FORECASTING EUROPEAN SOVEREIGN SPREADS USING MACHINE LEARNING

Roland Bouillot, Bertrand Candelon, Clemens Kool

LIDAM Discussion Paper LFIN 2025 / 04

LFIN

Voie du Roman Pays 34, L1.03.01 B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/lfin/publications

Forecasting European Sovereign Spreads using Machine Learning ⋆

Roland Bouillot[a,b], Bertrand Candelon[a], Clemens Kool[b]

[a] *UCLouvain, LIDAM - LFIN, Louvain-La-Neuve, Belgium*
[b] *Maastricht University, SBE - MILE, Maastricht, the Netherlands*

Abstract

Accurate forecasting constitutes a central objective for policymakers. This paper examines the application of advanced machine-learning techniques to predict the 10-year sovereign bond spreads vis-à-vis the German Bund, employing a novel high-dimensional dataset covering 10 European countries over the period 2008 − 2025. An exhaustive comparison of predictive performance, both in-sample and out-of-sample, demonstrates that XGBoost delivers the highest degree of accuracy. Building on these forecast-based spreads, we construct fragmentation matrices that capture the extent of asymmetry across euro area sovereign bond markets. Prior to the COVID-19 crisis, results confirm the well-documented clustering between core and peripheral countries. However, since 2023 this segmentation appears to have weakened, as French and Belgian spreads exhibit a synchronous trajectory. These findings contribute to the literature on financial integration and fragmentation within the euro area, offering new insights into the evolving dynamics of sovereign bond markets.

Keywords: *Machine learning; Financial fragmentation risk; XGBoost; Sovereign spreads.*

⋆The authors thank Iftekhar Hasan for his helpful comments and remarks as well as the participants at EcoSta (2023), the International Symposium on Money, Banking and Finance (2024), the Nederlandse Economendag (2024), the International Workshop on Financial Markets and Nonlinear Dynamics (2025) and the International Conference of the FEBS (2025). The authors also thank the National Bank of Belgium (NBB) for its financial support. Bertrand Candelon has performed this research as part of a research program titled “Financial and Extrafinancial Risk Modeling” under the aegis of the Europlace Institute of Finance, a joint initiative with insti7. The usual disclaimers apply.

Email addresses: `roland.bouillot@maastrichtuniversity.nl` (Roland Bouillot), `bertrand.candelon@uclouvain.be` (Bertrand Candelon), `c.kool@maastrichtuniversity.nl` (Clemens Kool)

1. Introduction

The introduction of a common currency was initially expected to lead to convergence in sovereign borrowing costs across euro-area countries. However, the global financial crisis and the subsequent European sovereign debt crisis exposed deep vulnerabilities in the euro area, revealing how quickly investors can begin to discriminate between member states once concerns over debt sustainability, growth, and institutional credibility intensify. Sovereign yield spreads widened sharply for a subset of countries, while remaining contained for others, and markets stopped treating euro-area sovereigns as close substitutes. This divergence in borrowing costs, commonly referred to as financial fragmentation, is at odds with the notion of a single capital market underpinning a monetary union.

Sovereign bond spreads perform a dual role. They price fiscal and macroeconomic risk at the country level, and they are a central channel through which a single monetary policy is transmitted to heterogeneous member states. When spreads move apart beyond what fundamentals can reasonably justify, fragmentation risk emerges. At one end of the spectrum, stress in high-debt or low-growth sovereigns can spill over to others through "contagion", with investors pricing in the possibility of broader instability. At the other end, investors may reallocate toward perceived safe havens, generating "flight-to-quality" episodes in which some sovereigns experience sharply lower yields. The resulting configuration of spreads reflects heterogeneous risk premia across countries sharing the same currency, and can produce a *de facto* "two-speed" euro area.

Policy makers have responded with a range of instruments aimed at containing fragmentation, most notably the European Central Bank's asset purchase programs and backstop facilities. These interventions have reduced tail risks, but they have not eliminated the underlying forces that drive spreads apart. The COVID-19 shock, the subsequent energy-price surge and the recent tightening cycle have all shown how quickly fragmentation concerns can re-emerge. Central banks and supervisory authorities therefore monitor an expanding set of market-based indicators of integration and segmentation. However, these indicators are largely backward-looking and typically rely on low-dimensional summaries of financial conditions. Much less is known about whether one can use high-dimensional macro–financial information to forecast sovereign spreads to provide near real-time signals about the evolution of the fragmentation risk.

This paper addresses two related questions. First, can modern machine-learning methods, when fed with rich macro–financial data, deliver accurate short-horizon forecasts of euro-area sovereign spreads? Second, can these model-based forecasts be used to construct informative indicators of financial fragmentation, capable of tracking how the cross-country structure of sovereign risk evolves over time? To answer these questions, we assemble a new monthly data set for ten euro-area sovereigns over December 2008 – February 2025, spanning a wide range of macroeconomic, financial and survey variables, and amounting to more than 900,000 country–variable–time observations. We focus on one-month-ahead forecasts of 10-year sovereign yield spreads vis-à-vis the German Bund, which are directly relevant for policy discussions on market stress and transmission.

The paper makes three contributions. First, we conduct a comprehensive horse race among thirteen forecasting models that span linear, nonlinear and rule-based approaches. Forecast accuracy is evaluated strictly out of sample using standard loss functions and tests of equal predictive accuracy. We find that a carefully tuned gradient-boosting specification (XGBoost) consistently lies on the performance frontier and never performs significantly worse than competing models.

Second, we adopt XGBoost as our baseline model and show that it delivers economically small one-month-ahead forecast errors across both core and peripheral sovereigns, including in more volatile issuers. These results indicate that a small class of flexible ensemble methods can provide systematically accurate short-horizon forecasts of euro-area sovereign spreads, and we use these forecasts as the basis for our fragmentation analysis.

Third, we open up the black box of the machine learning model by combining its forecasts with post-estimation diagnostics. SHAP-based decompositions reveal that, at the one-month horizon, lagged domestic spreads are the dominant predictors in every country, pointing to strong short-run persistence in risk premia. Beyond this autoregressive core, the model loads on macro–financial indicators related to financial conditions, prices, global markets and household or monetary sectors, with the relative importance of these drivers differing across the currency union. Correlation heatmaps of predicted spreads then trace how the cross-country structure of sovereign risk evolves through time. Over 2020–2025, the model-implied spreads suggest that the fragmentation dynamics between core and peripheral country is still at work. More worryingly, in more recent windows, French spreads gradually decouple from the main bloc, and by mid-2024–2025 France and Belgium form a distinct cluster, bringing back the fragmentation risk in the frontstage.

Relative to recent fragmentation indicators based on time-varying cointegration relationships, risk-neutral term premia and network-based measures of spillovers, our approach provides a forecast-based, model-driven indicator of fragmentation constructed from machine-learning predictions of sovereign spreads.

The remainder of the paper is organized as follows. Section 2 reviews the literature on euro-area sovereign spreads, financial fragmentation and machine-learning applications in macro–finance. Our novel (big) data set and its macro–financial variables are presented in Section 3. Section 4 presents the forecasting framework, the horse-race set up and results as well as the SHAP-based interpretation of country-specific drivers. Section 5 uses correlation heatmaps of predicted spreads to analyze the evolving clustering of sovereigns and to discuss implications for fragmentation. Finally, section 6 concludes with the main findings, policy implications and avenues for future research.

2. Literature Review

2.1. Determinants of Euro Area Sovereign Spreads and the Emergence of Fragmentation

Sovereign bond spreads in the euro area reflect compensation for a bundle of risks: fiscal and macroeconomic fundamentals, liquidity and market-risk premia, and institutional and political factors. Contributions in this field of literature document that higher debt and deficit ratios, weak growth and external imbalances are systematically associated with wider spreads, especially once markets reassessed sovereign credit risk after 2007 (Attinasi et al. (2010); Maltritz (2012); Afonso et al. (2015)). Fiscal news and revisions to budget projections are priced quickly, while global risk aversion, proxied by indicators such as the VIX or corporate credit spreads, amplifies the sensitivity of vulnerable issuers to shocks. Liquidity and market depth also matter: issues with smaller free float or thin secondary markets carry higher premia, particularly in periods of stress (Barbosa and Costa (2010)).

Beyond fundamentals and liquidity, spreads embed information about contagion, the bank–sovereign nexus and tail risks. During the crisis, sovereign and bank balance sheets became tightly intertwined and bank CDS premia and sovereign spreads mutually reinforced each other (Gerlach et al. (2010); Stanga (2011)). Work on contagion and "wake-up-call" effects shows that spreads react not only to domestic news but also to shocks to other sovereigns, with time-varying spillovers that are stronger for fiscally weak or highly leveraged countries (Giordano et al. (2013); Beirne and Fratzscher (2013); Favero (2013)). Redenomination risk emerged as an additional component of spreads: markets started to price the possibility that some member states might exit the currency union and redenominate their debt, particularly at the height of the crisis (De Santis (2012); Klose (2013)).

The run-up to the European sovereign debt (ESD) crisis is often viewed as a period in which sovereign risk premia were compressed relative to underlying fundamentals. In the first decade of EMU, long-term yields converged to German Bund levels and spreads remained small despite growing heterogeneity in fiscal positions, competitiveness and external balances. Empirical work shows that before 2008 spreads were weakly related to fundamentals, consistent with complacency and an underestimation of sovereign credit risk (Favero and Missale (2012); Maltritz (2012)). The global financial crisis of 2008–2009 acted as a wake-up call: risk premia began to differentiate more strongly across countries and investors started to reprice deficits, debt stocks and current-account positions.

The ESD crisis itself was triggered when Greece revised its deficit figures in late 2009–2010, leading to a rapid loss of market access and a sharp repricing of default risk. Subsequent research distinguishes between countries with clear solvency problems (e.g. Greece) and those primarily facing liquidity crises exacerbated by self-fulfilling expectations (De Grauwe (2012); Zoli (2009)). Spreads in Italy, Spain and Portugal rose to levels that implied default probabilities difficult to reconcile with medium-term fundamentals. Contagion operated both through trade and financial linkages and via investors' reassessment of the entire periphery as a cluster of correlated risks. At the same time, a pronounced flight-to-quality into German Bunds and, to a lesser extent, other core issuers widened cross-country differentials beyond what fundamentals alone would explain (Beber et al. (2009); Ehrmann and Fratzscher (2015)).

The policy response reshaped the mapping from fundamentals to spreads. Initial *ad hoc* packages for Greece were followed by the creation of the European Financial Stability Facility (EFSF) and later the European Stability Mechanism (ESM), but the decisive break in market dynamics arrived only with the ECB's "whatever it takes" commitment and the announcement of Outright Monetary Transactions (OMT). Empirical studies show that OMT, the Securities Markets Programme (SMP) and later large-scale asset purchase programs compressed term premia, reduced fragmentation and weakened the immediate pass-through from fiscal variables to spreads, particularly in stressed countries (Altavilla et al. (2015); Grandi and Taboga (2019); Moessner (2018)). At the same time, the crisis made clear that, in a monetary union without a common fiscal backstop, sovereign spreads also price institutional design and the credibility of the policy regime.

Post-crisis evidence indicates that fragmentation risk has not disappeared. Spreads remain more sensitive to fiscal and macro imbalances in high-debt countries, and stress episodes such as the COVID-19 shock or renewed geopolitical tensions have produced recurrent, if temporary, flare-ups in cross-country differentials (Candelon et al. (2022); Gilbert (2019)). Recent work proposes formal fragmentation indicators based on time-varying cointegration, risk-neutral term premia and network-based measures of spillovers, all of which show that the euro sovereign market oscillates between phases of integration and renewed segmentation. Overall, the literature suggests that euro area spreads today reflect a layered structure of risks: country-specific fiscal and macro conditions; global and euro-wide risk factors; redenomination and policy-regime risk; and evolving expectations about the ECB's willingness to contain disorderly market dynamics. Capturing how these components interact over time and across a large set of macro-financial variables is difficult with low-dimensional linear models and motivates the turn to more flexible machine learning approaches.

2.2. Machine learning Approaches to Sovereign Risk and Fragmentation

The macro–financial environment that underlies euro area sovereign spreads is high-dimensional and non-linear. Spreads react to a broad set of fiscal, real, financial and political indicators, as well as to global conditions and uncertainty measures. Traditional econometric approaches, such as linear panel regressions, factor models or Bayesian model averaging, impose strong parametric structure and are quickly strained when the cross-section of predictors becomes large, giving rise to the usual curse of dimensionality. In contrast, modern machine-learning methods combine regularization, flexible functional forms and systematic cross-validation to handle large predictor sets while prioritizing out-of-sample performance.

In asset pricing, this shift is already well established. Gu et al. (2020) show that tree-based methods and neural networks substantially improve the prediction of equity risk premia relative to leading linear factor models, with large economic gains to investors and gains traced to the ability to capture interactions and nonlinear effects. Similar results emerge in other contexts where returns or risk premia depend on many correlated predictors. Macroeconomic forecasting provides complementary evidence. Using a large U.S. macro–financial data set, Coulombe et al. (2022) decompose the sources of machine-learning gains and show that non-linear methods deliver systematically better forecasts than standard time-series and factor models, especially during periods of high uncertainty and financial stress. This pattern is consistent with the idea that standard linear approximations underperform precisely when nonlinearities and regime shifts matter most—the same environments in which fragmentation risk in sovereign bond markets is likely to be priced.

A rapidly growing literature applies machine learning to early-warning systems and crisis prediction. Liu et al. (2022) compare logistic regressions with several tree-based and ensemble algorithms in a global sample and find that random forests, gradient boosting and related models deliver more accurate early warnings of financial crises, while Shapley-value decompositions help identify the macro–financial indicators driving the predictions. Other contributions, such as Casabianca et al. (2019), use boosting methods to rank the determinants of banking crises and document that machine learning both outperforms logit models and reorders the relative importance of macro–financial indicators over time. Related work by Tölö (2020) shows that recurrent neural networks (RNN-LSTM and RNN-GRU) can further improve systemic-crisis prediction relative to simpler neural nets and logistic benchmarks.

At the same time, the literature is explicit that machine learning is not a free lunch. Beutel et al. (2019) demonstrate that in some early-warning settings, well-specified logit models match or even outperform more complex ML algorithms in recursive out-of-sample tests, highlighting the risk of overfitting and the importance of careful model validation. Recent studies respond by combining machine learning with rigorous forecast evaluation and explainability tools, showing that the gains from flexible methods are largest when models are tuned to policy-relevant loss functions and when the underlying signal-to-noise ratio is sufficiently high (e.g. Reimann, 2024; Bluwstein et al., 2023).

