocation decision. This interaction could be related to the varying sensitivity of different asset classes to economic conditions.

These dynamics can be interpreted in the context of different macroeconomic environments:

1. Economic Stress (e.g., 2008 Financial Crisis): During periods of low consumption growth and low earnings yields, the portfolio shifts towards a higher allocation in credits relative to commodities. This is because credits become more correlated with inflation during economic downturns, as detailed in Appendix H. This shift aligns with the flight-to-quality phenomenon described by Beber et al. (2009), where investors prefer safer, more liquid assets during times of market stress.

2. Economic Growth: As consumption growth increases and the economy expands, commodities become more favoured for inflation tracking. This is consistent with the findings of Gorton and Rouwenhorst (2006), who show that commodity returns are positively correlated with unexpected inflation. They also demonstrate that this correlation varies over time, although they did not explicitly link this to periods of positive economic consumption growth.

3. High Earnings Yield Environments: When earnings yields are high, the effect of consumption growth on the portfolio allocation is dampened. This reflects periods where corporate profitability is high relative to stock prices, potentially indicating a more stable economic environment where the inflation-tracking properties of both commodities and credits are more balanced.

Figure 5 provides a view of how different covariates influence the commodity-credit allocation over time. Consumption growth and earnings yield consistently show the largest effects, highlighting their importance in determining the portfolio's composition. Year-over-year inflation also plays a significant role, particularly during the late 1980s when inflation was high, leading to increased commodity allocation relative to credits.

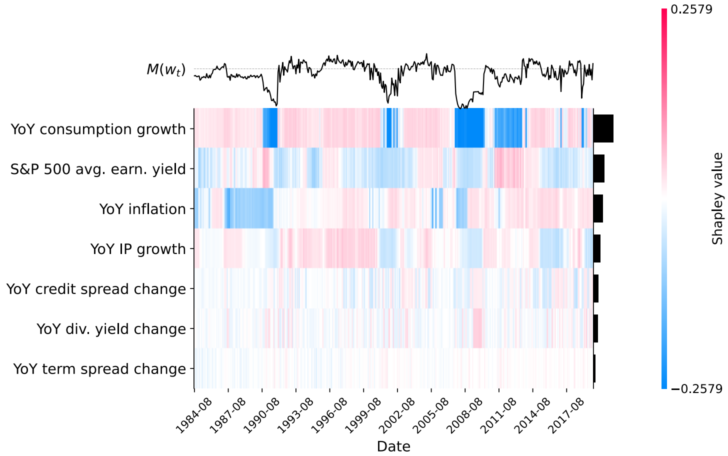


Figure 5: Heatmap of Shapley values with respect to $M(w_t) = \frac{w_{t,\text{comm}} - w_{t,\text{credits}}}{\|w_t\|_1}$ for the LLF inflation tracking portfolio per covariate over time. Data spans August 1983 to January 2020, colour-coded by Shapley value. Red represents shifts towards commodities and blue towards credits. Black bars show the average Shapley value as a proxy for feature importance.

The final part of the analysis concerns the influence of year-over-year industrial production (IP) growth on raw portfolio weights (Figure 6). The corresponding portfolio metric for instrument j is simply $M_j(w_t) = w_{t,j}$, allowing us to directly interpret how IP growth impacts the weight of each asset in the portfolio.

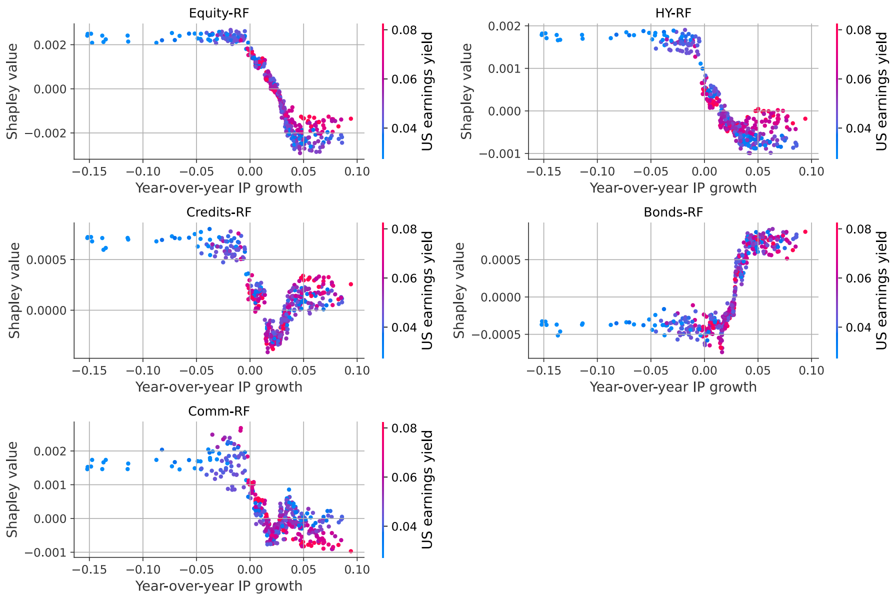


Figure 6: Shapley values illustrating the impact of year-over-year IP growth on individual asset weights in the LLF inflation tracking portfolio, August 1983 - January 2020. Each panel shows how year-over-year changes in IP growth affect the allocation to a specific asset class, with colour indicating the average S&P 500 earnings yield.

When year-over-year IP growth hovers around approximately 2%, which is near the average annual IP growth, the weight allocated to credits is relatively low. This suggests that credits become more correlated with inflation under conditions of economic expansion or contraction. During such periods, companies are more likely to experience fluctuations in their borrowing costs and default risks, which are closely tied to inflationary trends. For example, in an expanding economy, rising demand can drive up prices, leading to higher inflation and more demand for credits. Conversely, in a contracting economy, the risk of deflation might cause central banks to adjust interest rates, impacting credit markets.

Commodities show a similar effect, although with a stronger interaction with the average S&P 500 earnings yield. They are favoured for tracking inflation during periods of low IP growth, likely because they have a baseline demand that remains stable even when industrial production slows.

Bonds, however, became less preferred during periods of low IP growth. This could be because during economic downturns (like the 2008 financial crisis), bonds were largely unaffected while year-over-year inflation became negative due to reduced economic activity and lower consumer prices. By decreasing bond weights during times of low IP growth and economic uncertainty, the portfolio can better track inflation.

These findings suggest that a fixed decomposition of asset classes into macroeconomic factors such as inflation and growth, as proposed by Lohre et al. (2020), does not capture the full dynamics of how these relationships evolve over time. The interactions between asset classes and macroeconomic indicators are complex and vary with changing economic conditions. This result highlights the importance of dynamically adjusting tracking portfolio weights based on macroeconomic indicators to enhance tracking performance.

5.3 Shapley Value Interpretation: One-Year Ahead Consumption Growth LLF ETP

the LLF ETP demonstrates superior performance in predicting one-year consumption growth. To gain deeper insights into this model's behaviour, we employ Shapley value analysis, focusing on the portfolio metric $M(w_t) = |w_{t,\text{HY}}| + |w_{t,\text{equity}}| + |w_{t,\text{credits}}|$, which we refer to as the HEC (High-Yield, Equity, and Credits) leverage.

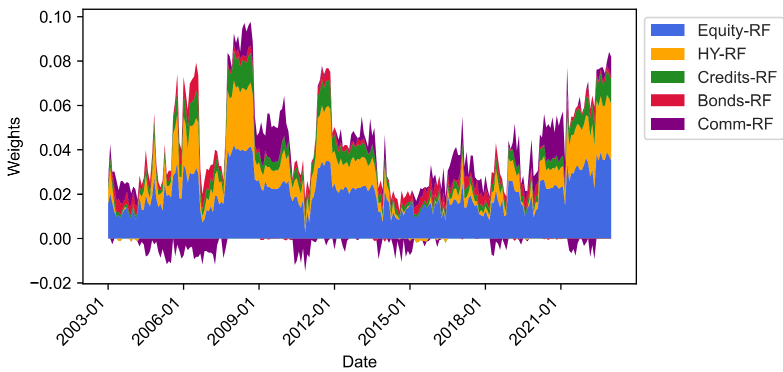


Figure 7: Portfolio weights of the LLF consumption growth tracking portfolio from January 2003 to January 2023. The model was fitted on data up to January 2020 without yearly retraining, which allows for easier portfolio interpretation.

