Economic Tracking Forests: Leveraging Tree-based Models for Macroeconomic Forecasts

Maurits van Altvorst, supervised by Anastasija Tetereva

Erasmus University Rotterdam

February, 2025

Motivation for Economic Tracking Portfolios

- **Low-frequency macro data:** Many macroeconomic variables (e.g. GDP, inflation) are measured infrequently and with a publication lag.
- **High-frequency market prices:** Stock and asset returns update daily and incorporate market expectations about future macro conditions.
- **Benefits:**
  - *Policy:* Real–time monitoring of economic conditions.
  - *Investors:* Hedging macroeconomic risk.

Lamont (2001) Economic Tracking Portfolios

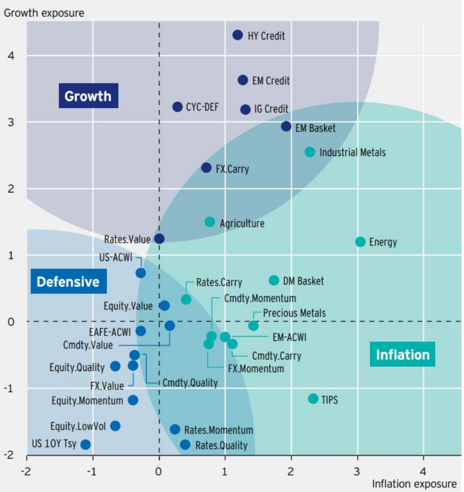
- **Basic idea:** Use asset returns to "track" macroeconomic news.
- Standard regression specification:

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

  where

  - $y_{t,t+h}$: Future macroeconomic target (e.g. inflation over h periods)
  - R_t: Asset returns (e.g. excess returns) from period $t-1$ to t
  - Z_{t-1}: Control variables known at timestamp $t-1$ (e.g. lagged macro factors)
- Outcome: A tracking portfolio that extracts real–time macro forecasts.

Market regimes and optimal portfolios



Lohre, Hixon, Raol, Swade, Tao, Wolle (2020)

Our Contribution

- We extend the linear framework using tree–based models:
  - **Random Forests (RFs)** and **Local Linear Forests (LLFs)**
- These models capture complex, non–linear dependencies between asset returns and macroeconomic factors.
- Our approach allows portfolio weights to adapt to different market regimes while preserving interpretability.

Market conditions-driven tracking portfolios

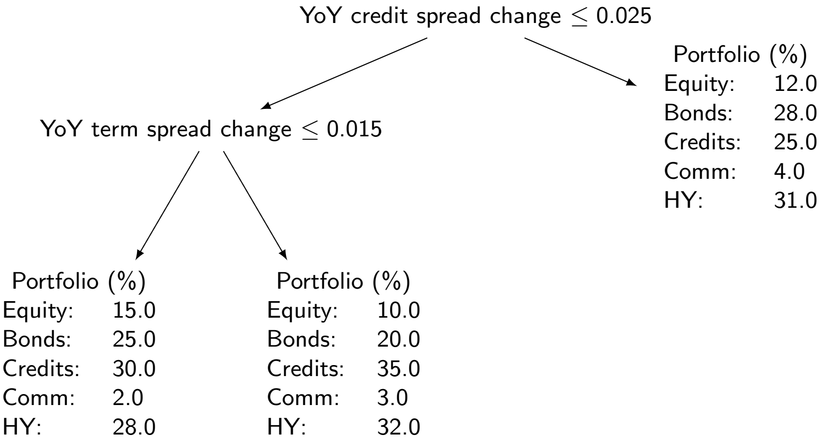


Figure: Asset allocation tree based on year-over-year (YoY) changes in credit spread and term spread, illustrating how different economic conditions lead to distinct portfolio allocations.

Methodology: Tracking LLF

- **Traditional RF:** In a standard random forest, we fit separate linear regressions within each leaf of every tree. This produces a set of local portfolio weights. To predict for new covariates Z_t, we simply take a weighted average of these local weights across all leaves and trees.

- **Local Linear Forest (LLF):** Instead of averaging local regressions, LLF conducts one global weighted least squares regression. The weights in this regression come directly from the structure of the fitted forest (i.e., the adaptive kernel weights), and the regression is inspired by Lamont's original equation.

- This method captures linear relationships between covariates and the target through the global regression, while the forest-derived weights account for non-linear effects.

Data

- August 1983 to January 2023
- Tracking portfolio consists of equities, high-yield instruments, credits, bonds, and commodities
- 7 market conditions variables
  - The year-over-year change in term spread
  - The year-over-year change in credit spread
  - The average S&P 500 dividend yield
  - Year-over-year industrial production growth, consumption growth and inflation
  - The year-over-year change in average S&P 500 earnings yield

Tracking portfolios' performance

		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**
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**
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**

Table: 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.