Applications to sovereign risk and euro area fragmentation are emerging. Arakelian et al. (2018) use recursive partitioning and random forests to stratify European sovereigns into risk zones based on fundamentals and contagion measures, documenting nonlinear threshold effects and time-varying sensitivity to global risk factors during and after the ESD crisis. Belly et al. (2023) compare several machine-learning algorithms, including XGBoost, support vector regression, elastic net, random forests and neural networks, to Bayesian model averaging in pricing euro area sovereign risk and show that machine-learning models better capture the joint influence of macro fundamentals, global financial conditions and public sentiment on CDS and bond spreads.

The evidence from asset pricing, macroeconomic forecasting and crisis prediction therefore suggests that machine learning is a natural candidate for modeling euro area sovereign spreads. In this paper, we build on this literature in two ways. First, we conduct a horse race among thirteen machine-learning algorithms to identify the techniques that deliver the most accurate forecasts of 10-year sovereign spreads for ten euro area countries. Second, focusing on the best-performing model, an XGBoost specification, we generate one-month-ahead spread forecasts during the January 2020 to February 2025 period and complement those forecasts with model-agnostic explainability tools to recover the key drivers of predicted spreads over time.

3. Data

In this paper, we assemble a new high-dimensional monthly macro–financial data set for ten euro area countries (Austria, Belgium, Finland, France, Greece, Ireland, Italy, the Netherlands, Portugal and Spain) over the period from December 2008 to February 2025. The data are drawn from the Bloomberg database and cover a broad spectrum of macroeconomic conditions, sectoral activity, financial markets and uncertainty.

Table 1 reports the number of series and observations per country. In total, the data set contains $4,948$ country–specific time series and $962,307$ individual observations. The breadth of coverage is a main reason for the use of high-dimensional prediction methods.

Table 1: Variable and observation count

Country	Number of variables	Number of observations
Austria	255	49,725
Belgium	637	123,537
Finland	510	99,184
France	1,278	247,723
Greece	397	77,415
Ireland	393	76,585
Italy	394	76,758
Netherlands	275	53,625
Portugal	292	56,940
Spain	517	100,815
Total	4,948	962,307

The $4,948$ series fall into 17 thematic blocks summarized in Table 2. Country-level variables span standard macroeconomic aggregates (national accounts, labor markets, price levels, government finances), sectoral indicators (industry, services, retail, housing and real estate, household balance sheets) and measures of whole-economy activity. These are complemented by financial variables (sovereign yields, credit activity, equity indices). Global market indices and proxies for macro-financial uncertainty are also added to the data set, to account for the external factors affecting the country under study's spread. The resulting data set provides a granular view of domestic conditions and their interaction with global factors at a monthly frequency.

The 10-year sovereign spread is a prevalent metric among market participants and academics for evaluating sovereign risk. We build our 10-year sovereign spread as the difference between a country's 10-year sovereign yield and the risk-free rate. For this purpose, we adopt the German 10-year bond yield as the risk-free rate, reflecting its widespread acceptance as the standard measure in both market practice and academic research for calculating euro area spreads.

The vastness and granularity of this data set not only provide a comprehensive view of each country's economy but also present unique challenges and opportunities for data processing, analysis, and modeling. By leveraging the data set's potential, this paper identifies intricate relationships between countries' idiosyncratic characteristics and their yield spreads, potentially producing more accurate and precise forecasts. Hence, the introduction of this novel data set represents a notable contribution to the field, enabling a more refined and comprehensive understanding of yield spread dynamics across euro area countries.

Table 2: Categories of variables for each country

National Accounts	Services Sector	Whole Economy Activity
Surveys & Cyclical Indicators	International Trade	Monetary Sector
Labor Market	Price Indexes	Government Finance & Debt
Retail Sector	Housing & Real Estate	Financial Indicators
Industrial Sector	Household Sector	Demographics
Global Markets Indexes	Uncertainty Measures	

Tables 3 and 4 report descriptive statistics for sovereign bond yields and bond spreads in eleven and ten euro area countries respectively. For each country, the mean, median, standard deviation, variance, range, inter-quartile range (IQR), skewness, kurtosis, minimum, and maximum values are reported. This subset of data consists of 195 monthly observations per country, spanning from December 2008 to February 2025, with all figures rounded to two decimals.

Table 3 presents the descriptive statistics for sovereign bond yields across eleven euro area countries. On average, 10-year yields display pronounced cross-country heterogeneity, with core countries exhibiting much lower and more stable levels than peripheral issuers. Mean yields for core sovereigns

(Germany, Austria, France, Finland and the Netherlands) cluster between 1.1 and 1.8 percentage points, while high-debt sovereigns such as Italy, Portugal and Spain exhibit mean yields between roughly 2.7 and 3.8 percentage points. Greece stands out with a mean yield of 7.5 percentage points and a range exceeding 30 percentage points, reflecting episodes of severe market stress during and after the ESD. Yield volatility, as measured by the standard deviation, is below 1.6 percentage points for core issuers and rises to 3.1 percentage points for Portugal and 6.4 percentage points for Greece. Inter-quartile ranges display a similar ranking, indicating more dispersed yield distributions in the periphery. Skewness is positive for all countries, and kurtosis is particularly elevated for Greece and Portugal, signaling thick tails and frequent extreme yield realizations.

Table 3: Descriptive statistics for euro area sovereign bond yields

Country	Mean	Median	Std. Dev	Variance	Range	IQR	Skew.	Kurt.	Min	Max
Austria	1.54	1.14	1.38	1.92	4.75	2.50	0.31	-1.27	-0.45	4.31
Belgium	1.76	1.22	1.49	2.21	5.36	2.56	0.28	-1.33	-0.39	4.97
Finland	1.43	1.05	1.31	1.71	4.51	2.30	0.36	-1.23	-0.43	4.08
France	1.60	1.27	1.30	1.68	4.36	2.32	0.17	-1.41	-0.41	3.95
Germany	1.15	0.86	1.22	1.49	4.29	1.98	0.33	-1.15	-0.70	3.59
Greece	7.49	5.75	6.41	41.14	32.01	5.68	1.89	4.09	0.60	32.60
Ireland	2.92	1.64	2.98	8.89	11.76	4.05	1.02	-0.05	-0.31	11.45
Italy	3.03	2.90	1.56	2.44	6.49	2.65	0.20	-1.00	0.54	7.03
Netherlands	1.39	1.07	1.31	1.71	4.56	2.08	0.31	-1.19	-0.55	4.00
Portugal	3.79	3.11	3.12	9.72	15.64	2.98	1.34	1.72	0.03	15.67
Spain	2.66	2.12	1.81	3.27	6.79	2.86	0.36	-1.06	0.04	6.83

Notes: Bond yields across 11 euro area countries under study, based on 195 monthly observations per country from December 2008 to February 2025.

Table 4: Descriptive statistics for euro area sovereign bond spreads

Country	Mean	Median	Std. Dev	Variance	Range	IQR	Skew.	Kurt.	Min	Max
Austria	0.40	0.35	0.23	0.05	1.11	0.30	1.25	1.39	0.07	1.18
Belgium	0.61	0.50	0.42	0.17	2.51	0.35	2.24	6.01	0.18	2.69
Finland	0.29	0.25	0.17	0.03	0.80	0.23	0.78	0.00	-0.01	0.79
France	0.47	0.39	0.21	0.05	1.11	0.24	1.72	3.41	0.21	1.31
Greece	5.98	3.80	6.04	36.45	29.97	6.44	2.10	4.95	0.82	30.79
Ireland	1.48	0.61	1.81	3.27	8.22	1.39	2.01	3.29	0.21	8.43
Italy	1.84	1.58	0.86	0.74	4.45	0.85	1.60	2.58	0.76	5.20
Netherlands	0.24	0.22	0.13	0.02	0.77	0.18	1.34	3.01	0.04	0.81
Portugal	2.50	1.49	2.54	6.48	13.48	2.43	2.01	4.04	0.40	13.88
Spain	1.46	1.08	1.03	1.05	4.98	0.69	1.95	3.42	0.52	5.50

Notes: Bond spreads across 10 euro area countries under study, based on 195 monthly observations per country from December 2008 to February 2025. Germany is used as the benchmark for each other country.

The descriptive statistics in Table 4 offer insights into the properties of the sovereign bond spreads of each euro area country under study relative to Germany, which serves as the benchmark country (risk-free). Spreads over German Bunds display again the expected core–periphery pattern. Average spreads for core issuers lie between roughly 25 and 60 basis points, whereas mean spreads range from about 1.5 to 2.5 percentage points for Ireland, Italy, Portugal and Spain, and reach almost 6 percentage points for Greece. Volatility is again concentrated in the periphery: the standard deviation of spreads is below 0.25 percentage points in Finland and the Netherlands, but exceeds 1 percentage point in Ireland, Italy, Portugal and Spain and reaches 6 percentage points in Greece. Ranges and IQRs tell a similar story, with wide dispersion for Greece, Portugal and Ireland and comparatively tight distributions for core countries. All spreads exhibit positive skewness, and kurtosis is particularly high for Belgium, Greece and Portugal, indicative of recurrent episodes of sharply widening premia.

These descriptive statistics highlight the substantial heterogeneity in bond yields and spreads behaviors across euro area countries with distinct contrasts in average levels, volatility and distributional characteristics. These cross-country differences in the level, dispersion and tail behaviour of spreads point to heterogeneous risk pricing within the currency union. Higher average spreads and more volatile, skewed distributions in Greece, Italy, Portugal and Spain are consistent with persistently higher perceived credit risk, whereas the lower and more tightly distributed spreads in Austria, Finland, France and the Netherlands reflect the safer status of core sovereigns. Ireland occupies an intermediate position, with characteristics that lie between the core and southern peripheral issuers. Our high-dimensional macro-financial data set is designed precisely to capture the drivers of this heterogeneity and to allow machine learning models to link the evolution of spreads to a rich set of domestic and global predictors.

4. Comparing Machine Learning techniques : a Horse Race

In this section, we begin by detailing each of the 13 machine learning techniques included in our horse race. Subsequently, we present the results of this horse race.

4.1. Overview of the machine learning techniques

In order to study financial fragmentation, we exploit our novel high-dimensional data set. In this environment, conventional econometric models quickly face the "curse of dimensionality" and specification constraints, whereas modern machine-learning methods are explicitly designed to handle many correlated predictors and to prioritize out-of-sample performance. This is why we employ machine learning. Indeed, machine learning offers a plethora of techniques for variable selection, aiding in the identification of the most informative predictors in high-dimensional data sets. In this paper, we perform a horse race of linear, non-linear and rule-based techniques to find which machine learning methodology best fits our data. The 13 machine learning techniques competing in this horse race are : (1) Standard LASSO, (2) Group LASSO, (3) Sparse-Group LASSO, (4) Ridge regression, (5) Elastic Net regression, (6) Support Vector Regression, (7) Feedforward Neural Network, (8) Decision Trees, (9) Regularized Trees, (10) Boosted Trees, (11) Stochastic Gradient Boosting, (12) Extreme Gradient Boosting and (13) LightGBM.

For each country, predictors are standardized to zero mean and unit variance and the sample is split into an 80% training subsample and a 20% hold-out test subsample. Hyperparameters are selected by K-fold cross-validation on the training set only, and out-of-sample performance on the test set is evaluated using the root mean squared error (RMSE), mean absolute error (MAE) and coefficient of determination R^2. Table 5 provides an overview of the selected models, the validation procedures and the evaluation metrics.

4.2. Linear techniques

The group of linear techniques encompasses the LASSO, the Ridge and the Elastic Net regressions. Before examining each linear technique more in detail, it is worth noting that they suffer from scale dependency. The algorithm's effectiveness in selecting variables and assigning coefficients is influenced by the scale of the predictor variables. Linear techniques penalize the absolute size of coefficients, shrinking less important coefficients toward zero. However, the magnitude of each variable's coefficient depends on the variable's scale. For example, in a data set with two predictors, one measured in thousands and the other in single units, LASSO would penalize the larger-scaled variable more heavily because its raw coefficient value would typically be larger, regardless of its relative importance. This imbalance can lead LASSO to shrink coefficients unevenly, selecting or discarding variables based on their scale rather than their true contribution to the model. To mitigate this effect, we standardize the predictor variables (mean of zero and standard deviation of one). By standardizing our data, we ensure that each variable's scale is comparable, allowing LASSO to more accurately apply penalties based on variable importance rather than scale differences.
The linear benchmark (LASSO) is a regularized regression of the 10-year spread on the standardized predictor set. Let y_i denote the spread at time i, x_i the p-dimensional predictor vector and β the coefficient vector. Penalized least squares solves

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \left(y_i - x_i'\beta\right)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j|, \tag{1}$$

which shrinks uninformative coefficients exactly to zero. Group LASSO extends this idea when predictors are partitioned into G groups $\{\mathcal{G}_g\}_{g=1}^{G}$, by penalizing the ℓ_2-norm of group coefficients, $\sum_g \|\beta_{\mathcal{G}_g}\|_2$, inducing sparsity at the group level. Sparse-group LASSO combines both penalties to allow for simultaneous group and within-group selection.

Table 5: Machine learning techniques, validation procedures and accuracy metrics

Machine learning technique	Nb	Validation, hyperparameter tuning and evaluation metrics
Linear		
Standard LASSO	1	10-fold cross-validation on the training data to select the penalty parameter λ.
Group LASSO	2	10-fold cross-validation on the training data to choose the group penalty parameter.
Sparse-group LASSO	3	10-fold cross-validated sparse-group LASSO on the training data to select λ along the group path.
Ridge	4	10-fold cross-validation on the training data to select λ (ridge penalty).
Elastic Net	5	10-fold cross-validation with mixing parameter $\alpha = 0.5$ on the training data to choose λ.
Non-linear		
Support Vector Regression (SVR)	6	ε-SVR with default radial kernel and tuning parameters estimated on the training data; predictions are computed on the test set.
Feedforward Neural Network (FNN)	7	Two hidden layers FNN (10 and 5 units) trained on the training data; out-of-sample predictions are obtained on the test set.
Rule-based		
Trees		
Decision Trees	8	Regression tree estimated by cost-complexity pruning on the training data; performance is evaluated on the test set.
Bagged Trees (Random Forest)	9	Random forest with 3,000 trees fitted on the training data, using out-of-bag error for stabilization; evaluated on the test set.
Boosted Trees	10	Gradient-boosted regression trees (3,000 base learners, depth 6, shrinkage 0.05) with 5-fold cross-validation on the training data to select the effective number of trees; the full procedure is repeated 50 times and performance averaged.
Stochastic Gradient Machines		
SG Boosting	11	Stochastic gradient boosting with bag fraction 0.7, depth 4 and 5-fold cross-validation for tree number; the train–validate–test cycle is repeated 50 times and performance averaged.
XGBoost	12	Random search over a hyperparameter grid with 10-fold cross-validation and early stopping on the training data; the configuration is re-estimated 50 times on the full training set and test performance averaged.
LightGBM	13	LightGBM with standard regression objective; in each of 50 runs, 10-fold cross-validation with early stopping selects the number of iterations before training on the full training data and evaluating on the test set.