Figure 7 illustrates that while relative proportions of high-yields, equities, and credits remain fairly stable over time, the overall leverage of these assets fluctuates considerably. This suggests that the magnitude of these assets' exposure to consumption growth changes over time, reflecting the model's ability to capture varying economic conditions.

Figure 8 reveals that year-over-year inflation has the most substantial effect on the HEC leverage, followed by year-over-year changes in credit spread. This aligns with the findings of Bansal and Yaron (2004), who demonstrate that economic uncertainty, as captured by variables such as inflation and credit spreads, are significant factors in determining long-run consumption and growth risks.

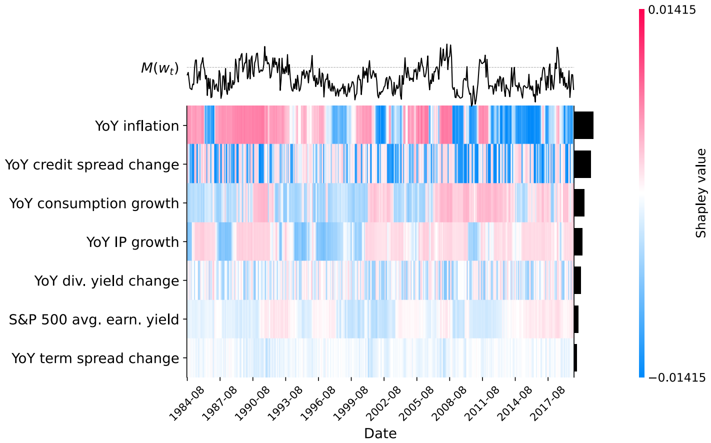


Figure 8: Heatmap of Shapley values with respect to $M(w_t) = |w_{t,\text{HY}}| + |w_{t,\text{equity}}| + |w_{t,\text{credits}}|$ (HEC leverage) for the LLF consumption tracking portfolio, August 1983 - January 2020. The colour represents the Shapley value, with red indicating increased HEC leverage and blue decreased HEC leverage. Black bars show the average Shapley value as a proxy for feature importance.

The relationship between year-over-year inflation and HEC leverage is particularly noteworthy. Figure 9 illustrates a positive, non-linear effect between year-over-year inflation and HEC leverage. The impact of inflation on HEC leverage diminishes when it exceeds approximately 3% per year. Interestingly, there is a slight interaction effect with consumption growth. During periods of low consumption growth combined with high inflation (a sign of stagflation), there is a modest additional increase in HEC leverage. Conversely, when inflation is near its typical level of approximately 2%, the impact of consumption growth on HEC leverage is reversed.

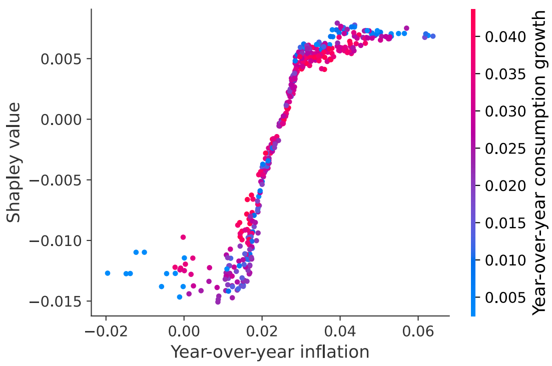


Figure 9: Shapley values with respect to $M(w_t) = |w_{t,\text{HY}}| + |w_{t,\text{equity}}| + |w_{t,\text{credits}}|$ as a function of year-over-year inflation for the local linear consumption tracking portfolio, August 1983 - January 2020. Shapley values indicate the magnitude and direction of inflation's influence on HEC leverage and colour represents year-over-year consumption growth.

The year-over-year change in credit spread also significantly impacts HEC leverage. Figure 10 shows an almost binary response of HEC leverage to this covariate: when the credit spread increases, the HEC leverage of the portfolio rises sharply, whereas if the credit spread decreases, the HEC leverage drops. This finding is consistent with the work of Gilchrist and Zakrajšek (2012), who demonstrate that credit spreads contain significant predictive power for economic activity, including consumption growth.

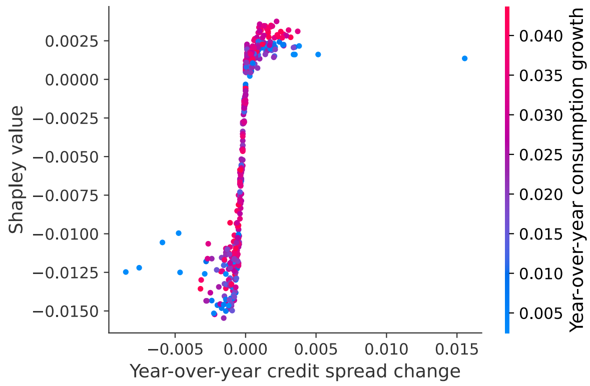


Figure 10: Shapley values with respect to $M(w_t) = |w_{t,\mathrm{HY}}| + |w_{t,\mathrm{equity}}| + |w_{t,\mathrm{credits}}|$ as a function of year-over-year credit spread change for the local linear consumption tracking portfolio, August 1983 - January 2020. Shapley values indicate the magnitude and direction of inflation's influence on HEC leverage and colour represents the year-over-year consumption growth.

The Shapley value analysis highlights the influence of inflation and credit spreads on the HEC leverage of the LLF ETP. This reflects how periods of high inflation and increased credit spreads signal heightened long-run macrooeconomic risk, as proposed by Bansal and Yaron (2004). During such periods, the sensitivity of equities, high-yields, and credits to short-term economic conditions, such as consumption growth, increases. The LLF ETP captures this relationship and increases the HEC leverage during periods of increased long-run macroeconomic risk to better track consumption growth.

5.4 Portfolio Robustness to the Number of Assets and to Uninformative Assets

The performance and stability of ETPs can be significantly influenced by the selection of underlying assets. While linear ETPs use the entire historical dataset to obtain optimal portfolio weights, RF and LLF-based ETPs consider a smaller effective dataset through the means of localised regressions. This makes it crucial to assess the robustness of these portfolios to both the number of assets and the inclusion of potentially uninformative assets.

Figure 11 illustrates the sensitivity of different ETP methods to the number of assets included. The linear ETP shows relative insensitivity to the number of assets for growth and inflation factors. However, for consumption, the linear ETP's performance deteriorates sharply as more assets are included, suggesting potential overfitting.

In contrast, RF and LLF ETPs demonstrate more consistent performance across different numbers of assets for the consumption factor, indicating their ability to handle high-dimensional asset spaces effectively. However, for the IP growth factor, RF and LLF ETPs show higher

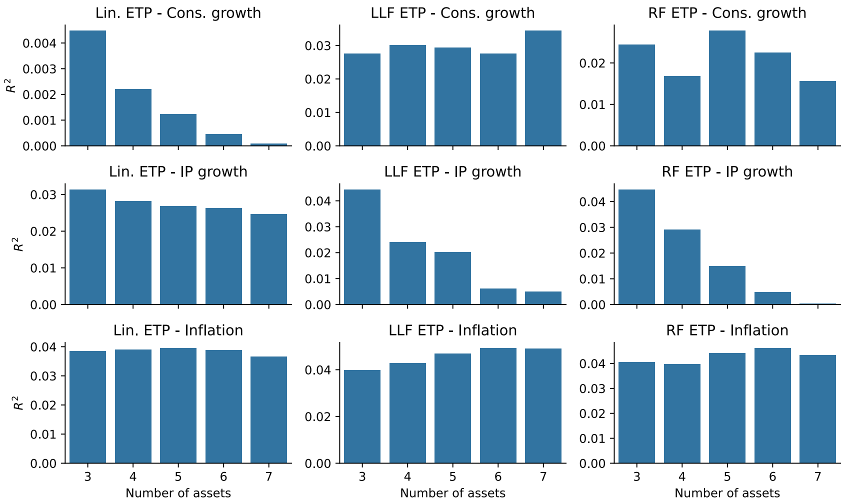


Figure 11: Average MZ R^2 per portfolio per factor per number of assets. Portfolio fitted on data from 31 August 1983 to 31 December 2009, with out-of-sample MZ R^2 from 31 January 2010 to 31 December 2019, 12-month horizon.

sensitivity to the number of assets.