Consumption-tracking portfolio weights over time

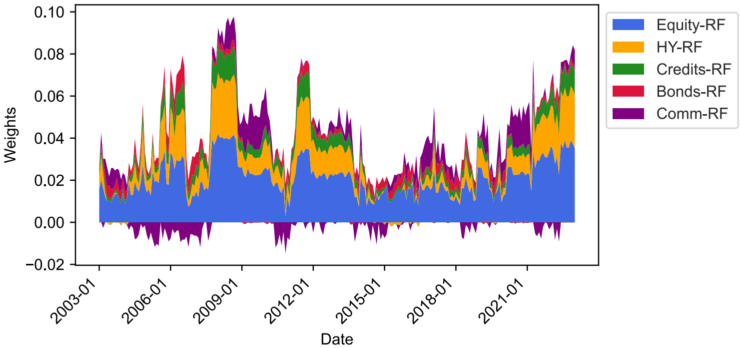


Figure: Portfolio weights of the LLF consumption growth tracking portfolio from January 2003 to January 2023. The model was fitted on data up to January 2020 without yearly retraining, which allows for easier portfolio interpretation.

Introduction to Shapley Values

Shapley values, originating in cooperative game theory, offer a way to explain the prediction of a non-linear model for a given datapoint. The Shapley value for a given feature is the *average marginal contribution* for that feature if we introduce the other features one-by-one.

Example: House Prices

Suppose we predict a house's price using three features: area (m^2), number of bedrooms, and distance to the nearest school.

To determine the contribution of the *area* feature for a specific house, we consider all possible orders in which the features can be added into our predictive model:

- If we first include the number of bedrooms, then add the area, and finally the distance to school, the marginal contribution of the area is the difference in the predicted price when area is added after the number of bedrooms.
- If we first include the distance to the nearest school, then the area, and finally the number of bedrooms, we might get a different marginal contribution.

Shapley values for portfolio metrics

In our application, we compute Shapley values with respect to a portfolio metric: $M(w_t) = \frac{w_{t,\text{comm}} - w_{t,\text{credits}}}{\|w_t\|_1}$. This allows us quantify the average effect of each macroeconomic covariate on portfolio weights for a given timestamp.

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

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

where:

- $N = \{1, 2, \ldots, J\}$ is the set of all covariates.
- Q is any subset of N that does not include j.
- $v(Q)$ is the value function, representing the expected portfolio metric when only the covariates in Q are known:

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

In practice, we approximate these values via Monte Carlo methods by shuffling the values of covariates not in Q and averaging the resulting portfolio metrics.

Inflation-tracking portfolio weights analysis

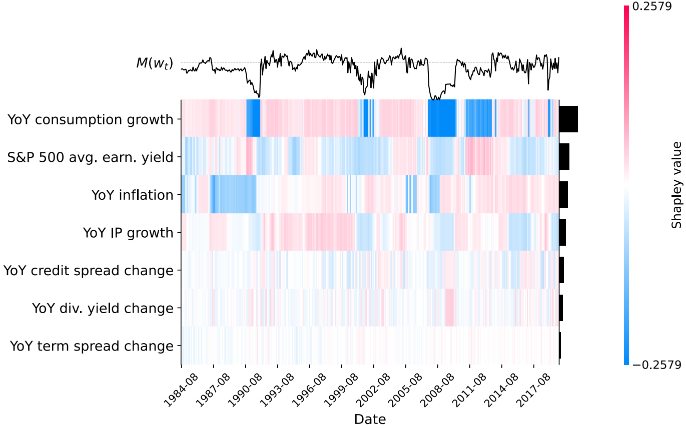


Figure: Heatmap of Shapley values with respect to $M(w_t) = \frac{w_{t,\text{comm}} - w_{t,\text{credits}}}{\|w_t\|_1}$ for the LLF inflation tracking portfolio per covariate over time. Data spans August 1983 to January 2020, colour-coded by Shapley value. Red represents shifts towards commodities and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

Interaction effect between YoY consumption growth and US earnings yield

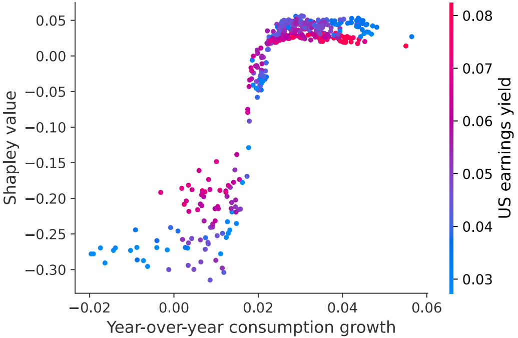


Figure: Shapley values with respect to $M(w_t) = \frac{w_{t,\text{comm}} - w_{t,\text{credits}}}{\|w_t\|_1}$ for the LLF inflation tracking portfolio.

- Non-linear effect: precise value of YoY consumption growth doesn't matter if it's above 2%.
- Interaction effect: a low US earnings yield amplifies the effect of YoY consumption growth (blue points have more extreme Shapley values).