Notes: All models are estimated on an 80% training subsample and evaluated on a 20% hold-out test subsample. Unless otherwise stated, cross-validation on the training data uses squared-error loss to select regularization strengths or the number of trees/iterations. Out-of-sample accuracy is measured by Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and the coefficient of determination R^2 on the test set; for Monte Carlo procedures, we report averages over 50 repetitions.

Ridge regression replaces the ℓ_1 penalty by an ℓ_2 penalty

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \left(y_i - x_i'\beta\right)^2 + \lambda_2 \sum_{j=1}^{p} \beta_j^2, \tag{2}$$

which shrinks coefficients towards zero but typically retains all predictors, a desirable property when many predictors are correlated and potentially relevant.

Elastic Net nests LASSO and ridge by combining both penalties,

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \left(y_i - x_i'\beta\right)^2 + \lambda \left[\alpha \sum_{j=1}^{p} |\beta_j| + (1-\alpha) \sum_{j=1}^{p} \beta_j^2\right], \tag{3}$$

with mixing parameter $\alpha \in [0,1]$. In our baseline specification we set $\alpha = 0.5$, giving equal weight to sparsity and shrinkage.

4.3. Non-linear techniques

Linear specifications restrict the conditional mean of spreads to be an affine function of predictors. To allow for richer functional forms we consider Support Vector Regression (SVR) and a feedforward neural network (FNN).

SVR fits a function $f(x) = w'\phi(x) + b$ that is as flat as possible while keeping prediction errors within an ε-insensitive tube whenever feasible. In its primal form, SVR solves

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \tag{4}$$

subject to

$$|y_i - w'\phi(x_i) - b| \le \varepsilon + \xi_i, \qquad \xi_i, \xi_i^* \ge 0,\ i = 1, \ldots, n, \tag{5}$$

where C controls the trade-off between flatness and violations of the tube and $\phi(\cdot)$ is an implicit feature mapping induced by a kernel $K(x_i, x_j) = \phi(x_i)'\phi(x_j)$. We use a Gaussian kernel and tune C and ε via cross-validation.

The FNN is a flexible non-linear approximator composed of an input layer, two hidden layers and a linear output layer. Let $\mathbf{a}^{(0)} = x_i$ denote the standardized inputs and, for layer l,

$$\mathbf{z}^{(l)} = \mathbf{W}^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}, \qquad \mathbf{a}^{(l)} = \tanh\left(\mathbf{z}^{(l)}\right), \tag{6}$$

with $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ the weight matrix and bias vector. The output layer is linear,

$$\hat{y}_i = \mathbf{W}^{(L)}\mathbf{a}^{(L-1)} + \mathbf{b}^{(L)}, \tag{7}$$

so that the network can produce spreads on the original scale. Parameters are estimated by minimizing the mean squared error

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{8}$$

using stochastic gradient descent with backpropagation.

4.4. Rule-based techniques

Rule-based techniques infer a set of if–then rules from the data, typically implemented by decision trees and their ensembles. They are well suited to capturing interactions and non-linearities and offer a natural route to model explainability.

4.4.1. Tree-based models

A regression tree partitions the predictor space recursively into disjoint regions and assigns a constant prediction to each terminal node. At a candidate split $x_k \leq s$, the algorithm chooses the threshold that minimizes the weighted within-node variance of the target variable,

$$\mathrm{MSE}(s,k) = \frac{|D_{\mathrm{L}}|}{|D|}\mathrm{Var}(D_{\mathrm{L}}) + \frac{|D_{\mathrm{R}}|}{|D|}\mathrm{Var}(D_{\mathrm{R}}), \tag{9}$$

where D_{L} and D_{R} are the left and right child nodes resulting from the split of sample D. The prediction in a terminal leaf is the mean of y_i in that region. Depth and minimum leaf size are controlled to avoid overfitting. Decision trees are complemented by bagged trees (random forests), which average many trees fitted to bootstrap resamples of the training data, thereby reducing variance.

Boosted trees combine many shallow trees in an additive fashion. Starting from an initial prediction $F_0(x)$, gradient boosting iteratively fits trees $h_t(x)$ to the residuals and updates the prediction function as

$$F_t(x) = F_{t-1}(x) + \eta h_t(x), \tag{10}$$

where η is the learning rate. In our implementation we use small trees, shrinkage and cross-validated early stopping to control complexity.

4.4.2. Stochastic Gradient Machines

Stochastic Gradient Machines extend boosting by incorporating randomization and explicit regularization. Stochastic gradient boosting trains each tree on a random subsample of the data, which stabilizes the ensemble and mitigates overfitting. XGBoost further augments the objective with a penalty on tree complexity,

$$\mathrm{Obj} = \sum_{i=1}^{n} L(y_i, F_{t-1}(x_i) + h_t(x_i)) + \Omega(h_t), \tag{11}$$

where $L(\cdot,\cdot)$ is a differentiable loss function and $\Omega(h_t)$ penalizes the number of leaves and the magnitude of leaf weights. The use of first- and second-order derivatives of L yields efficient, numerically stable updates.

Finally, LightGBM is a gradient-boosting framework designed to scale gradient-tree boosting to large and sparse predictor sets. As in XGBoost, each iteration adds a new tree that minimizes a regularized quadratic approximation of the loss, but LightGBM grows trees in a leaf-wise fashion: at each step it selects the leaf whose split delivers the largest reduction in the objective and expands only that leaf, subject to depth and minimum-leaf constraints.

4.5. The Horse Race

Given the specific characteristics of our data set and the empirical objectives of this paper, comparing various machine learning techniques is an indispensable practice which provides insights into the efficacy and applicability of different algorithms. This sub-section aims to systematically compare the selected machine learning techniques.

4.5.1. The race setup

To compare the thirteen forecasting models on a common footing, we organize a "horse race" based on three standard accuracy measures: the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE) and the coefficient of determination R^2. The preferred specification is the one that delivers the lowest RMSE and MAE and the highest R^2 on a hold-out test sample. For a given model, let y_i denote the realized spread, $\hat{y}_i$ the forecast, and n the number of test observations. The loss functions are

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \qquad \text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{12}$$

and the goodness-of-fit measure is

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{13}$$

where $\bar{y}$ is the mean of y_i. RMSE and MAE summarize the magnitude of forecast errors (with RMSE penalizing large errors more heavily), while R^2 indicates how much of the variation in spreads the model explains relative to a naïve mean forecast.

All models are estimated on a training subsample comprising 80% of the data and evaluated on the remaining 20% test subsample. The split is fixed throughout and the test set is never used for model fitting or tuning. Within the training set, we rely on K-fold cross-validation to select regularization and complexity parameters where appropriate. For penalized linear models (Standard LASSO, Group LASSO, Sparse-group LASSO, Ridge and Elastic Net), we use 10-fold cross-validation to choose the penalty strength λ (with the mixing parameter in Elastic Net fixed at $\alpha = 0.5$).

The nonlinear models are specified as follows. Support Vector Regression is estimated with the default radial kernel and tuning parameters on the training set, and forecasts are computed on the test set. The Feedforward Neural Network is a two hidden layers network with 10 and 5 units, respectively; this architecture is fixed *ex ante* and trained on the training data before generating test-set forecasts.

The tree-based and ensemble models use a combination of cross-validation and internal resampling. The single decision tree is estimated using cost-complexity pruning on the training data, and its performance is evaluated on the test set. The bagged-tree model is implemented as a random forest with 3,000 trees; bootstrap aggregation and out-of-bag error provide stabilization and final accuracy is measured on the test set. Gradient-boosted trees (GBM) and stochastic gradient boosting (SGB) are estimated on the training data with fixed depth and shrinkage parameters; in each case, 5-fold cross-validation determines the effective number of trees. Because these boosting algorithms are stochastic, the train–validate–test procedure is repeated 50 times with different random seeds and we aggregate their performance by averaging RMSE, MAE and R^2 across repetitions.

For XGBoost, we first conduct a random search over a hyperparameter grid. Each candidate configuration is evaluated using 10-fold cross-validation and early stopping on the training set to select the optimal number of boosting rounds. We then fix the best configuration and re-estimate the model 50 times on the full training sample, generating test-set forecasts in each run. LightGBM is estimated in a similar Monte Carlo fashion: in each of 50 runs, we perform 10-fold cross-validation with early stopping on the training data to determine the number of iterations, train the final model at that stopping point, and compute test-set forecasts.

Forecast performance is always reported on the hold-out test sample. For deterministic models, we report a single set of RMSE, MAE and R^2. For stochastic ensemble methods (GBM, SGB, XGBoost and LightGBM), we report averages of these metrics over 50 repetitions, which smooths out the effect of algorithmic randomness and yields a more stable ranking of models.

4.5.2. The race results

In this horse race, each of the thirteen forecasting models is a competing "horse" and each country constitutes a separate race. For every country we run three heats, one for each accuracy metric (RMSE, MAE and R^2), which yields 30 country–metric combinations in total. Table 6 presents the percentage of wins by model and metric while tables 7–9 report the full set of test-sample RMSE, MAE and R^2 values. Finally, Table 10 shows the results of the Diebold-Mariano tests.

Table 6: Percentage of wins by model and metric

Model	RMSE wins (%)	MAE wins (%)	R^2 wins (%)	Overall wins (%)
XGBoost	40.0	70.0	40.0	50.0
LightGBM	20.0	10.0	20.0	16.7
Standard LASSO	10.0	10.0	10.0	10.0
Elastic Net	10.0	10.0	10.0	10.0
Group LASSO	10.0	0.0	10.0	6.7
SG Boosting	10.0	0.0	10.0	6.7
Decision Trees	0.0	0.0	0.0	0.0
Bagged Trees (Random Forest)	0.0	0.0	0.0	0.0
Boosted Trees (GBM)	0.0	0.0	0.0	0.0
Ridge	0.0	0.0	0.0	0.0
Sparse-group LASSO	0.0	0.0	0.0	0.0
Support Vector Regression	0.0	0.0	0.0	0.0
Feedforward Neural Network	0.0	0.0	0.0	0.0

Notes: Percentages are computed over 10 country-level races per metric (RMSE, MAE, R^2), and 30 races in total for the overall column. In each country–metric race, ties are split equally among the models sharing the best score.

RMSE results

Table 7 summarizes the Root Mean Squared Error (RMSE) for all models and countries. XGBoost delivers the lowest RMSE in four of the ten countries (Finland, France, Ireland and the Netherlands). In the remaining cases, the winning model is another tree-based method or a penalized linear benchmark: Stochastic Gradient Boosting in Austria, Standard LASSO in Belgium, Group LASSO in Greece, LightGBM in Italy and Portugal, and Elastic Net in Spain. In core countries (Austria, Belgium, Finland, France and the Netherlands), XGBoost wins 6 out of 15 core-country races when RMSE, MAE and R^2 are considered jointly, with the residual victories split mainly between Standard LASSO and Stochastic Gradient Boosting. In peripheral countries (Greece, Ireland, Italy, Portugal and Spain), XGBoost faces stronger competition from LightGBM and Elastic Net in terms of RMSE, but it remains among the top-performing models in every market.

The magnitude of the RMSE gains over linear benchmarks is economically meaningful in several countries. For example, in Finland XGBoost attains an RMSE of about 0.054, compared with roughly 0.069 for the best penalized linear model (Group LASSO). In France the gap is larger: the RMSE of XGBoost is approximately 0.078 versus 0.120 for the best linear competitor, implying a reduction of roughly one third. In Ireland, where spreads are comparatively volatile, XGBoost cuts RMSE from about 0.34 (Standard LASSO) to 0.14. By contrast, in Belgium and Spain the best penalized linear regression (Standard LASSO in Belgium, Elastic Net in Spain) achieves lower RMSE than XGBoost, while in Austria, Greece, Italy and Portugal the differences between XGBoost and the RMSE winner are small in absolute terms.

MAE results

Turning to the Mean Absolute Error (MAE) in Table 8, XGBoost is even more dominant. It achieves the lowest MAE in seven of the ten countries (Austria, Finland, France, Greece, Ireland, Italy and the Netherlands). In Portugal, LightGBM attains the smallest MAE, and in Belgium and Spain the best performers are Standard LASSO and Elastic Net, respectively. Again, the improvements relative to linear and kernel-based benchmarks are sizable in several markets. In Greece, for instance, XGBoost reduces MAE to about 1.05, whereas the best penalized linear benchmark (Group LASSO) records an MAE around 1.36. In Portugal, XGBoost attains an MAE close to 0.39 compared with roughly 0.49 for the best linear alternative. In core countries, XGBoost consistently dominates Support Vector Regression and the Feedforward Neural Network, which display larger and less stable MAE values despite their flexibility. Taken together, the MAE results show that XGBoost yields the most accurate forecasts in terms of typical error size in the majority of countries under study.

The R^2 results

Table 9 reports the out-of-sample R^2, which measures the proportion of test-sample variance explained relative to a naïve mean forecast. XGBoost achieves the highest R^2 in four of the ten countries (Finland, France, Ireland and the Netherlands). In the remaining markets, the R^2 winner varies: Stochastic Gradient Boosting in Austria, Standard LASSO in Belgium, Group LASSO in Greece, LightGBM in Italy and Portugal, and Elastic Net in Spain. Although a few penalized linear and alternative tree-based methods occasionally match or slightly exceed XGBoost's R^2 in individual countries, XGBoost never performs poorly; it remains in the top tier of models for variance explanation across all markets. The magnitude of the R^2 differences is again non-trivial in several cases. In Finland, XGBoost attains an out-of-sample R^2 around 0.91, compared with roughly 0.85 for the best linear benchmark. In Ireland, the gap is particularly pronounced: XGBoost reaches an R^2 close to 0.99, while the best linear model explains about 0.93 of the variance. Even in countries where another model attains the highest R^2, such as Belgium or Spain, the advantage over XGBoost is modest and comes at the cost of higher RMSE or MAE.