We conclude that all portfolios are somewhat sensitive to the number of assets included, but that it depends a lot on the considered factor. In general, we observe that including too more assets usually leads to worse performance.

To assess robustness to uninformative assets, we introduce a white noise (WN) asset into our analysis and evaluate its impact on portfolio performance.

Table 2: Impact of including a white noise (WN) asset on ETP performance across different forecast horizons and portfolio sizes. Values represent the change in average MZ R^2 across all macroeconomic factors when the WN asset is added to portfolios with varying numbers of real assets (3 to 7). Negative values indicate performance deterioration.

Horizon	Portfolio	3 Assets	4 Assets	5 Assets	6 Assets	7 Assets
1-month	Linear ETP	-0.056	-0.056	-0.056	-0.057	-0.055
	RF ETP	-0.066	-0.061	-0.065	-0.060	-0.056
	LLF ETP	-0.118	-0.124	-0.128	-0.119	-0.108
12-month	Linear ETP	-0.019	-0.020	-0.018	-0.019	-0.017
	RF ETP	-0.040	-0.028	-0.029	-0.024	-0.022
	LLF ETP	-0.045	-0.038	-0.040	-0.035	-0.041

Table 2 reveals that for both short and long-term horizons, the linear ETP consistently demonstrates the highest robustness against the inclusion of a WN asset. The LLF ETP shows the least robustness, with the RF offering a middle ground.

The lower robustness of RF and LLF ETPs to uninformative assets can be explained by their localised approach to portfolio construction. Forests rely on partitioning the dataset, which can lead to overfitting when performing regressions within the leaves. This is particularly pronounced in the LLF, where the macroeconomic covariates are also included in the localised regression similarly to the linear ETP regression equation as introduced by Lamont (2001).

To conclude, we find that all three portfolios are sensitive to the number of available assets to some extent. The linear ETP, however, demonstrates significantly greater robustness to uninformative assets than the RF and LLF ETPs. These findings highlight a trade-off in using machine learning-based ETPs. While they offer improved tracking performance by capturing non-linear relationships and interactions, they are more susceptible to overfitting when irrelevant assets are included.

5.5 LLF Kernel Interpretation

The LLF generates a kernel that implicitly defines a high-dimensional feature space. To interpret this complex kernel and gain insights into the economic regimes captured by the model, we employ Kernel Principal Component Analysis (KPCA) (Schölkopf et al., 1997). This technique allows us to project the high-dimensional feature space onto lower dimensions, allowing for visualisation and interpretation of the economic regimes identified by the LLF.

We first examine whether the kernels differ significantly across different macroeconomic target factors. To investigate this, we calculate the kernel matrices from the one-year horizon LLFs for inflation, consumption, and growth factors. We fit these portfolios on all data from August 1983 to January 2023.

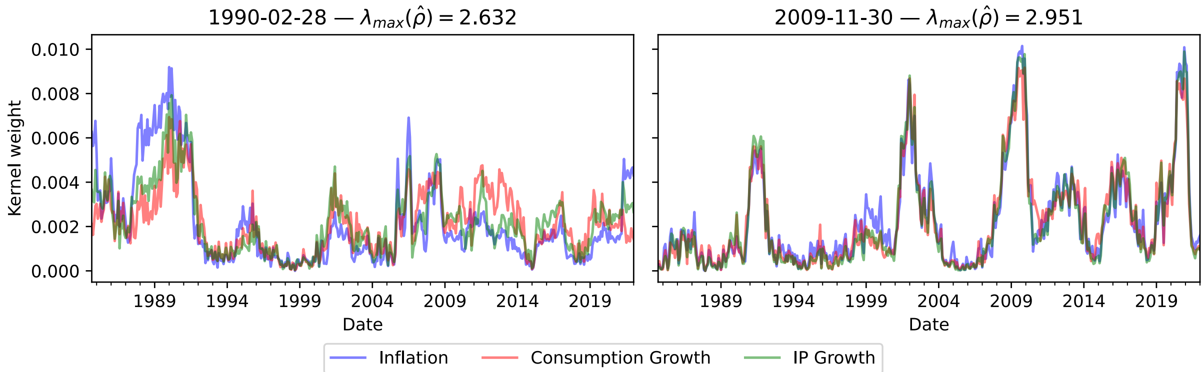


Figure 12: Comparison of LLF kernel weights across three ETPs (inflation, consumption, and growth) for two specific timestamps: February 1990 and November 2009. The portfolios were fitted on data from 31 August 1983 to 31 January 2023, 12-month horizon. The maximum eigenvalue of the 3x3 sample correlation matrix of kernel weights over time is also shown.

Figure 12 presents the kernel weights of two different timestamps over time for the three different LLF ETPs. To quantify the similarity between the learned kernels, we calculate the maximum eigenvalue of the sample correlation matrix between kernel weights. For the kernel weights with respect to November 2009, the maximum correlation matrix eigenvalue is 2.951, indicating that a single factor explains approximately $2.951/3 \approx 98.4\%$ of the standardised variance. This high similarity suggests that during the financial crisis, all three tracking portfolios consider approximately the same data points for constructing their portfolios. The portfolios placed high weights on other periods of economic stress, such as the savings and loan crisis from the early 1990s, the dot-com bubble around 2002, and the Covid-19 pandemic around 2021.

In contrast, for February 1990, the similarity is much lower, with a maximum eigenvalue of 2.632, suggesting that only about 87.7% of the standardised variance is explained by a single factor. This indicates that during more stable economic periods, the portfolios diverge more in their consideration of historical data points. For example, the consumption tracking portfolio puts significantly more weight on the period 2011-2015 than the other two portfolios, while the inflation tracking portfolio places more emphasis on the period 1987-1991.

These findings align with the concept of time-varying correlations in financial markets, as discussed by Longin and Solnik (2001). During periods of economic stress, correlations between different asset classes and economic variables tend to increase, a phenomenon often referred to as "correlation breakdown". Our results suggest that the LLF model captures this effect, leading to more similar kernel weights across different economic tracking portfolios during crisis periods.

To further investigate which periods are considered to be in the same economic regime, we use KPCA to project the high-dimensional latent Hilbert space generated by the LLF kernel onto two dimensions. This visualisation allows us to identify which timestamps are perceived as similar and thus in the same regime. We focus on the inflation tracking portfolio, as kernel weights are generally close across different factors.

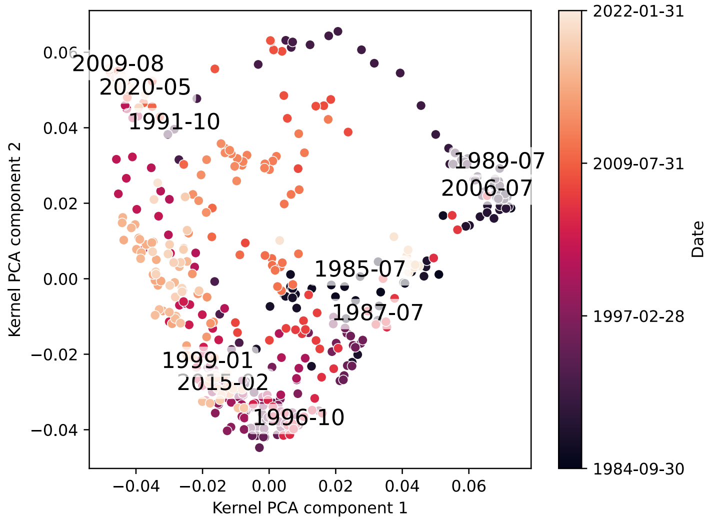


Figure 13: Kernel PCA projection of LLF kernel weights for the 12-month inflation tracking portfolio. The plot shows data points from 31 August 1983 to 31 January 2023 projected onto the first two principal components, revealing distinct clusters corresponding to different economic regimes. Colour represents the date of a data point.