Explanation of interaction effect

- Hamilton (2009) describes that strong global economic growth can make oil prices more responsive to inflationary pressures $\implies$ during periods of high economic growth, commodities can serve as a more effective hedge against inflation. This explains the positive slope.
- Interaction effect:
  - Stress: flight-to-quality phenomenon described by Beber et al. (2009), where investors prefer safer, more liquid assets during times of market stress, explains why commodities become more correlated with inflation when US earnings yield and consumption growth are low.
  - High yield environment: periods where corporate profitability is high relative to stock prices. This potentially indicates a more stable economic environment where the inflation-tracking properties of both commodities and credits are more balanced.

Consumption-tracking portfolio weights over time

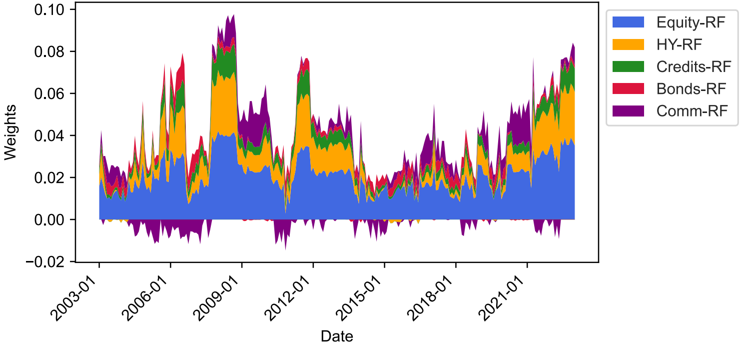


Figure: Portfolio weights of the LLF consumption growth tracking portfolio from January 2003 to January 2023. The model was fitted on data up to January 2020 without yearly retraining, which allows for easier portfolio interpretation.

Consumption-tracking portfolio Shapley values over time

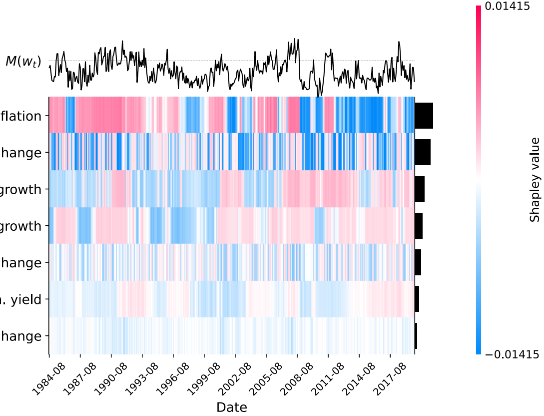


Figure: Heatmap of Shapley values with respect to $M(w_t) = |w_{t,\text{HY}}| + |w_{t,\text{equity}}| + |w_{t,\text{credits}}|$ (HEC leverage) for the LLF consumption tracking portfolio.

Impact of inflation and credit spread change on consumption-tracking portfolio

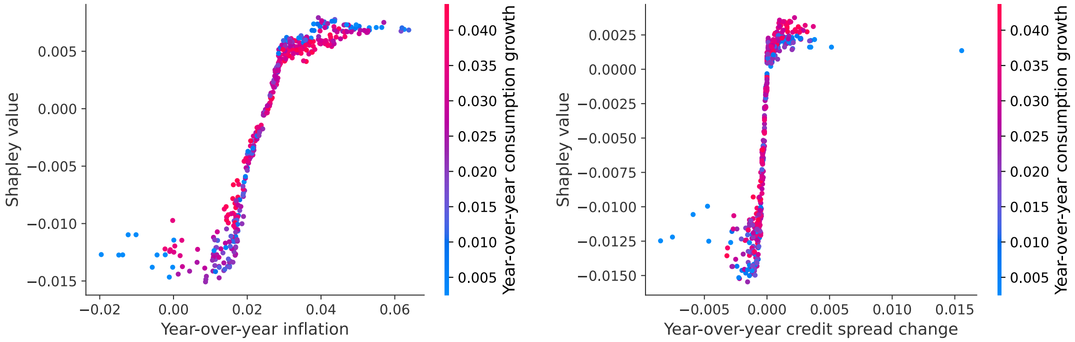


Figure: Shapley values with respect to $M(w_t) = |w_{t,\text{HY}}| + |w_{t,\text{equity}}| + |w_{t,\text{credits}}|$ as a function of year-over-year inflation and year-over-year credit spread change for the local linear consumption tracking portfolio, August 1983 - January 2020. Shapley values indicate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

Conclusion

- ETFs for inflation, consumption growth, and industrial production growth outperform linear ETPs
  - For industrial production growth,the RF ETP achieves an out-of-sample Mincer-Zarnowitz R^2 of 13% compared to 4.7% for the linear ETP
  - For the one-year horizon, the LLF ETP shows statistically significant improvements over the linear ETP for both inflation and consumption

- The relationship between commodities, credits, and inflation is not static
  - For the inflation tracking portfolio, during periods of low consumption growth and economic stress, such as the 2008 financial crisis, credits become more correlated with inflation than commodities
  - For consumption growth tracking portfolios, our analysis shows significant dependencies on inflation and credit spreads