Table 7: Horse Race - Root Mean Square Error (RMSE)

Machine Learning Techniques					**Country**					
Linear	*Austria*	*Belgium*	*Finland*	*France*	*Greece*	*Ireland*	*Italy*	*Netherlands*	*Portugal*	*Spain*
Standard LASSO	0.111	**0.099**	0.073	0.121	2.324	0.344	0.387	0.060	0.897	0.431
Group LASSO	0.103	0.214	0.069	0.120	**2.042**	0.485	1.041	0.076	0.671	0.440
Sparse-group LASSO	0.122	0.113	0.072	0.128	2.276	0.380	0.415	0.060	1.008	0.452
Ridge	0.163	0.264	0.086	0.147	3.537	0.547	0.633	0.072	1.402	0.657
Elastic Net	0.115	0.104	0.072	0.127	2.228	0.361	0.402	0.058	0.901	**0.411**
Non-linear										
Support Vector Regression (SVR)	0.144	0.261	0.072	0.155	2.772	0.385	0.566	0.072	1.232	0.606
Feedforward Neural Network (FNN)	0.208	0.381	0.178	0.216	7.242	1.468	1.024	0.124	2.464	1.340
Rule-based										
Trees										
Decision Trees	0.126	0.184	0.083	0.107	2.856	0.277	0.547	0.067	1.529	0.771
Bagged Trees (Random Forest)	0.107	0.195	0.064	0.096	2.294	0.143	0.412	0.057	0.699	0.497
Boosted Trees (GBM)	0.096	0.194	0.057	0.080	2.194	0.189	0.382	0.053	0.616	0.467
Stochastic Gradient Machines (SGM)										
SG Boosting	**0.089**	0.202	0.057	0.078	2.057	0.163	0.387	0.054	0.622	0.449
XGBoost	0.090	0.166	**0.054**	**0.078**	2.143	**0.138**	0.372	**0.052**	0.647	0.508
LightGBM	0.111	0.195	0.061	0.088	2.657	0.184	**0.355**	0.053	**0.543**	0.480

Table 8: Horse Race - Mean Absolute Error (MAE)

Machine Learning Techniques					**Country**					
Linear	*Austria*	*Belgium*	*Finland*	*France*	*Greece*	*Ireland*	*Italy*	*Netherlands*	*Portugal*	*Spain*
Standard LASSO	0.082	**0.069**	0.059	0.080	1.656	0.263	0.266	0.050	0.685	0.275
Group LASSO	0.078	0.137	0.052	0.097	1.355	0.269	0.757	0.063	0.490	0.295
Sparse-group LASSO	0.081	0.079	0.059	0.091	1.253	0.292	0.282	0.049	0.730	0.272
Ridge	0.107	0.130	0.068	0.095	1.833	0.416	0.452	0.057	1.006	0.394
Elastic Net	0.081	0.071	0.059	0.090	1.262	0.283	0.291	0.049	0.679	**0.259**
Non-linear										
Support Vector Regression (SVR)	0.092	0.119	0.054	0.090	1.405	0.293	0.372	0.055	0.699	0.340
Feedforward Neural Network (FNN)	0.149	0.221	0.137	0.151	4.494	0.836	0.785	0.105	1.893	0.914
Rule-based										
Trees										
Decision Trees	0.074	0.100	0.063	0.077	1.680	0.197	0.360	0.049	0.928	0.492
Bagged Trees (Random Forest)	0.072	0.095	0.049	0.065	1.148	0.098	0.278	0.045	0.474	0.293
Boosted Trees (GBM)	0.069	0.093	0.044	0.058	1.271	0.129	0.279	0.043	0.424	0.296
Stochastic Gradient Machines (SGM)										
SG Boosting	0.065	0.090	0.045	0.059	1.120	0.105	0.261	0.043	0.414	0.283
XGBoost	**0.063**	0.080	**0.042**	**0.057**	**1.052**	**0.089**	**0.245**	**0.041**	0.393	0.275
LightGBM	0.082	0.088	0.047	0.064	1.460	0.112	0.252	0.043	**0.372**	0.288

Table 9: Machine learning techniques – R^2

Machine learning techniques	**Country**									
	Austria	*Belgium*	*Finland*	*France*	*Greece*	*Ireland*	*Italy*	*Netherlands*	*Portugal*	*Spain*
Linear										
Standard LASSO	0.814	**0.955**	0.835	0.669	0.898	0.932	0.860	0.749	0.889	0.886
Group LASSO	0.839	0.788	0.851	0.676	**0.921**	0.865	-0.012	0.595	0.938	0.882
Sparse-group LASSO	0.777	0.941	0.842	0.630	0.902	0.917	0.839	0.750	0.859	0.875
Ridge	0.597	0.677	0.773	0.514	0.763	0.828	0.625	0.635	0.728	0.735
Elastic Net	0.802	0.950	0.839	0.637	0.906	0.925	0.849	0.766	0.888	**0.896**
Non-linear										
Support Vector Regression (SVR)	0.686	0.685	0.840	0.455	0.854	0.915	0.701	0.638	0.790	0.775
Feedforward Neural Network (FNN)	0.344	0.329	0.029	-0.056	0.005	-0.240	0.021	-0.085	0.159	-0.100
Rule-based										
Trees										
Decision Trees	0.762	0.844	0.787	0.740	0.845	0.956	0.721	0.688	0.676	0.637
Bagged Trees (Random Forest)	0.826	0.823	0.874	0.793	0.900	0.988	0.841	0.774	0.932	0.849
Boosted Trees (GBM)	0.862	0.825	0.900	0.855	0.908	0.979	0.864	0.806	0.947	0.866
Stochastic Gradient Machines (SGM)										
SG Boosting	**0.880**	0.811	0.900	0.862	0.920	0.985	0.860	0.795	0.946	0.876
XGBoost	0.878	0.872	**0.911**	**0.864**	0.913	**0.989**	0.870	**0.808**	0.942	0.842
LightGBM	0.813	0.824	0.887	0.824	0.866	0.981	**0.882**	0.806	**0.959**	0.859

Diebold–Mariano tests and final ranking

The pointwise ranking based on RMSE, MAE and R^2 is complemented by Diebold-Mariano (DM) tests for equal predictive accuracy. For each country, we compare every model to XGBoost using one-step-ahead squared-error loss. This yields $12 \times 10 = 120$ pairwise comparisons. In 46 of these cases the alternative model is significantly worse than XGBoost at the 5% level, while in the remaining 74 cases the DM statistic is not significant. Crucially, no model ever significantly outperforms XGBoost in any country. In other words, whenever another method attains a slightly lower RMSE or higher R^2 in Tables 6–8, the DM tests treat the difference as statistically indistinguishable from XGBoost, whereas in many instances XGBoost is statistically superior.

The strength of XGBoost is particularly visible in countries such as Finland, Ireland, the Netherlands and Portugal, where DM tests reject equal accuracy in favor of XGBoost against a majority of competitors. For example, in Finland and Ireland, 8 and 7 out of 12 alternative models, respectively, are significantly dominated by XGBoost. Neural networks and some ridge-type specifications are especially fragile: they are frequently and significantly outperformed by XGBoost despite their nominal flexibility.

Table 10: Diebold–Mariano tests of equal predictive accuracy vs XGBoost

	# significantly worse		# significantly better		# not significantly different	
Model	**Count**	**Share**	**Count**	**Share**	**Count**	**Share**
Standard LASSO	3	30%	0	0%	7	70%
Group LASSO	5	50%	0	0%	5	50%
Sparse-group LASSO	5	50%	0	0%	5	50%
Ridge	7	70%	0	0%	3	30%
Elastic Net	4	40%	0	0%	6	60%
Support Vector Regression (SVR)	5	50%	0	0%	5	50%
Feedforward Neural Network (FNN)	9	90%	0	0%	1	10%
Decision Trees	6	60%	0	0%	4	40%
Random Forest (Bagged Trees)	0	0%	0	0%	10	100%
Boosted Trees (GBM)	0	0%	0	0%	10	100%
SG Boosting	0	0%	0	0%	10	100%
LightGBM	2	20%	0	0%	8	80%

Notes: Entries report, for each competing model, the number and share of countries for which Diebold-Mariano tests (at the 5% level, two-sided) indicate significantly higher, significantly lower or statistically indistinguishable forecast accuracy relative to XGBoost.

Overall, XGBoost attains the best metric in 15 out of the 30 country–metric combinations (4 RMSE races, 7 MAE races and 4 R^2 races) as shown in Table 6, and is never significantly dominated in any Diebold-Mariano test. Competing tree-based methods such as LightGBM and Stochastic Gradient Boosting occasionally outperform XGBoost in individual countries, and penalized linear regressions (Standard LASSO and Elastic Net) can match it in Belgium and Spain. However, these gains are small and statistically non-significant, while XGBoost's advantage is both systematic and statistically supported across many comparisons.

Therefore, we view XGBoost as the winner of the horse race. Its boosted tree architecture is well suited to the high-dimensional, nonlinear and potentially interacting determinants of sovereign spreads. For the remainder of the paper, we adopt XGBoost as our forecasting model and use its predictions as the basis for subsequent analyses of financial fragmentation in euro area sovereign bond markets.

5. Empirical Results

This empirical results section implements the forecasting framework developed in the previous section. Since the primary objective of this paper is to assess whether modern machine-learning methods can deliver accurate forecasts in a novel high-dimensional macro-financial setting, we take as our forecasting target the 10-year sovereign bond yield spreads vis-à-vis the German Bund.

Using the XGBoost model, we generate one-month-ahead forecasts of 10-year sovereign bond yield spreads for Austria, Belgium, Finland, France, Greece, Ireland, Italy, the Netherlands, Portugal and Spain over the period from January 2020 to February 2025, conditioning on the full macro-financial data set. The empirical analysis proceeds in two steps. First, we conduct a country-specific examination of the forecast paths, comparing predicted spreads with realized outcomes to gauge the accuracy of the model and to infer how sovereign risk evolves at the national level. Second, we analyze the cross-sectional structure of the predicted spreads by constructing correlation heatmaps based on the dissimilarity measure $1 - \rho_{i,j}$, which reveal the extent and persistence of the core-periphery divide within the euro area. Fragmentation patterns thus emerge as an economically relevant consequence of the forecasting exercise. Taken together, these results both validate the XGBoost benchmark in a genuinely out-of-sample setting and provide a novel forward-looking view of sovereign risk and financial fragmentation in the euro Area.

5.1. Country-specific forecasting performance

This first subsection assesses whether machine learning, and XGBoost in particular, can forecast euro area sovereign spreads with economically meaningful accuracy when fed with a rich macro–financial information set. For each of the ten countries under study, we estimate a country–specific XGBoost model and generate one–month–ahead forecasts of the 10-year sovereign bond yield spread vis-à-vis the German Bund over the period from January 2020 to February 2025. The models are trained on the full panel of monthly macro–financial variables and evaluated strictly out-of-sample, using realized spreads that the algorithm does not observe at the time of prediction.

5.1.1. Forecast design and accuracy

The forecasting design closely mirrors best practice in the time–series machine–learning literature. For each country, we first construct a monthly data set in which all predictors are aligned at the end of the month and the target variable is the one–month–ahead spread, y_{t+1}. We implement an expanding-window, rolling-origin scheme: for each forecast origin t, the model is trained on all data up to month $t-1$ and then used to predict the spread at $t+1$. This avoids any look-ahead bias and ensures that the information set available to the algorithm coincides with what a real-time forecaster would have observed.

Forecasts are generated with a XGBoost model using country-specific hyperparameters. Similarly to the horse race, for each forecast origin we fit $B = 50$ separate XGBoost models with different random seeds and draw of subsamples and feature subsets. Predictions from these 50 models are then averaged to produce a bagged forecast $\hat{y}_{t+1}$. The cross-model dispersion is used to construct approximate forecast intervals. This bagging step reduces the variance of individual trees and stabilizes predictions in the presence of complex nonlinear interactions.

Across countries, the resulting one–month–ahead forecasts track realized spreads closely. As presented in table 11, in core economies, forecast errors are small in both absolute and relative terms. For Austria, Belgium, Finland, France, Ireland and the Netherlands, the root mean squared error (RMSE) lies in a narrow band of roughly 4 to 8 basis points, while the mean absolute error (MAE) ranges from about 3 to 7 basis points. The Netherlands sits at the lower end of this range (around 4 bp RMSE and 3 bp MAE), whereas Austria, Belgium and Ireland are closer to 7–8 bp RMSE and 6–7 bp MAE. Figures 1 to 6 confirm that the predicted paths align tightly with the observed trajectories, including during periods of heightened volatility around the COVID-19 shock, the onset of the war in Ukraine and the subsequent inflationary episode.

Table 11: One-month-ahead XGBoost forecast accuracy

Country	**RMSE (bps)**	**MAE (bps)**
Austria	7.67	5.61
Belgium	8.02	6.01
Finland	6.06	4.67
France	7.27	5.50
Greece	40.06	33.82
Ireland	8.07	6.67
Italy	19.60	16.15
Netherlands	4.02	3.09
Portugal	40.01	21.95
Spain	13.08	10.68

Notes: RMSE and MAE are reported in basis points for one-month-ahead forecasts of 10-year sovereign bond yield spreads vis-à-vis the German Bund. Forecasts are generated from country-specific XGBoost models over the period January 2020 - February 2025, and evaluated strictly out of sample.

Forecasting is naturally more challenging in peripheral countries, where spreads are larger and more volatile. For Italy, Greece, Portugal and Spain, the RMSE of the XGBoost forecasts ranges from about 13 basis points in Spain to roughly 20 basis points in Italy and close to 40 basis points in

Greece and Portugal. The corresponding MAE values span from roughly 11 basis points (Spain) to about 16 basis points (Italy) and exceed 30 basis points in Greece, with Portugal around the low-20s. Even in these more turbulent environments, the model captures the main swings in spreads, including the sharp but short-lived tensions during 2020–2022, while inevitably missing some of the most abrupt month-to-month moves. Taken together, these results indicate that a well-tuned XG-Boost model paired with a big data set can deliver precise one-month-ahead predictions of sovereign spreads across a heterogeneous set of euro area issuers.

Compared with traditional smaller-scale linear specifications, this approach offers two key advantages. First, the gradient-boosting architecture flexibly approximates nonlinearities and threshold effects that are difficult to capture with standard affine term-structure or linear spread models. Second, because the model is estimated on hundreds of domestic and global predictors, it can exploit information from a much richer state space than conventional regressions without overfitting, thanks to its built-in regularization and the expanding-window evaluation design. The combination of low forecast errors at short horizons and robustness across very different sovereign risk profiles suggests that machine learning can indeed be a useful tool for forecasting sovereign bond spreads, and hence for monitoring fragmentation risk in real time.