Figure 13 illustrates the data points projected onto two dimensions via KPCA. The clustering of data points reveals distinct economic regimes:

1. Economic Crises (top left): The 2008 financial crisis, Covid-19 pandemic, and the 1991 recession are clustered together. This grouping aligns with the findings of Reinhart and Rogoff (2009), who argue that financial crises, despite their varied origins, often share similar patterns.

2. Overheated Economy (right side): Data points corresponding to 2006 (right before the 2008 financial crisis) and 1989 (preceding the 1991 recession) are grouped together. This cluster likely represents periods of economic overheating, characterised by rapid economic growth and excessive credit growth (Kaminsky & Reinhart, 1999).

3. Reaganomics (bottom right): This cluster represents the economic policies of the 1980s, characterised by tax cuts, deregulation, and increased military spending.

4. Stable Growth (bottom): The late 1990s and the period of growth around 2015 are clustered here. This region likely represents periods of stable, moderate economic growth.

Interestingly, the Reaganomics cluster is positioned between stable growth and economic overheating. This intermediate positioning suggests that the LLF model identifies Reaganomics as sharing some characteristics with both stable growth periods and overheating economies. This aligns with the mixed outcomes of the period, as described by Niskanen (1988). Niskanen notes that while Reaganomics led to significant economic growth, it also contributed to increased federal debt and trade deficits, signs of potential economic instability.

The clear separation of these clusters demonstrates the LLF model's ability to identify and differentiate between distinct economic regimes. Moreover, the clustering patterns provide insights into the model's understanding of economic dynamics. For instance, the proximity of pre-crisis periods (e.g., 2006, 1989) to each other suggests that the model captures early warning signs of economic instability. This aligns with research on leading indicators of financial crises (Kaminsky & Reinhart, 1999).

5.6 Differences Between LLFs and RFs in Splitting Decisions

The splitting decisions made by RFs and LLFs have significant implications for the performance and interpretability of ETPs. A key difference between these methods lies in how they handle the covariates originally used in Lamont (2001)'s regression equation, leading to distinct splitting behaviours.

Lamont's original regression equation includes covariates to adjust for expected returns:

$$y_{t+h} = w' R_t + c Z_{t-1} + \epsilon_{t,t+h}, \tag{15}$$

where Z_{t-1} represents the covariates known at time $t-1$. The LLF incorporates this structure directly into its leaf models, allowing it to focus its splits on covariates that have non-linear effects on the relationship between returns and future macroeconomic factors. In contrast, the RF must use splits to capture both linear and non-linear effects of covariates, as it does not explicitly model the linear component in its leaves.

Figure 14 illustrates this difference for the 12-month horizon inflation tracking portfolio. The LLF predominantly splits on year-over-year inflation and consumption growth, which our Shapley value analysis (Section 5.2) identified as having significant non-linear effects. The RF, however, shows a more uniform distribution of splits across covariates.

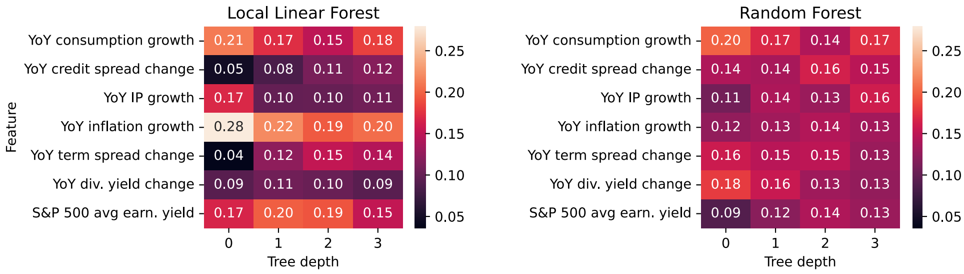


Figure 14: Comparison of LLF and RF split frequencies at each tree depth for the 12-month horizon inflation tracking portfolio. Each column sums to 1, representing relative performance of covariates at each tree depth. Fitted on data from 31 August 1983 to 31 January 2023.

To validate this interpretation, we examine the economic significance of the LLF's most frequent and least frequent splits. Figure 15 shows the split thresholds for the most popular LLF split (year-over-year inflation) and the least popular LLF split (year-over-year term spread change).

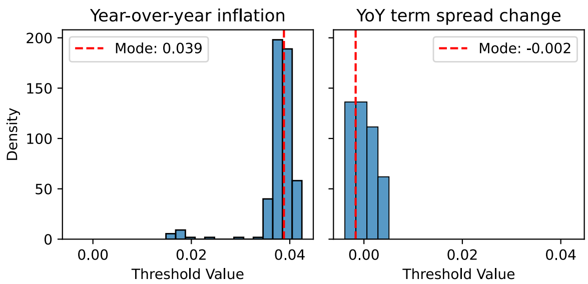


Figure 15: Histogram of split thresholds for the 0-depth (root) splits in the LLF model for the 12-month horizon inflation tracking portfolio. The figure compares the split distribution of thresholds for year-over-year inflation (most frequent split) and term spread change (least frequent split). Fitted on data from 31 August 1983 to 31 January 2023.

We test the significance of these thresholds using an auxiliary regression based on Lamont's regression equation:

$$y_{t+h} = w' R_t + w'_{\text{right-conditional}} R_t \cdot \mathbb{1}(Z_{t-1,j} \geq \tau) + c Z_{t-1} + u_{t,t+h}, \tag{16}$$

where τ is the split threshold and $Z_{t-1,j}$ is the j-th covariate.

For the most popular LLF split (year-over-year inflation), an F-test for $H_0 : w_{\text{right-conditional}} = 0$ yields an F-statistic of 2.518 (p value $= 2.9\%$). This indicates a significant change in the relationship between returns and future inflation when lagged year-over-year inflation passes the 3.9% threshold.

In contrast, for the least popular LLF split (year-over-year term spread change), we find an F-statistic of 0.507 (p value $= 77.1\%$). This suggests no significant non-linear effect for this covariate, explaining why the LLF rarely splits on it.

These results demonstrate that the LLF, by incorporating Lamont's linear covariate adjustment directly in its leaves, can focus its splits on capturing economically significant non-linear effects. The LLF effectively distinguishes between covariates with important non-linear effects (like inflation) and those without (like year-over-year term spread change). The RF, lacking this structure, must use its splits to capture both linear and non-linear covariate effects, leading to a more uniform split distribution.

This analysis highlights a key advantage of the LLF in constructing ETPs: by adhering more closely to Lamont's original framework, it can more effectively capture complex non-linear relationships while still accounting for linear effects of covariates on expected returns. This capability seems to be particularly beneficial for long-term horizon tracking portfolios, where LLFs show superior performance.

6 Conclusion

This study advanced the construction of economic tracking portfolios (ETPs) by integrating machine learning techniques, specifically random forests (RFs) and local linear forests (LLFs), to address the limitations of traditional linear models in capturing complex, non-linear relationships between asset returns and macroeconomic factors. Our findings demonstrate that these machine learning-based approaches consistently outperform linear ETPs across various benchmark indicators and time horizons, offering improved tools for economic forecasting and risk management.

RF-based ETPs show superior performance for short-term (+1 month) horizons across all examined portfolio factors: inflation, consumption growth, and industrial production (IP) growth. The improvement is particularly pronounced for IP growth, where the RF ETP achieves an out-of-sample Mincer-Zarnowitz (MZ) R^2 of 13% compared to 4.7% for the linear ETP. For longer-term (+1 year) horizons, LLF ETPs consistently demonstrate the highest MZ R^2 values, taking advantage of their ability to capture both global non-linear structures and local linear trends.

To assess the statistical significance of these enhanced performances, we conduct one-sided Diebold-Mariano tests. The results reveal that for the one-year horizon, the LLF ETP shows statistically significant improvements over the linear ETP for both inflation and consumption tracking at the 5% significance level. However, for IP growth tracking and all short-term (1-month) horizons, the improvements, while often present, are not statistically significant at the 5% level. This suggests that the enhancements of machine learning-based ETPs are most pronounced for long-term inflation and consumption tracking.

To understand the underlying dynamics of these improved performances, we employ Shapley value analysis, revealing complex relationships and interaction effects between macroeconomic variables and asset allocations. This analysis provides new insights into how economic conditions influence the relative importance of different assets in tracking portfolios. For instance, our findings indicate that the relationship between commodities, credits, and inflation is not static but varies dynamically with economic conditions, challenging the fixed factor decomposition proposed by previous studies.

The inflation tracking portfolio analysis reveals that during periods of low consumption growth and economic stress, such as the 2008 financial crisis, credits become more correlated with inflation than commodities. This shift likely occurs because rising default risks and changing borrowing costs in such periods directly impact credit markets, making them more sensitive to inflation. Conversely, in stable or growing economies, commodities serve as better inflation hedges due to their inherent value and demand stability.

For consumption growth tracking portfolios, our analysis shows significant dependencies on inflation and credit spreads. The LLF ETP tends to increase its allocations to equities, high-yield instruments, and credits (collectively named HEC leverage) during periods of high inflation or increased credit spreads. This behaviour aligns with previous studies that show that such periods signal heightened long-run macroeconomic risk, which in turn increases the sensitivity of these assets to short-term economic conditions like consumption growth.

The application of kernel principal component analysis (KPCA) to the LLF model's kernel weights provides a new approach to identify distinct economic regimes. This technique reveals clusters corresponding to different economic conditions, such as financial crises, stable growth periods, and overheating economies. The clear separation of these clusters and the LLF model's ability to interpolate between regimes demonstrates the LLF model's ability to adapt to different economic conditions effectively.

Despite their superior tracking performance, RF and LLF-based ETPs have certain limitations. While they demonstrate similar robustness to the number of assets included as linear ETPs, they show greater sensitivity to the inclusion of uninformative assets. This highlights the importance of careful asset selection when constructing portfolios using these machine learning techniques. Additionally, the increased transaction costs associated with frequent weight adjustments in RF and LLF-based portfolios present practical challenges for their implementation in hedging scenarios.

This study makes several key contributions to the literature on economic tracking portfolios and the application of machine learning in finance. Firstly, it demonstrates the potential of RF and LLF techniques to enhance the performance of ETPs, providing more accurate tools for economic forecasting and risk management. Second, through Shapley value analysis, it offers new insights into the dynamic relationships between macroeconomic variables and asset allocations, challenging some established views on asset factor decompositions. Third, the application of KPCA to LLF kernels provides a new method for identifying economic regimes based on changing interactions between asset returns and macroeconomic conditions, which could have broad applications in economic research.

Future research could extend this work in several directions. Firstly, incorporating hyperparameter tuning techniques could potentially further improve the out-of-sample performance of RF and LLF-based ETPs. Second, exploring other machine learning techniques, such as gradient boosting or support vector machines, could provide additional insights and performance improvements. Third, incorporating additional regularisation in forest construction could improve the robustness of RF and LLF-based ETPs to uninformative assets.

In conclusion, this study demonstrates the potential of machine learning techniques, particularly RFs and LLFs, in improving the performance of ETPs. By capturing complex relationships between macroeconomic covariates, these models offer improved tools for economic forecasting and risk management. Challenges remain, particularly in terms of robustness to uninformative assets and practical considerations such as high turnover, but the performance improvements and insights provided by these methods show the potential of these methods.

References

Amit, Y. & Geman, D. (1997). Shape quantization and recognition with randomized trees. *Neural computation*, *9*(7), 1545–1588.

Athey, S., Tibshirani, J. & Wager, S. (2019). Generalized random forests. *Annals of Statistics*, *47*(2), 1148–1178. doi: 10.1214/18-AOS1709

Ban, G.-Y., El Karoui, N. & Lim, A. E. (2018). Machine learning and portfolio optimization. *Management Science*, *64*(3), 1136–1154.

Bansal, R. & Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The journal of Finance*, *59*(4), 1481–1509.

Basu, S. (1983). The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence. *Journal of financial economics*, *12*(1), 129–156.

Beber, A., Brandt, M. W. & Kavajecz, K. A. (2009). Flight-to-quality or flight-to-liquidity? Evidence from the euro-area bond market. *The Review of Financial Studies*, *22*(3), 925–957.

Bekaert, G. & Engstrom, E. (2010). Inflation and the stock market: Understanding the "Fed Model". *Journal of Monetary Economics*, *57*(3), 278–294.

Bettencourt, L. O., Tetereva, A. & Petukhina, A. (2024). Advancing Markowitz: Asset Allocation Forest. *Available at SSRN 4781685*.

Breeden, D. T., Gibbons, M. R. & Litzenberger, R. H. (1989). Empirical tests of the consumption-oriented CAPM. *The Journal of Finance*, *44*(2), 231–262.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*, 123–140.

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

Breiman, L., Friedman, J., Olshen, R. & Stone, C. (1984). Cart. *Classification and regression trees*.

Bryzgalova, S., Pelger, M. & Zhu, J. (2020). Forest through the trees: Building cross-sections of stock returns. *Available at SSRN 3493458*.

Campbell, J. Y. (1991). A variance decomposition for stock returns. *The economic journal*, *101*(405), 157–179.

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

Cong, L. W., Feng, G., He, J. & He, X. (2023). Asset pricing with panel tree under global split criteria. *National Bureau of Economic Research*.

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

Fama, E. F. (1981). Stock returns, real activity, inflation, and money. *The American economic review*, *71*(4), 545–565.

Fama, E. F. & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of financial economics*, *22*(1), 3–25.

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

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

Federal Reserve Bank of St. Louis. (2024). *Federal Reserve Economic Data (FRED).* Retrieved from `https://fred.stlouisfed.org` (Accessed: 2024-05-16)

Friedberg, R., Tibshirani, J., Athey, S. & Wager, S. (2020). Local linear forests. *Journal of Computational and Graphical Statistics*, *30*(2), 503–517.

Gilchrist, S. & Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American economic review*, *102*(4), 1692–1720.

Gordon, R. J. (2011). The history of the Phillips curve: Consensus and bifurcation. *Economica*, *78*(309), 10–50.

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

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

Hamilton, J. D. (2009). *Causes and Consequences of the Oil Shock of 2007-08* (Tech. Rep.). National Bureau of Economic Research.

Hamilton, J. D. (2016). Macroeconomic regimes and regime shifts. *Handbook of macroeconomics*, *2*, 163–201.

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

Hothorn, T., Lausen, B., Benner, A. & Radespiel-Tröger, M. (2004). Bagging survival trees. *Statistics in medicine*, *23*(1), 77–91.

Kaminsky, G. L. & Reinhart, C. M. (1999). The twin crises: the causes of banking and balance-of-payments problems. *American economic review*, *89*(3), 473–500.

Keim, D. B. & Stambaugh, R. F. (1986). Predicting returns in the stock and bond markets. *Journal of financial Economics*, *17*(2), 357–390.

Koop, G. & Potter, S. M. (1999). Dynamic asymmetries in us unemployment. *Journal of Business & Economic Statistics*, *17*(3), 298–312.

Lamont, O. A. (2001). Economic tracking portfolios. *Journal of Econometrics*, *105*(1), 161–184.

Lis, E. & Porqueddu, M. (2018). The role of seasonality and outliers in HICP inflation excluding food and energy. *Economic Bulletin Boxes*, *2*.

Lohre, H., Hixon, R. S., Raol, J., Swade, A., Tao, H. & Wolle, S. (2020). Investing through a macro factor lens. *Risk & Reward*, 26–34.

Longin, F. & Solnik, B. (2001). Extreme correlation of international equity markets. *The journal of finance*, *56*(2), 649–676.

Lundberg, S. M. & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Medeiros, M. C., Vasconcelos, G. F., Veiga, Á. & Zilberman, E. (2021). Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *Journal of Business & Economic Statistics*, *39*(1), 98–119.