Figure 1: Austria

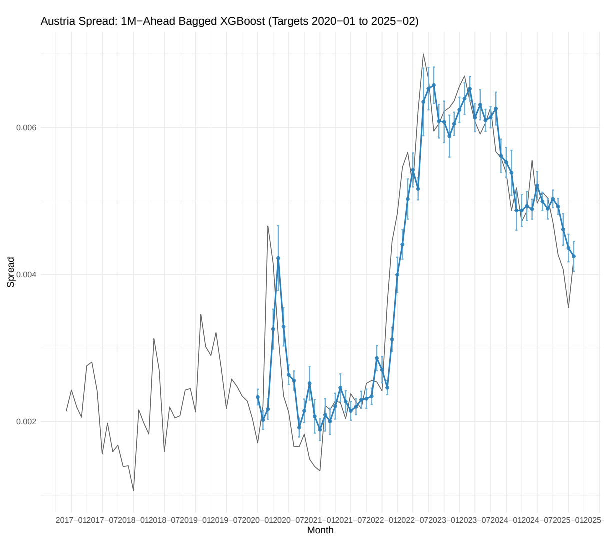


Figure 2: Finland

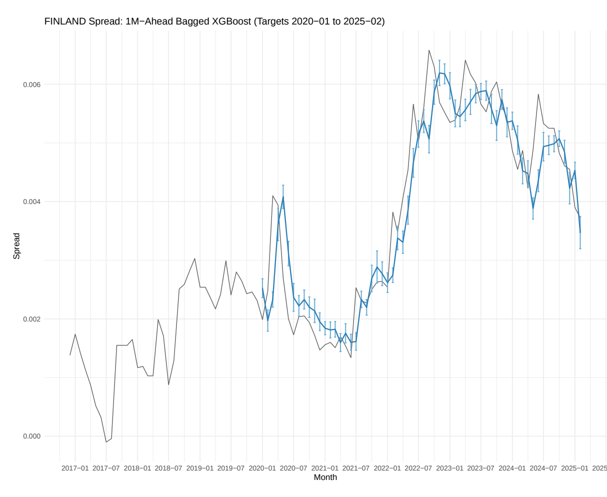


Figure 3: The Netherlands

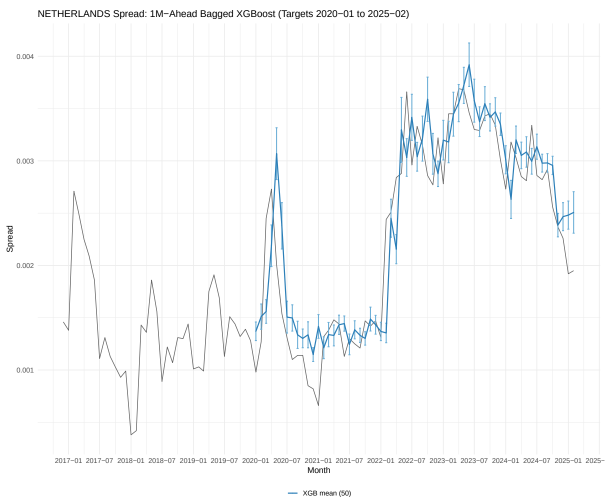


Figure 4: France

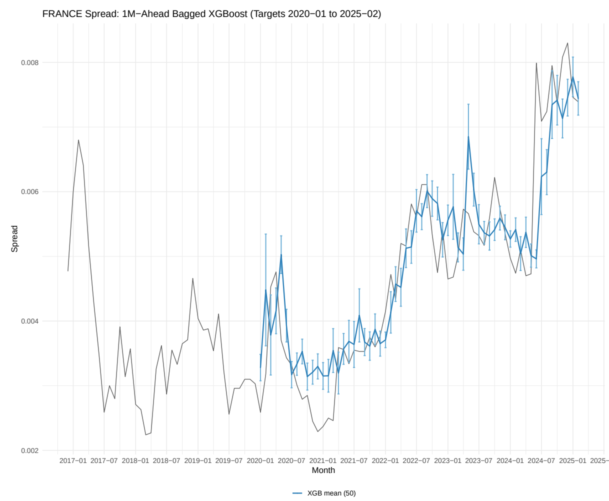


Figure 5: Belgium

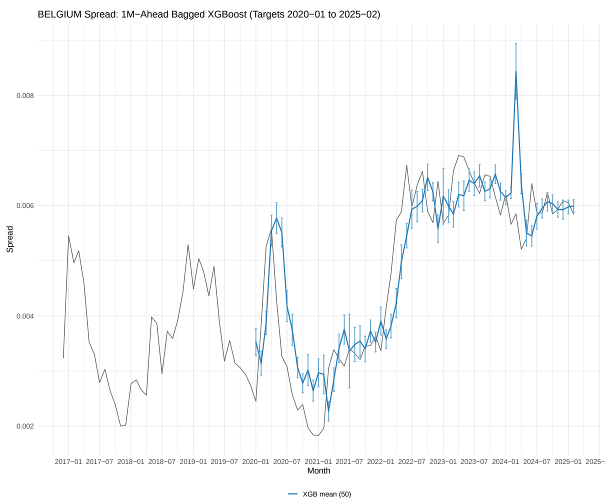


Figure 6: Spain

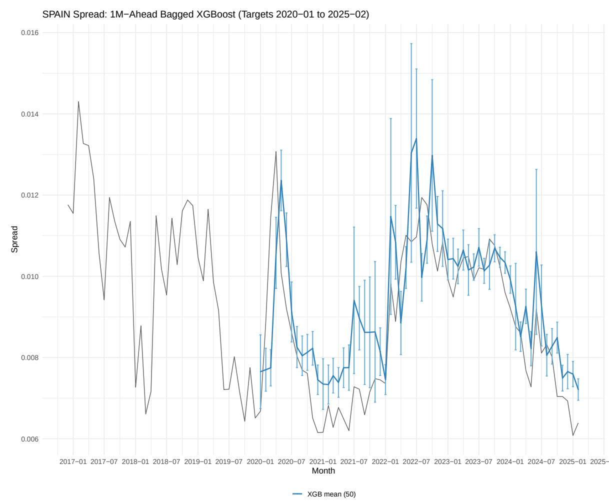


Figure 7: Ireland

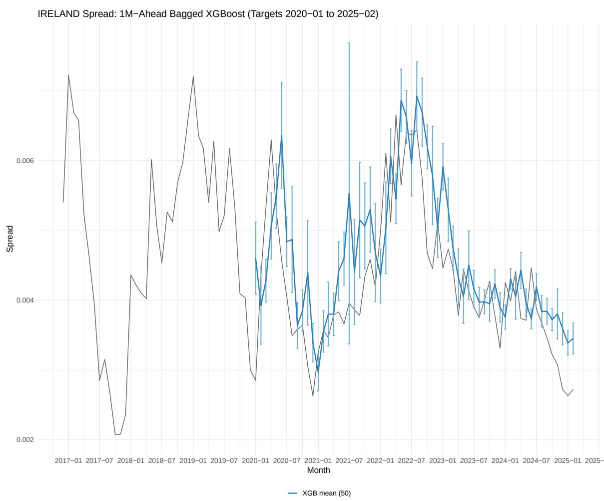


Figure 8: Portugal

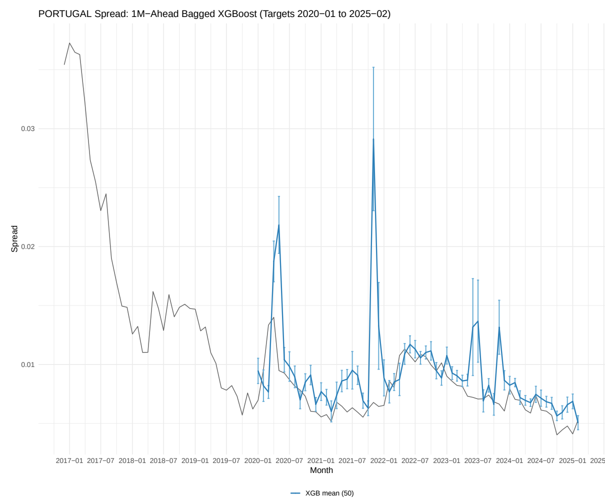


Figure 9: Italy

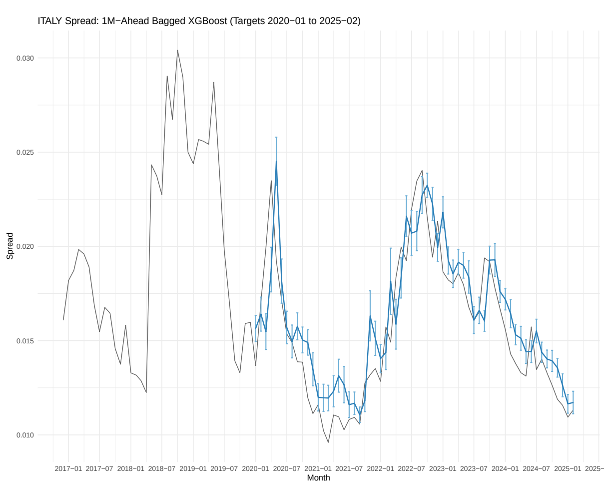
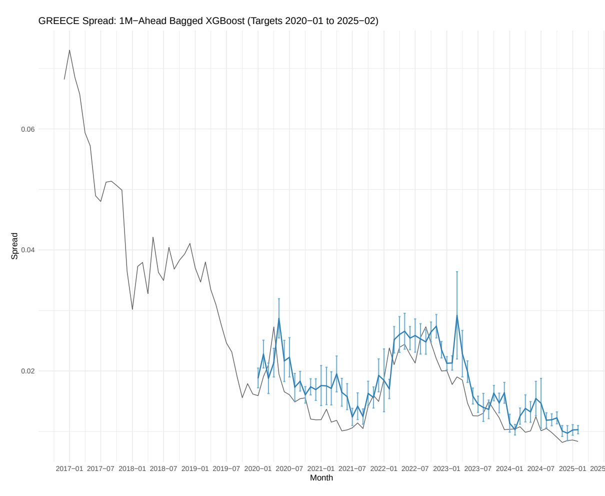


Figure 10: Greece



5.1.2. SHAP-based economic interpretation

The flexibility of XGBoost comes at the cost of reduced transparency relative to linear models. To recover economic structure from the "black box", we compute SHAP values for each monthly training window and each country. For a given forecast origin, the SHAP decomposition attributes to every predictor a marginal contribution to the one-month-ahead spread forecast; averaging absolute contributions over time and across models produces a ranking of the variables that systematically drive the predictions (Table 13).

The SHAP diagnostics reveal a very stark pattern. In every country, the own lagged sovereign spread is the single most important predictor by a wide margin. In the peripheral block (Greece, Portugal, Spain, Italy and Ireland) the aggregate SHAP contribution of the domestic spread dwarfs that of any other variable: it is several times larger than the next-ranked feature and dominates the top-10 list throughout the sample. This confirms that at the one-month horizon risk premia are highly persistent in high-spread markets, and that XGBoost exploits this autoregressive structure very aggressively. In the core economies (Austria, Finland and the Netherlands, and to a lesser extent France and Belgium), the lagged spread still ranks first, but its advantage over other predictors is smaller, consistent with more muted and less volatile spread dynamics.

Beyond this autoregressive component, out of our seventeen categories, six broad groups of variables emerge as systematically important: Financial Indicators, Prices, Global Macro, Personal/Household Sector, Monetary Sector, and Surveys/Cyclical Indicators (Table 12). Financial Indicators are by far the most pervasive block, appearing 29 times and accounting for 58% of all leading SHAP features. Prices come next with 22 occurrences (44%), followed by Global Macro variables with 14 occurrences (28%). Personal/Household Sector indicators appear 11 times (22%), while Monetary Sector variables and Surveys/Cyclical Indicators each appear 8 times (16%). These frequencies make clear that, once the autoregressive component is accounted for, the model relies primarily on macro-financial indicators but also takes into account expectations available in survey-based cyclical indicators to generate forecasts as accurate as possible.

First, Financial Indicators form the dominant block. This group comprises interest rates across the term structure, lending and credit rates and measures of loans and deposits. Their prevalence in the SHAP ranking shows that the model systematically conditions sovereign spread forecasts on the price and quantity of credit. Widening money-market, corporate or bank funding spreads, or higher lending rates relative to benchmarks, raise predicted spreads and are attributed to this block; compressions in key reference spreads or easier credit conditions lower the forecast. Financial Indicators therefore capture the direct repricing of sovereign risk as intermediaries adjust funding costs and credit supply. Price Indexes represent the second most important block. This category includes consumer price indices and subcomponents, together with export and import price indices. High SHAP contributions indicate that XGBoost uses domestic inflation and external price pressures to refine its one-month-ahead sovereign risk assessments. Third, Global Market Indexes (stock market indices,

Table 12: Breakdown of leading SHAP features

Category	Occurrences	Share of top features (%)
Financial Indicators	29	58
Price Indexes	22	44
Global Market Indexes	14	28
Household Sector	11	22
Monetary Sector	8	16
Surveys & Cyclical Indicators	8	16
International Trade	2	4
Government Finance & Debt	1	2
Housing & Real Estate	1	2
Industrial Sector	1	2
Labor Market	1	2
Retail Sector	1	2
Services Sector	1	2

Notes: The table reports how often each category appears among the top 5 leading SHAP features across countries; higher counts indicate that the corresponding category is frequently used by the model when forecasting sovereign spreads.

uncertainty measures and foreign-exchange indicators) appear 14 times (28%) among the leading SHAP features. We could consider equity indices as proxies for global and regional risk appetite, uncertainty measures capturing tail-risk perceptions and exchange rates vis-à-vis major and regional currencies proxy external competitiveness and safe-haven flows. Fourth, indicators related to the Household Sector cover consumer credit outstanding and broader business and consumer credit aggregates, capturing leverage and risk-bearing capacity of the private sector. Increases in household or broad credit, particularly when not backed by strong income or confidence, raise model-implied spreads by signaling a greater likelihood that private stress spills over to public balance sheets. When households and firms deleverage or maintain stable balance sheets, SHAP contributions from this block dampen predicted sovereign risk. Fifth, Monetary Sector variables (money supply aggregates, bank balance-sheet items and international reserves) summarize the liquidity environment and the size and structure of the banking system relative to the sovereign, while reserves proxy the external buffer against funding shocks. Tighter liquidity, weaker bank balance sheets or falling reserves are associated with higher predicted spreads; ample liquidity and robust banking assets reduce them. Sixth and finally, Surveys and Cyclical Indicators, including consumer confidence and business conditions provide forward-looking information on activity and employment. Deteriorating survey readings increase the forecasted spread by anticipating weaker growth and a higher probability of fiscal stress, while strong and broad-based survey outcomes reduce it.

Taken together, the SHAP evidence shows that XGBoost does not rely on isolated statistical artefacts. Instead, it recovers an economically coherent structure. Domestic spread persistence remains the primary driver of one-month-ahead risk premia, but the model also systematically exploits a broad set of macro–financial indicators to uncover country-specific, sometimes non-obvious relationships between fundamentals and spreads. This ability to combine a large, heterogeneous information set with a flexible forecasting technology is what distinguishes our approach from traditional smaller-scale econometric models: it allows the algorithm to uncover country-specific, and sometimes non-obvious, macro–financial linkages that are difficult to detect or accommodate within conventional, low-dimensional specifications.