Meinshausen, N. & Ridgeway, G. (2006). Quantile regression forests. *Journal of machine learning research*, *7*(6).

Mincer, J. A. & Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic forecasts and expectations: Analysis of forecasting behavior and performance* (pp. 3–46). NBER.

Mullainathan, S. & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica: journal of the Econometric Society*, 315–335.

Niskanen, W. A. (1988). *Reaganomics: An insider's account of the policies and the people* (Vol. 292). Oxford University Press New York.

Reinhart, C. M. & Rogoff, K. S. (2009). *This time is different: Eight centuries of financial folly.* Princeton University Press.

Schölkopf, B., Smola, A. & Müller, K.-R. (1997). Kernel principal component analysis. In *International conference on artificial neural networks* (pp. 583–588).

Shapley, L. S. (1953). A Value for n-Person Games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the theory of games ii* (pp. 307–317). Princeton: Princeton University Press.

Stock, J. H. & Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER macroeconomics annual*, *4*, 351–394.

Stock, J. H. & Watson, M. W. (2007). Why has US inflation become harder to forecast? *Journal of Money, Credit and banking*, *39*, 3–33.

Štrumbelj, E. & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems*, *41*, 647–665.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the american Statistical association*, *89*(425), 208–218.

Teräsvirta, T. & Anderson, H. M. (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of applied econometrics*, *7*(S1), S119–S136.

Vassalou, M. (2003). News related to future GDP growth as a risk factor in equity returns. *Journal of financial economics*, *68*(1), 47–73.

Wager, S. & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242.

Zhu, L., Qiu, D., Ergu, D., Ying, C. & Liu, K. (2019). A study on predicting loan default based on the random forest algorithm. *Procedia Computer Science*, *162*, 503–513.

A Data Sources and Descriptive Statistics

Table 3: Data sources and series identifiers for asset classes used in the ETPs.

Asset Class	Data Source	Series Name
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
Commodities	S&P GSCI Commodity TR	S&P GSCI Commodity TR
Gold	Bloomberg Gold Spot Rate	XAUUSD
Oil	Bloomberg WTI Crude Oil Subindex	BCOMCL
Risk-Free	Fama French library	RF

Table 4: Data sources and series identifiers for macroeconomic factors tracked and other covariates used in the ETPs.

Variable	Data Source	Series Name
	Macroeconomic Factors Tracked	
Inflation	FRED	Consumer Price Index for All Urban Consumers: All Items in U.S. City Average (CPIAUCSL)
Consumption Growth	FRED	Real Personal Consumption Expenditures: Nondurable Goods (DNDGRA3M086SBEA) + Services (DSERRA3M086SBEA)
Industrial Production Growth	FRED	Industrial Production Index (INDPRO)
	Other Macroeconomic Covariates	
US Dividend Yield	Quandl	MULTPL/SP500_DIV_YIELD_MONTH
US Earnings Yield	Quandl	MULTPL/SP500_EARNINGS_YIELD_MONTH
Credit Spread	FRED	'GS20' - 'GS30'

Table 5: Pearson correlation coefficients between asset returns and future macroeconomic indicators (1-month and 1-year horizons), August 1983 - January 2023. The upper panel presents correlations with raw assets returns, while the lower panel shows correlations with excess returns (asset returns minus risk-free rate).

	Equity	Bonds	Credits	Comm	HY
+1m inflation	0.14	-0.14	0.07	0.64	0.22
+1y inflation	0.10	-0.06	0.01	0.26	0.05
+1m cons. growth	0.23	-0.04	0.21	0.19	0.22
+1y cons. growth	0.18	0.00	0.10	0.02	0.13
+1m IP growth	0.18	-0.14	0.10	0.30	0.21
+1y IP growth	0.22	-0.08	0.08	0.09	0.23
	Equity minus RF	Bonds minus RF	Credits minus RF	Comm minus RF	HY minus RF
+1m inflation	0.13	-0.16	0.04	0.64	0.20
+1y inflation	0.08	-0.09	-0.03	0.25	0.03
+1m cons. growth	0.23	-0.04	0.21	0.19	0.22
+1y cons. growth	0.18	-0.00	0.09	0.01	0.12
+1m IP growth	0.18	-0.14	0.10	0.30	0.20
+1y IP growth	0.22	-0.09	0.08	0.09	0.23

Table 6: Descriptive statistics of the covariates. Time period is August 1983 to January 2023.

	Min	Max	Med	Avg	SD
YoY term spread change	-0.025	0.032	-0.000	-0.000	0.006
YoY credit spread change	-0.009	0.016	-0.000	-0.000	0.002
YoY div. yield change	-0.004	0.006	-0.000	-0.000	0.001
YoY inflation	-0.020	0.090	0.027	0.028	0.016
YoY consumption growth	-0.116	0.207	0.026	0.025	0.019
YoY IP growth	-0.173	0.162	0.026	0.020	0.043
S&P 500 avg. earn. yield	0.008	0.108	0.050	0.051	0.017

B Hyperparameter Details

The performance of machine learning models, including our RF and LLF ETPs, can be sensitive to the choice of hyperparameters. Table 7 provides a comprehensive list of the hyperparameters used in our RF and LLF models, along with their descriptions and chosen values. These parameters were used consistently across all models to ensure fair comparison and to avoid potential look-ahead bias that could arise from extensive hyperparameter tuning.

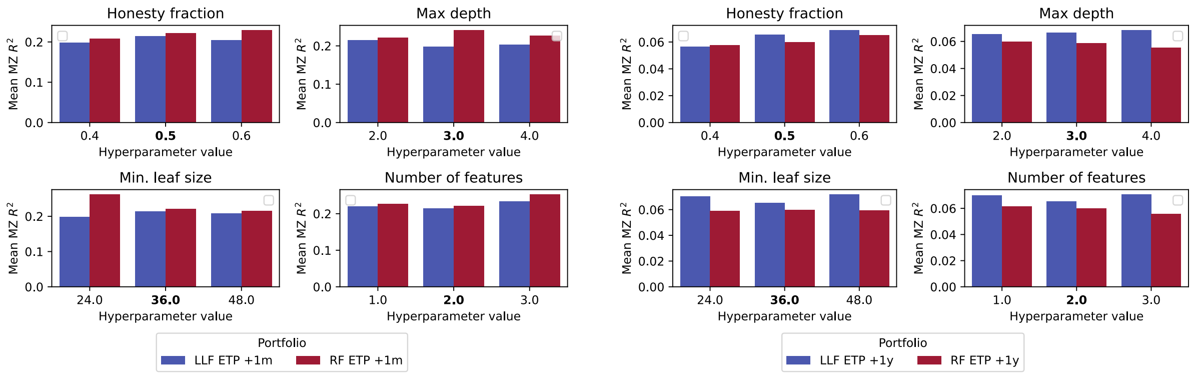
Table 7: Hyperparameters, descriptions, and corresponding values for the RF and LLF ETPs.

Hyperparameter	Description	(Recommended) Value
Number of Trees (B)	Number of trees in the ensemble	200
Honesty Fraction	Fraction of data used for honest estimation	0.5
Minimum Observations per Leaf	Minimum number of samples required to be in a leaf node	36
Maximum Depth of Trees	Maximum depth of the individual trees	3
Inner regression L_2 penalty (λ)	Controls the trade-off between tracking error minimisation and portfolio weight magnitudes	5-fold cross-validation per leaf
Covariate Subset Size	Number of covariates to consider when looking for the best split	$\left\lfloor \sqrt{\text{number of covariates}} \right\rfloor = 2$

C Robustness Check: Hyperparameter Sensitivity

Ideally, annual expanding window hyperparameter optimisation would be conducted through cross-validation, ensuring the hyperparameters are always tuned on the data available at that time. However, due to computation limitations, this approach is impractical. Instead, we chose a fixed set of hyperparameters recommended by prior studies, avoiding any tuning to prevent look-ahead bias (Appendix B).