5.2. The euro area's yield spread dynamics

To complement the country-by-country analysis, we examine how the joint dynamics of predicted spreads evolve over time. For each subperiod under study, we construct a correlation heatmap based on the XGBoost one-month-ahead forecasts of 10-year sovereign spreads. Let $\hat{s}_{i,t}$ denote the forecast for country i at forecast origin t. For a given sample window, we compute the Pearson correlation matrix ρ_{ij} of $\hat{s}_{i,t}$ and $\hat{s}_{j,t}$ across months, and transform it into a dissimilarity matrix

$$d_{ij} = (1 - \rho_{ij})^{\beta},$$

with $\beta = 2$. This transformation penalizes weak or negative correlations and magnifies the distance between countries whose forecasted spread paths move in opposite directions. We then apply hierarchical clustering with complete linkage to the distance matrix and cut the resulting dendrogram at $K = 2$ clusters. We set the number of clusters to $K = 2$ in order to mirror the standard core-periphery dichotomy in the euro area sovereign spread literature. The correlation matrix is reordered according to the dendrogram and the two endogenously determined clusters are highlighted by rectangles in the heatmaps (Figures 11–14).

The rationale for this exercise is straightforward. Fragmentation is not a static characteristic of the euro area; it is a dynamic property of how sovereign risk co-moves across countries over time. A country can move closer to, or further away from, the core depending on how investors reassess its fundamentals, fiscal position and exposure to common shocks. By recomputing the correlation structure on rolling sample windows, using only the model-based forecast paths, we obtain a simple but powerful device for monitoring fragmentation in (near) real time. In particular, changes in cluster membership or in the composition of the "core" and "periphery" blocks provide an early indication that the underlying spread dynamics, as perceived by the forecasting model, are reorganizing in a non-trivial way.

The full-sample heatmap for January 2020–January 2025 (Figure 11) displays a familiar core–periphery structure as described in the literature. One cluster groups Austria, Finland, the Netherlands, France and Belgium, which exhibit high intra-cluster correlations in their predicted spreads and relatively weak comovement with the remaining members. This group corresponds to the conventional core, characterized by lower and more stable spreads and by a stronger "flight-to-quality" component in investor behavior. The second cluster brings together Greece, Italy, Portugal, Spain and Ireland, which show strong internal cohesion and markedly different dynamics relative to the core. This configuration is consistent with a periphery block whose spreads are jointly influenced by higher debt burdens, more fragile fiscal positions and greater sensitivity to global risk repricing.

Table 13: Top five SHAP features by country

Bloomberg Ticker	**Category 1**	**Category 2**	**Total |SHAP|**
Austria			
SPREAD_AUSTRIA	Financial Indicators	Spreads	0.0583
ATMODPL3 Index	Monetary Sector	Money Supply	0.0113
GM_CAC StockM	Global Market Indexes	Stock Markets	0.0083
GM_NDX StockM	Global Market Indexes	Stock Markets	0.0032
GM_DAX StockM	Global Market Indexes	Stock Markets	0.0032
Belgium			
SPREAD_BELGIUM	Financial Indicators	Spreads	0.0717
CP55BE Index	Price Indexes	Harmonised Index of Consumer Prices	0.0554
BGLDNFOO Index	Financial Indicators	Lending & Credit	0.0238
CP82BE Index	Price Indexes	Harmonised Index of Consumer Prices	0.0232
CPCTBE Index	Price Indexes	Harmonised Index of Consumer Prices	0.0141
Finland			
SPREAD_FINLAND	Financial Indicators	Spreads	0.0281
ECOYEFIN Index	International Trade	International Trade Balance	0.0116
GM_Financial_Uncertainty_1M	Global Market Indexes	Uncertainty measures	0.0038
CP1FFIYY Index	Price Indexes	Harmonised Index of Consumer Prices	0.0016
CP9MFIYY Index	Price Indexes	Harmonised Index of Consumer Prices	0.0014
France			
SPREAD_FRANCE	Financial Indicators	Spreads	0.0393
IPIPEDCA Index	Price Indexes	Import Price Indexes	0.0214
IPIPJJSI Index	Price Indexes	Import Price Indexes	0.0085
GM_EURDKK Curncy	Global Market Indexes	FOREX	0.0056
IPIPPFAV Index	Price Indexes	Import Price Indexes	0.0038
Greece			
SPREAD_GREECE	Financial Indicators	Spreads	1.3895
GIRELPWC Index	Financial Indicators	Interest Rates	0.8084
LDHHOLGR Index	Household Sector	Consumer Credit Outstanding	0.1686
R L N F L 1 G R Index	Financial Indicators	Interest Rates	0.1074
IC29MGRY Index	Price Indexes	Import Price Indexes	0.0663
Ireland			
SPREAD_IRELAND	Household Sector	Consumer Credit Outstanding	0.4180
LIHPOLO3 Index	Household Sector	Consumer Credit Outstanding	0.3912
ICADCAIC Index	Household Sector	Business & Consumer Credit Outstanding	0.1045
GM_SPX StockM	Global Market Indexes	Stock Markets	0.0693
BISNIER Index	Financial Indicators	Exchange Rates	0.0245
Italy			
SPREAD_ITALY	Financial Indicators	Spreads	0.2670
ASSITREA Index	Monetary Sector	Bank Balance Sheets	0.0572
ITLOOM5 Index	Household Sector	Consumer Credit Outstanding	0.0426
ITLOG Index	Financial Indicators	Lending & Credit	0.0363
GM_CAC StockM	Global Market Indexes	Stock Markets	0.0143
Netherlands			
SPREAD_NETHERLANDS	Financial Indicators	Spreads	0.0358
GM_EURRUB Curncy	Global Market Indexes	FOREX	0.0067
RDHNO1NT Index	Financial Indicators	Interest Rates	0.0024
NERSCEVM Index	Retail Sector	Retail Sales	0.0024
R L N F L 2 N L Index	Financial Indicators	Interest Rates	0.0023
Portugal			
SPREAD_PORTUGAL	Financial Indicators	Spreads	0.6812
PTCCPE Index	Surveys & Cyclical Indicators	Consumer Confidence	0.1065
EUCOPT Index	Surveys & Cyclical Indicators	Business Conditions	0.1025
PTMODM2Y Index	Monetary Sector	Money Supply	0.0975
PTMOREPO Index	Monetary Sector	Money Supply	0.0716
Spain			
SPREAD_SPAIN	Financial Indicators	Spreads	0.3356
VLHBBORL Index	Household Sector	Consumer Credit Outstanding	0.0385
SPIFNDGP Index	Monetary Sector	International Reserves	0.0212
GM_Elections_and_Political_Governance_EMV_Tracker	Global Market Indexes	Uncertainty Measures	0.0155
GM_FTSEMIB StockM	Global Market Indexes	Stock Markets	0.0133

Notes: For each country, we report the five predictors with the largest SHAP importance. The last column reports, for each feature, the sum of absolute SHAP values over all forecast origins from December 2019 to January 2025 (corresponding to targets January 2020 to February 2025). Category 1 and Category 2 follow the macro–financial classification described in Table 2.

The fact that this dichotomy emerges from model-based forecasts, rather than from ex-post realized spreads, indicates that the machine-learning model internalizes the traditional core–periphery divide when forming expectations over the recent period.

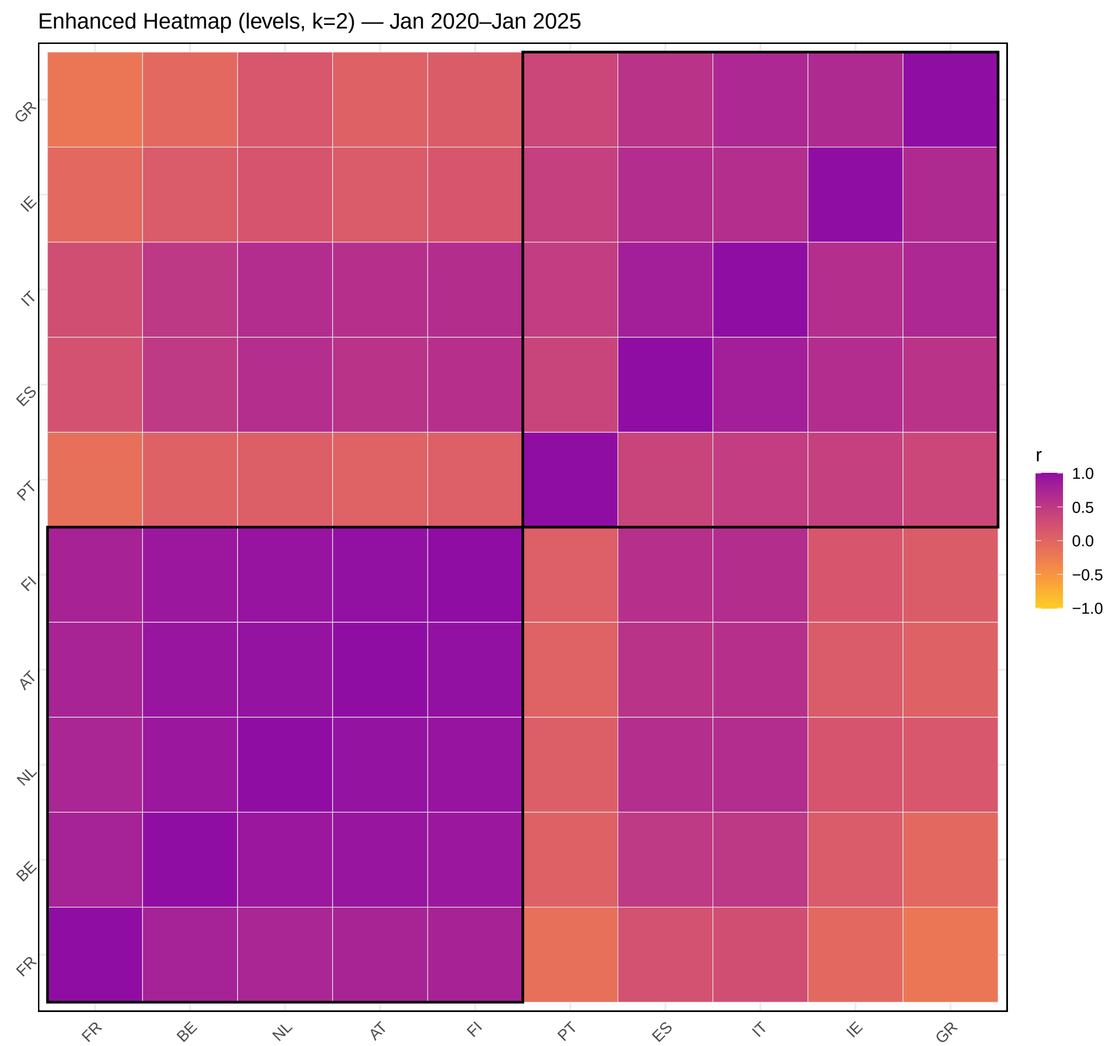
Restricting the window to the post-COVID and post-energy-shock period (January 2022–February 2025) does not overturn this picture. The heatmap for 2022–2025 (Figure 12) continues to produce two well-separated clusters that closely resemble the standard core–periphery split. Despite the large common shocks associated with the pandemic, the energy price spike and the subsequent tightening of ECB policy, the correlation structure of predicted spreads still segregates high-spread, high-volatility sovereigns from low-spread, safe-haven issuers. In other words, the major macro–financial shocks of 2020–2022 did not eliminate the structural fragmentation of the euro area sovereign market; they temporarily affected all countries but left the underlying clustering of risk premia intact.

The situation changes once we move to more recent subsamples. In the heatmap for January 2023–February 2025 (Figure 13), France no longer fits cleanly into either the core or the periphery block. The hierarchical clustering algorithm isolates France as a stand-alone cluster, with a correlation pattern vis-à-vis all other countries that is sufficiently distinct to prevent it from being grouped with any of the two standard blocks. Economically, this "single-country cluster" means that the model perceives French spread dynamics as increasingly idiosyncratic: they neither co-move strongly with the traditional core nor with the usual periphery. From a fragmentation perspective, this is a worrying development. A large, systemically important member of the euro area whose sovereign spread pattern decouples from both groups can become a separate source of instability, as market participants reassess its fiscal trajectory, political risk and long-term debt sustainability.

The most recent window, June 2024–February 2025 (Figure 14), intensifies this signal. In this heatmap, France and Belgium jointly form a separate cluster, distinct from the remaining eight countries. The interpretation is that the predicted spread dynamics of France and Belgium have started to move together, but in a way that is systematically different from both the traditional core and the high-spread periphery. In practice, this points to a third group emerging in the euro area: a set of large or mid-sized sovereigns that are no longer treated by the model as safe-haven issuers, but that also do not fully share the risk profile of the historical periphery. Such a configuration amplifies fragmentation risk: shocks hitting France or Belgium are unlikely to be absorbed smoothly by the rest of the core, and may instead propagate in a more disorderly fashion across the union.

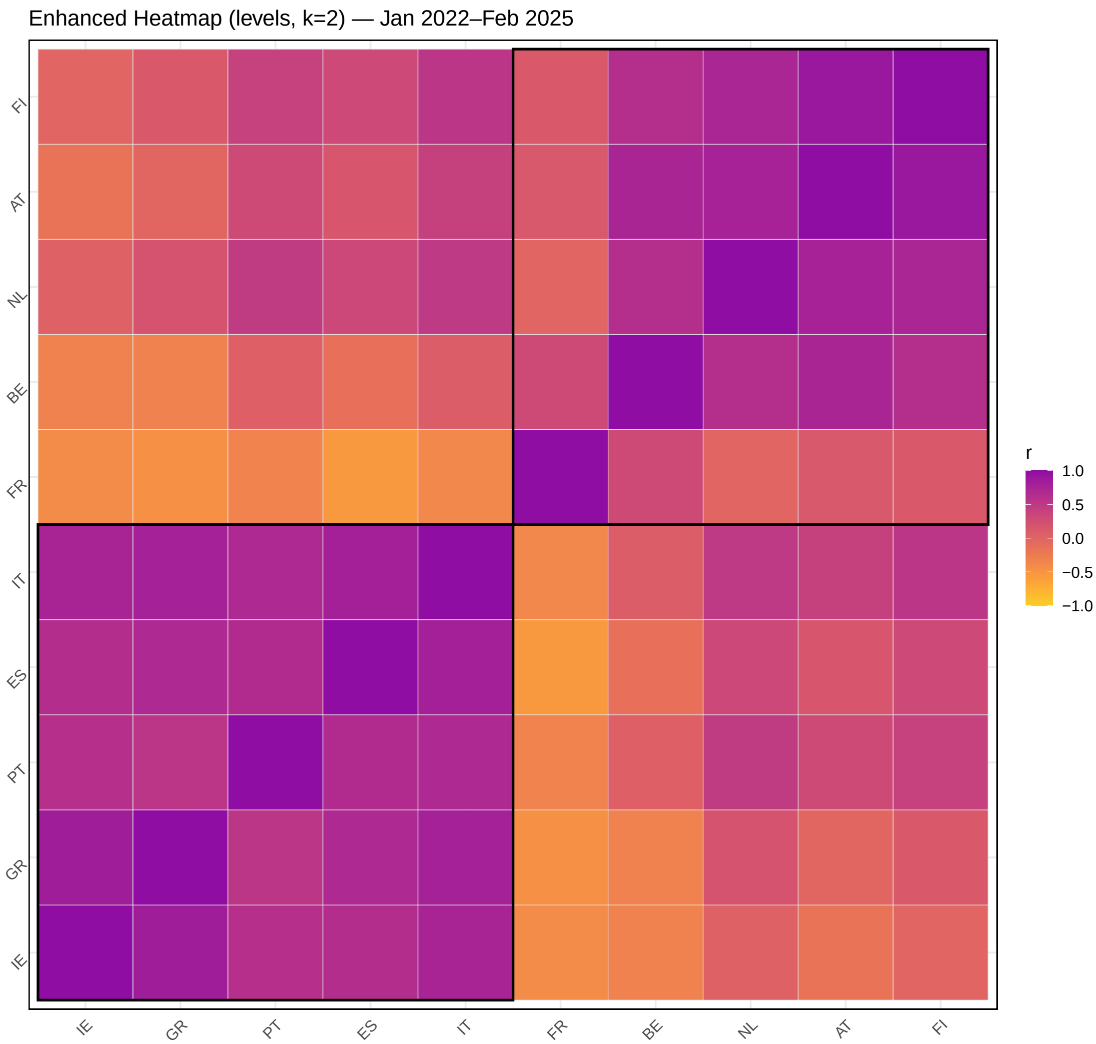
These patterns have not straightforward policy implications. If a major member state such as France starts to detach from the core and to form a separate cluster with Belgium, the "single" monetary policy of the ECB becomes even harder to defend. A more heterogeneous correlation structure of sovereign spreads undermines the uniform transmission of policy rates, increases the likelihood of persistent risk-premium differentials and raises the bar for instruments such as targeted bond purchase programs to be effective. Moreover, a sustained decoupling of France and Belgium from the core cluster could re-activate bank–sovereign feedback loops in those countries and heighten redenomination or break-up risk premiums in the wider euro-denominated asset space. It therefore stresses the importance for both France and Belgium to come back to credible medium-term fiscal objectives, compatible with the other euro members and in line with levels imposed by the Maastricht treaty.

Figure 11: Correlation heatmap of predicted spreads (2 clusters, levels, Jan 2020–Jan 2025)



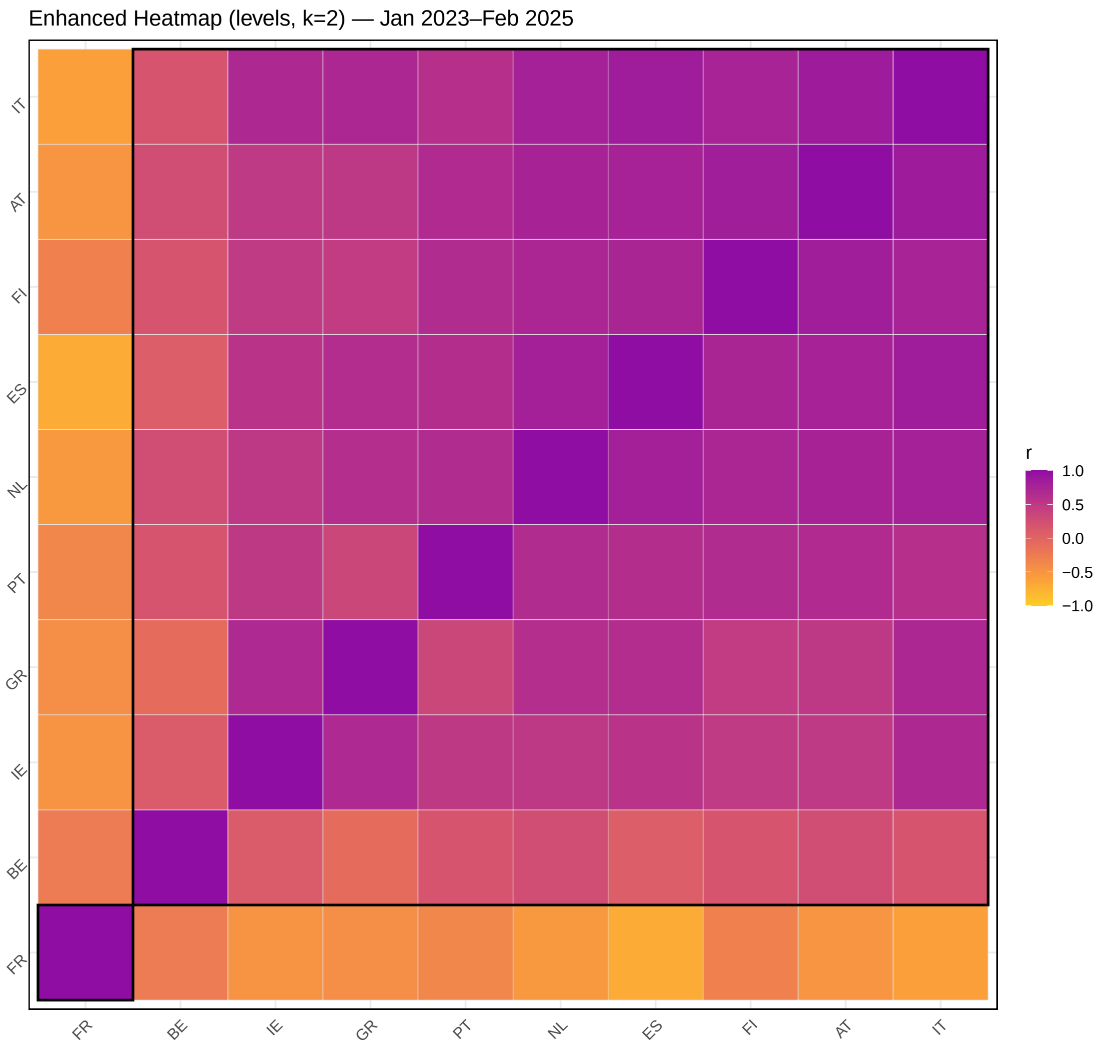
This panel reports the correlation matrix of one-month-ahead predicted 10-year sovereign yield spreads over the full sample, with countries ordered by hierarchical clustering into two groups.

Figure 12: Correlation heatmap of predicted spreads (2 clusters, levels, Jan 2022–Feb 2025)



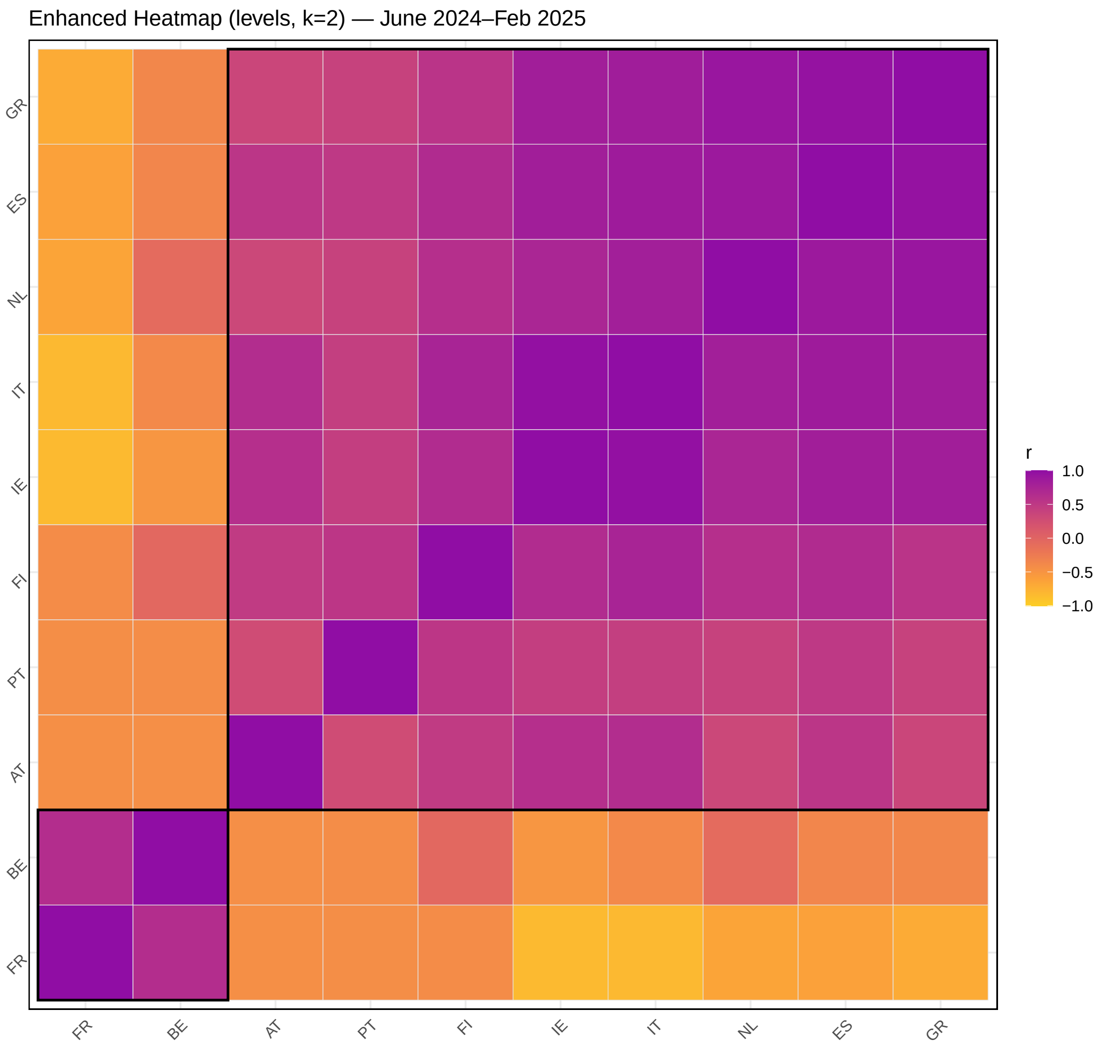
This panel focuses on the post-COVID and energy-crisis period, showing that the core–periphery structure in predicted spreads remains mostly intact.

Figure 13: Correlation heatmap of predicted spreads (2 clusters, levels, Jan 2023–Feb 2025)



Over this more recent window, the clustering indicates an incipient decoupling of France from the main euro area spread dynamics.

Figure 14: Correlation heatmap of predicted spreads (2 clusters, levels, Jun 2024–Feb 2025)



In the most recent period, France and Belgium form a separate cluster, signalling a pronounced divergence in their predicted spread behavior relative to the rest of the euro area.

6. Conclusion

This paper asks whether modern machine-learning methods can generate accurate short-horizon forecasts of euro area sovereign bond spreads and how such forecasts can be used to track financial fragmentation in real time. Using a high-dimensional macro–financial data set for ten euro area sovereigns over 2008-2025, we compare a wide range of linear, nonlinear and ensemble models to forecast one-month-ahead 10-year spreads relative to the German Bund, and we use SHAP-based diagnostics and correlation heatmaps of predicted spreads to recover an economically interpretable cross-country structure of sovereign risk.

The paper makes three contributions that mirror the structure of the analysis. First, we conduct a comprehensive horse race among thirteen forecasting models that span linear, nonlinear and rule-based approaches. Forecast accuracy is evaluated strictly out of sample using standard loss functions and tests of equal predictive accuracy. We find that a carefully tuned gradient-boosting specification (XGBoost) consistently lies on the performance frontier and is never significantly dominated by competing models.

Second, we adopt XGBoost as our baseline model and show that it delivers economically small one-month-ahead forecast errors across both core and peripheral sovereigns, including more volatile issuers. These results indicate that a small class of flexible ensemble methods can provide systematically accurate short-horizon forecasts of euro area sovereign spreads, and we use these forecasts as the basis for our fragmentation analysis.

Third, we open up the black box of the machine-learning model by combining its forecasts with post-estimation diagnostics. SHAP-based decompositions reveal that, at the one-month horizon, lagged domestic spreads are the dominant predictors in every country, pointing to strong short-run persistence in risk premia. Beyond this autoregressive core, the model loads on macro–financial indicators related to financial conditions, prices, global markets and household or monetary sectors, with the relative importance of these drivers differing across the currency union. Correlation heatmaps of predicted spreads then trace how the cross-country structure of sovereign risk evolves through time. Over 2020-2025, the model-implied spreads suggest that the fragmentation dynamics between core and peripheral countries are still at work; more worryingly, in more recent windows French spreads gradually decouple from the main bloc, and by mid-2024-2025 France and Belgium form a distinct cluster.

These findings have important implications for monitoring financial fragmentation in the euro area. Fragmentation emerges as a shifting configuration rather than a fixed core–periphery split, with potential changes in country membership and co-movement over time. This is highly relevant for monetary policy, as fragmentation may significantly distort the transmission of ECB decisions to sovereign borrowing costs. Machine-learning forecasts and correlation heatmaps may provide a high-frequency, model-based signal of emerging segmentation that could complement existing policymakers' dashboards. The recent decoupling of France and Belgium, traditionally "core" countries, is noteworthy, as persistent divergence in these markets suggests that a single policy stance may translate into heterogeneous financing conditions. As a result, our proposed framework can serve as an additional analytical layer for assessing fragmentation risk, complementing policymakers' efforts to detect and address such risks before they escalate.

Although highly informative, our framework remains predictive and reduced-form by design: it is not constructed to identify structural determinants or causal transmission channels, and our focus on one-month-ahead forecasts of 10-year yields leaves open questions about longer horizons and the full term structure. These limitations naturally suggest avenues for future work. One is to integrate identified macro-financial shocks into the forecasting framework, linking predictive content more directly to specific monetary, fiscal or geopolitical instability. Another is to extend the approach to multi-step and term-structure forecasts, including projections beyond our sample, and to embed the resulting indicators into scenario and stress-testing exercises.

Overall, the evidence demonstrates that modern machine-learning techniques can deliver accurate, interpretable short-horizon forecasts of euro area sovereign spreads, which can serve as informative indicators of financial fragmentation. We view this as a step toward illustrating how high-dimensional predictive tools could be incorporated into the surveillance toolkit of policymakers and market participants.