This makes it important to assess whether our chosen hyperparameters are representative of typical RF and LLF performance. To evaluate this, we conducted robustness checks by varying one hyperparameter at a time from our fixed hyperparameter set. We then fit the portfolio using all data available up to 31 December 2009 and evaluate the performance up to 31 January 2023 without retraining. The average MZ R^2 across all three factors was used to assess performance.



Panel A: one month ahead

Panel B: one year ahead

Figure 16: Sensitivity analysis of RF and LLF ETPs to hyperparameter variations. The plots show the average out-of-sample MZ R^2 across all macroeconomic factors when varying individual hyperparameters. Panel A: one month ahead, Panel B: one year ahead, both fitted on 31 August 1983 to 31 December 2009, evaluated on 1 January 2010 to 31 January 2023. The original hyperparameter values are highlighted in bold.

Figure 16 demonstrates that for both one-month and one-year horizons, the average MZ R^2 is relatively insensitive to hyperparameter variations. For the one-year horizon portfolios, the impact on MZ R^2 is minimal, with absolute changes in MZ R^2 amounting to only fractions of a per cent. For the one-month horizon portfolios, the variations are slightly more pronounced but still within a few per cent.

For context, linear ETPs typically perform significantly worse. Table 1 reports a 3.1% MZ R^2 for the 1-year consumption tracking portfolio and 7.4% for the 1-year LLF ETP. Additionally, the MZ R^2 of the 1-month linear growth tracking portfolio is only 4.7% whereas the RF ETP achieves 13%. This indicates that both RF and LLF ETPs consistently outperform their linear counterparts, even with suboptimal hyperparameters.

Finally, the figures suggest that the originally chosen hyperparameters are rarely the optimal ones. This implies that our models' out-of-sample performance can potentially be improved using cross-validation, indicating room for further optimisation.

D Augmented Dickey-Fuller test results

Table 8 presents the test statistics and corresponding p values of the augmented Dickey-Fuller (ADF) test with trend applied to the data. The results indicate that for all covariates, the null hypothesis of non-stationarity is rejected against 5% significance, leading us to conclude that all covariates are stationarity. This finding supports the use of these variables in our models.

E Robustness Check: Covariate Lookbacks

To further validate our findings, we conducted a robustnes check using alternative specifications for the macroeconomic covariates. Instead of year-over-year differences, we use first differences (monthly changes) for inflation, consumption, industrial production, dividend yeild, credit spread and term spread covariates.

Table 8: Results of augmented Dickey-Fuller tests with trend for stationarity of macroeconomic covariates, August 1983 - January 2023. Lower p values indicate stronger evidence against the null hypothesis of non-stationarity.

	ADF test statistic	p value
Year-over-year term spread change	-5.81	0.00^{***}
Year-over-year credit spread change	-13.83	0.00^{***}
Year-over-year average dividend yield change	-5.04	0.00^{***}
Year-over-year inflation	-3.61	0.01^{**}
Year-over-year consumption growth	-6.13	0.00^{***}
Year-over-year IP growth	-5.77	0.00^{***}
S&P 500 average earnings yield	-3.49	0.01^{**}

Notes: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

The results in Table 9 largely reinforce our main findings, with tree-based methods outperforming linear ETPs in most cases. However, an anomaly is observed for 1-month consumption forecasts, where the linear ETP shows superior performance.

Table 9: Performance metrics for inflation, consumption, and growth ETPs. August 31, 2003 to January 31, 2023 with yearly retraining for the RF and LLF ETP and monthly retraining for the linear ETP. Bolded values indicate the higest MZ R^2 for each horizon and economic factor.

		1-month			1-year		
		Lin. ETP	RF ETP	LLF ETP	Lin. ETP	RF ETP	LLF ETP
Infl.	Avg. turnover	0.015	0.039	0.083	0.051	0.167	0.272
	Avg. leverage	0.027	0.013	0.030	0.075	0.037	0.061
	MZ R^2	0.22	**0.37**	0.35	0.043	**0.059**	0.056
	MZ intercept	0.0022	0.0021	0.0020	0.025	0.025	0.025
	MZ slope	2.13	2.79	1.26	1.67	3.08	1.68
Cons.	Avg. turnover	0.016	0.040	0.097	0.034	0.113	0.219
	Avg. leverage	0.029	0.008	0.014	0.074	0.023	0.052
	MZ R^2	**0.092**	0.0073	0.0055	0.043	0.066	**0.061**
	MZ intercept	0.0011	0.0015	0.0016	0.019	0.019	0.019
	MZ slope	4.71	3.08	0.961	3.30	12.1	4.23
Growth	Avg. turnover	0.030	0.091	0.17	0.037	0.467	0.355
	Avg. leverage	0.042	0.017	0.034	0.074	0.150	0.181
	MZ R^2	0.045	**0.126**	0.068	0.064	0.060	**0.065**
	MZ intercept	0.0009	0.0002	0.0002	0.0052	0.0045	0.0045
	MZ slope	3.66	5.86	1.62	8.73	3.28	2.41

Further investigation reveals that this anomaly is primarily driven by the extreme economic conditions during the Covid-19 pandemic. When excluding April, May, and June 2020 from the analysis, the performance of RF and LLF ETPs for 1-month consumption forecasts improved significantly:

Table 10: Performance metrics for the 1-month consumption ETP. Data spans August 31, 2003 to January 31, 2023, excluding the period from April 2020 to June 2020 to account for Covid-19 economic disruptions. The highest MZ R^2 is in bold.

		1-month		
		Lin. ETP	RF ETP	LLF ETP
Cons.	Avg. turnover	0.016	0.040	0.086
	Avg. leverage	0.029	0.008	0.016
	MZ R^2	0.035	0.054	**0.056**
	MZ intercept	0.0015	0.0015	0.0015
	MZ slope	1.31	3.78	1.06

F Diebold-Mariano Test Results

To rigorously assess the significance of the performance improvements by the machine learning-based ETPs, we conduct one-sided Diebold-Mariano (DM) tests. These tests compare the forecasting accuracy of the best-performing machine learning ETP against the linear ETP for each macroeconomic factor and forecast horizon.

Table 11: Diebold-Mariano test results comparing the forecasting accuracy of the best-performing machine learning-based ETP against linear ETPs. The table presents results for 1-month and 1-year forecast horizons across inflation, consumption, and growth factors.

Horizon	Factor	Best Model	Best MZ R^2	Lin. ETP MZ R^2	DM Stat.	p value
1 month	Inflation	RF ETP	0.4089	0.4017	0.1031	0.5412
	Consumpt.	RF ETP	0.0021	0.0199	1.0848	0.8606
	Growth	RF ETP	0.1281	0.0467	-1.1850	0.1183
1 year	Inflation	LLF ETP	0.0685	0.0436	-1.9672	0.0249**
	Consumption	LLF ETP	0.0744	0.0309	-2.0097	0.0225**
	Growth	LLF ETP	0.0705	0.0677	-0.1811	0.4282

Notes: ** indicates significance at the 5% level for a one-sided test.

In this one-sided test, the null hypothesis is that the forecast accuracy of the machine learning-based ETP and the linear ETP are equal. The alternative hypothesis is that the machine learning-based ETP outperforms the linear ETP.

The results in Table 11 demonstrate the statistically significant performance improvements of the LLF ETPs in long-term (1-year) inflation and consumption tracking, with p values below the 5% significance threshold.

G Portfolio Weights Over Time

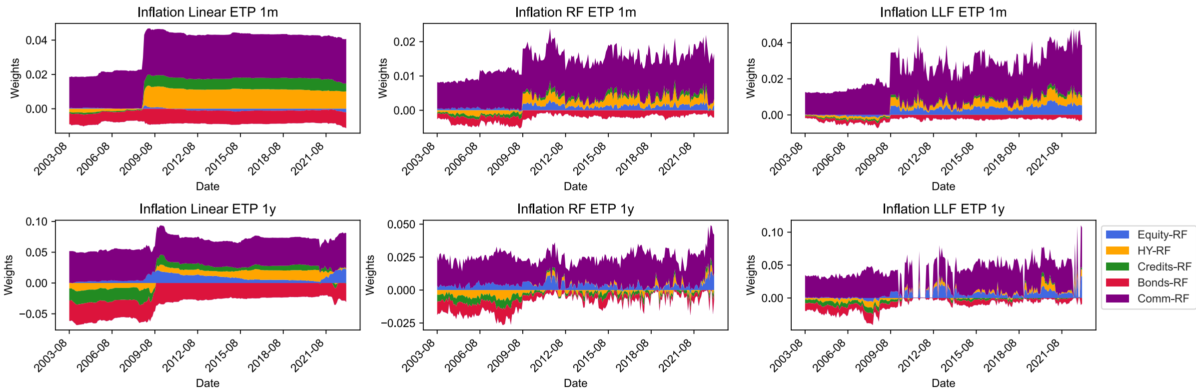


Figure 17: Portfolio weights of inflation tracking portfolios, August 2003 - January 2023 with yearly retraining for the RF and LLF ETPs and monthly retraining for the linear ETP.

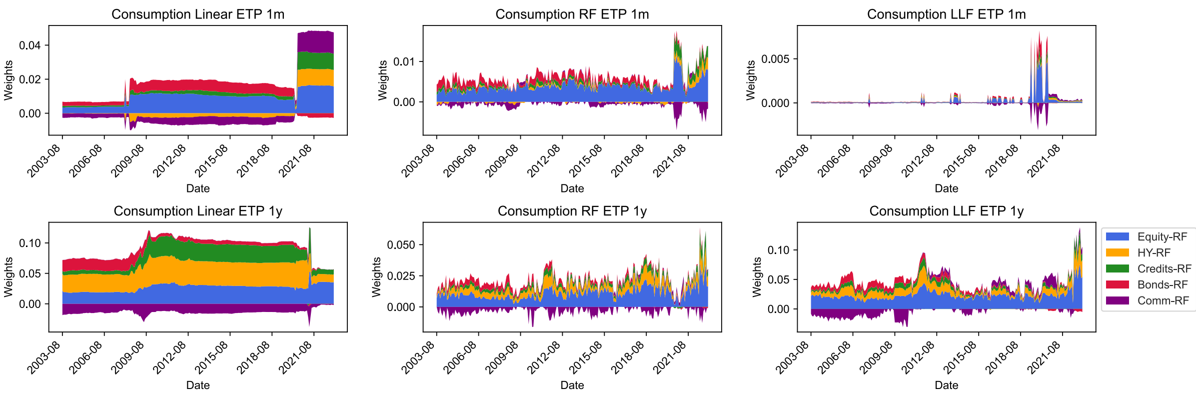


Figure 18: Portfolio weights of consumption tracking portfolios, August 2003 - January 2023 with yearly retraining for the RF and LLF ETPs and monthly retraining for the linear ETP.

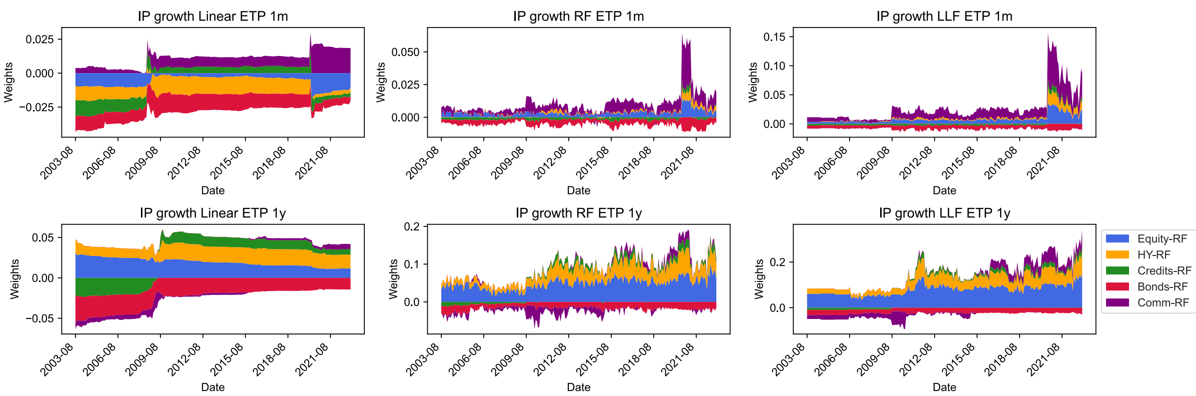


Figure 19: Portfolio weights of IP growth tracking portfolios, August 2003 - January 2023 with yearly retraining for the RF and LLF ETPs and monthly retraining for the linear ETP.

H Commodities, Credits and Inflation during the 2008 Financial Crisis

To provide additional context for the Shapley value analysis in Section 5.2, we examine the changing correlations between inflation, commodities, and credits during the 2008 financial crisis.

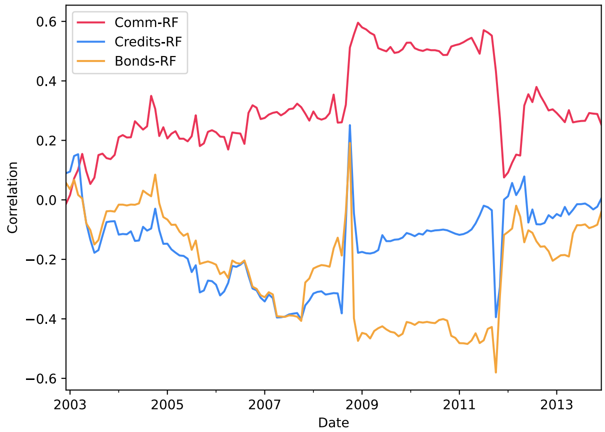


Figure 20: Rolling 3-year correlations of commodities and credits with inflation from January 2001 to January 2014. This period contains significant economic events, including the 2008 financial crisis.

Figure 20 shows a significant increase in the correlation between inflation and credits during the 2008 financial crisis that is higher than the increase in correlation between inflation and commodities. This shift in correlations explains the LLF ETP's tendency to favour credits more during economic downturns, as observed in the Shapley value analysis (Figure 4) and the portfolio weights over time (Figure 21).

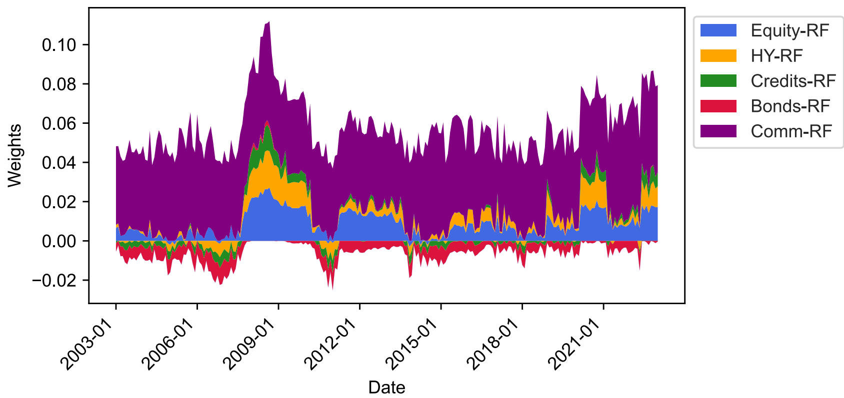


Figure 21: Portfolio weights of LLF ETP from January 2003 to January 2023, model fitted with data until January 2020 which makes the portfolio weights easier to interpret.

I Programming code

The accompanying code repository contains the following components:

- A Python package called `llaaf` that implements core functionalities, including portfolio construction, backtesting, and evaluation methods.
- Jupyter notebooks in the `notebooks/` directory that execute runs and generate plots and tables used in the report.
- Data assets and intermediate output files stored in the `assets/` and `output/` directories, respectively.
- R code for comparisons with the `grf` package for univariate LLFs and data preparation.

The `llaaf` package is structured into modules for data handling, model implementation, plotting, portfolio management, simulation, optimisation, and interpretation. Notebooks are organised to cover data exploration, portfolio backtesting, result analysis, and robustness checks.

For a comprehensive overview of the repository structure, installation instructions, and detailed descriptions of each component, refer to the `README.md` file included in the repository.