References

Afonso, A., Arghyrou, M. G., and Kontonikas, A. (2015). The determinants of sovereign bond yield spreads in the EMU. ECB Working Paper 1781, European Central Bank.

Altavilla, C., Giannone, D., and Lenza, M. (2015). The financial and macroeconomic effects of the OMT announcements. *International Journal of Central Banking*, 11(2):29–57.

Arakelian, V., Dellaportas, P., Savona, R., and Vezzoli, M. (2018). Sovereign risk zones in europe during and after the debt crisis. Ssrn working paper, SSRN.

Attinasi, M. G., Checherita, C., and Nickel, C. (2010). What explains the surge in euro area sovereign spreads during the financial crisis of 2007–09? *Public Finance and Management*, 10(4):595–645.

Barbosa, L. and Costa, S. (2010). Determinants of sovereign bond yield spreads in the euro area in the context of the financial crisis. *Banco de Portugal Economic Bulletin*.

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

Beirne, J. and Fratzscher, M. (2013). The pricing of sovereign risk and contagion during the european sovereign debt crisis. *Journal of International Money and Finance*, 34:60–82.

Belly, G., Guégan, D., and Hassani, B. (2023). Forecasting sovereign risk in the euro area via machine learning. *Journal of Forecasting*, 42(3):649–673.

Beutel, J., List, S., and von Schweinitz, G. (2019). Evaluating leading indicators of systemic banking crises: A mixed-cross-section global approach. *Journal of International Money and Finance*, 94:1–27.

Bluwstein, K., Buckmann, M., Joseph, A., Kapadia, S., and Şimşek, Ö. (2023). Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. *Journal of International Economics*, 145:103773.

Candelon, B., Luisi, A., and Roccazzella, F. (2022). Fragmentation in the european monetary union: Is it really over? *Journal of International Money and Finance*, 122:102545.

Casabianca, E., Catalano, M., Forni, L., Passeri, S., and Giarda, E. (2019). An early warning system for banking crises: From regression-based analysis to machine learning techniques.

Coulombe, P. G., Leroux, M., Stevanović, D., and Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Economic Dynamics and Control*, 135:104200.

De Grauwe, P. (2012). The governance of a fragile eurozone. *Australian Economic Review*, 45(3):255–268.

De Santis, R. A. (2012). The euro area sovereign debt crisis: Safe haven, credit rating agencies and the spread of the fever from greece, ireland and portugal. ECB Working Paper 1419, European Central Bank.

Ehrmann, M. and Fratzscher, M. (2015). Euro area government bonds: Integration and fragmentation during the sovereign debt crisis. Staff Working Paper 2015-13, Bank of Canada.

Favero, C. A. (2013). Modelling and forecasting government bond spreads in the euro area: A GVAR model. *Journal of Econometrics*, 177(2):343–356.

Favero, C. A. and Missale, A. (2012). Sovereign spreads in the euro area: Which prospects for a eurobond? *Economic Policy*, 27(70):231–273.

Gerlach, S., Schulz, A., and Wolff, G. B. (2010). Banking and sovereign risk in the euro area. Discussion Paper Series 1: Economic Studies 2010/09, Deutsche Bundesbank.

Gilbert, N. (2019). Euro area sovereign risk spillovers before and after the ECB's OMT announcement. DNB Working Paper 636, De Nederlandsche Bank.

Giordano, R., Pericoli, M., and Tommasino, P. (2013). Pure or wake-up-call contagion? another look at the EMU sovereign debt crisis. CONSOB Working Paper 71, Commissione Nazionale per le Società e la Borsa (CONSOB).

Grandi, P. and Taboga, M. (2019). The effects of the ECB's asset purchase programme on the sovereign bond market. *Journal of Empirical Finance*, 52:190–210.

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

Klose, J. (2013). Sovereign yield spreads during the euro-crisis: Fundamental factors versus redenomination risk. IWH Discussion Paper 8/2013, Halle Institute for Economic Research (IWH).

Liu, L., Chen, C., and Wang, B. (2022). Predicting financial crises with machine learning methods. *Journal of Forecasting*, 41(5):871–910.

Maltritz, D. (2012). Determinants of sovereign yield spreads in the eurozone: A bayesian approach. *Journal of International Money and Finance*, 31(3):657–672.

Moessner, R. (2018). Effects of asset purchases and financial stability measures on sovereign bond yields. BIS Working Paper 721, Bank for International Settlements.

Reimann, C. (2024). Predicting financial crises: An evaluation of machine learning algorithms and model explainability for early warning systems. *Review of Evolutionary Political Economy*, 5.

Stanga, I. (2011). Sovereign and bank credit risk during the global financial crisis. DNB Working Paper 314, De Nederlandsche Bank.

Tölö, E. (2020). Predicting systemic financial crises with recurrent neural networks. *Journal of Financial Stability*, 49:100746.

Zoli, E. (2009). Euro area sovereign risk during the crisis. IMF Working Paper 09/222, International Monetary Fund.

Credit authorship contribution statement

Bouillot, Roland:
Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing (original draft), Writing (review and editing), Visualization,

Candelon, Bertrand:
Conceptualization, Methodology, Writing (review and editing), Supervision, Funding acquisition.

Kool, Clemens:
Conceptualization, Methodology, Writing (review and editing), Supervision.

Data availability
The authors do not have permission to share data.

7. Appendix

Supplementary explanations, figures and tables

Table 14: Overview of appendix tables and figures

Item	Title	Content / purpose
Table 15	Diebold–Mariano tests of equal predictive accuracy vs XGBoost	Reports, for each competing model and country, Diebold–Mariano t-statistics and p-values for one-step-ahead forecasts under squared-error loss, testing equality of predictive accuracy relative to XGBoost. The table covers the full set of linear, non-linear and rule-based specifications used in the horse race.
Figures 15–34	DM-test panels vs XGBoost by country	For each of the ten euro area countries (Austria, Belgium, Finland, France, Greece, Ireland, Italy, Netherlands, Portugal and Spain), side-by-side panels show (i) the significance classification of Diebold–Mariano tests versus XGBoost across all competing models and (ii) the corresponding DM t-statistics. These figures provide a visual summary of where XGBoost significantly outperforms or matches alternative forecasting methods.
Figures 35–54	SHAP summary panels by country	For each country, side-by-side panels display (i) the top predictors ranked by their aggregate SHAP contributions and (ii) the frequency with which each predictor enters the monthly top-10 by absolute SHAP value over 2020–2025. These figures summarize which macro–financial variables systematically drive the XGBoost forecasts and how their importance differs between core and peripheral sovereigns.

Table 15: Diebold–Mariano tests of equal predictive accuracy vs XGBoost

Model	*Austria*	*Belgium*	*Finland*	*France*	*Greece*	*Ireland*	*Italy*	*Netherlands*	*Portugal*	*Spain*
Linear										
	1.70	-0.95	2.76	1.84	0.56	3.96	0.34	1.85	2.92	-1.05
Standard LASSO	(0.097)	(0.348)	(0.009)	(0.075)	(0.581)	(0.000)	(0.733)	(0.072)	(0.006)	(0.303)
	1.50	3.10	2.15	2.71	-0.49	1.31	2.74	3.85	0.39	-0.70
Group LASSO	(0.143)	(0.004)	(0.038)	(0.010)	(0.629)	(0.198)	(0.010)	(0.000)	(0.700)	(0.490)
	1.77	-0.85	2.82	2.28	0.81	3.62	0.71	2.06	2.18	-1.36
Sparse-group LASSO	(0.085)	(0.402)	(0.008)	(0.029)	(0.425)	(0.001)	(0.482)	(0.047)	(0.035)	(0.182)
	2.14	1.51	3.17	1.90	1.94	4.01	2.06	3.80	2.17	2.16
Ridge	(0.039)	(0.139)	(0.003)	(0.066)	(0.060)	(0.000)	(0.047)	(0.001)	(0.036)	(0.037)
	1.74	-0.94	2.78	2.17	0.57	4.01	0.78	1.15	2.59	-1.29
Elastic Net	(0.090)	(0.351)	(0.009)	(0.037)	(0.570)	(0.000)	(0.438)	(0.256)	(0.014)	(0.206)
Non-linear										
	1.70	1.23	2.13	1.66	1.99	3.10	2.15	2.20	1.56	2.03
Support Vector Regression (SVR)	(0.097)	(0.227)	(0.040)	(0.105)	(0.053)	(0.004)	(0.038)	(0.034)	(0.128)	(0.049)
	2.60	1.78	4.14	2.15	2.44	2.22	2.90	5.24	3.60	3.18
Feedforward Neural Network (FNN)	(0.013)	(0.084)	(0.000)	(0.039)	(0.019)	(0.033)	(0.006)	(0.000)	(0.001)	(0.003)
Rule-based										
Trees										
	1.14	1.65	3.43	1.52	2.79	3.01	1.70	2.16	2.73	2.21
Decision Trees	(0.261)	(0.107)	(0.002)	(0.137)	(0.008)	(0.005)	(0.097)	(0.037)	(0.010)	(0.033)
	1.77	1.32	1.73	1.19	1.52	0.27	0.88	1.66	0.93	-0.31
Random Forest (Bagged Trees)	(0.084)	(0.195)	(0.091)	(0.243)	(0.138)	(0.789)	(0.382)	(0.105)	(0.357)	(0.758)
	1.11	1.35	0.66	0.13	0.29	1.44	0.51	-0.22	-0.40	-0.62
Boosted Trees (GBM)	(0.273)	(0.186)	(0.516)	(0.898)	(0.774)	(0.160)	(0.616)	(0.827)	(0.694)	(0.536)
Stochastic Gradient Machines (SGM)										
	0.09	1.25	1.37	0.12	-0.82	1.43	1.06	0.53	-0.19	-0.74
SG Boosting	(0.926)	(0.220)	(0.181)	(0.903)	(0.417)	(0.162)	(0.299)	(0.597)	(0.853)	(0.462)
	2.61	1.29	1.59	0.83	2.14	1.30	-0.39	0.04	-1.00	-0.53
LightGBM	(0.013)	(0.204)	(0.120)	(0.411)	(0.039)	(0.200)	(0.699)	(0.966)	(0.323)	(0.599)

Notes: Entries report Diebold–Mariano t-statistics and p-values for one-step-ahead forecasts, comparing each model to XGBoost using squared-error loss. Negative (positive) statistics indicate lower (higher) average squared error than XGBoost. p-values are for a two-sided null of equal predictive accuracy.

Figure 15: DM-test significance vs XGBoost Austria

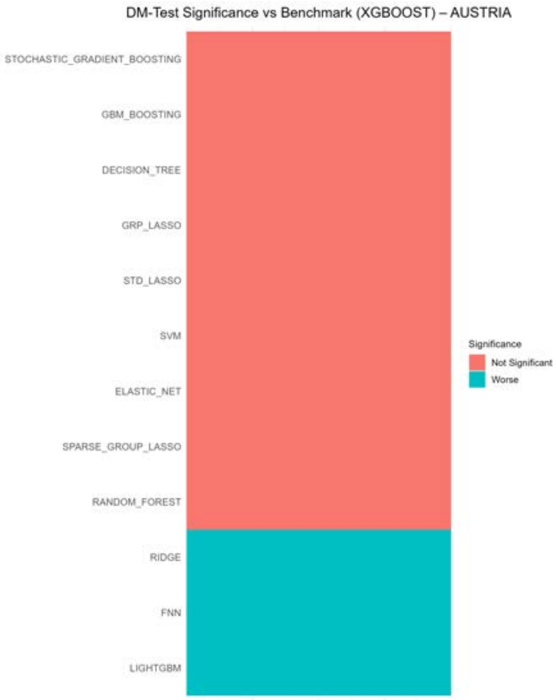


Figure 16: DM-test t-statistics vs XGBoost Austria

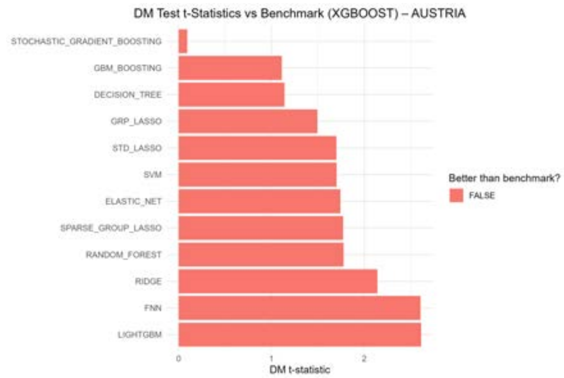


Figure 17: DM-test significance vs XGBoost Belgium

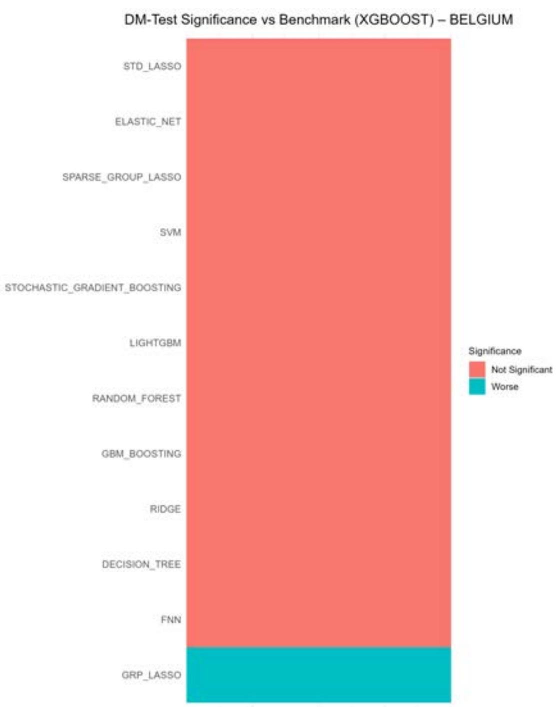


Figure 18: DM-test t-statistics vs XGBoost Belgium

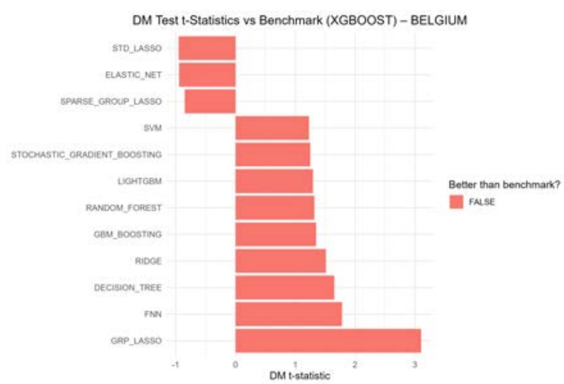


Figure 19: DM-test significance vs XGBoost Finland

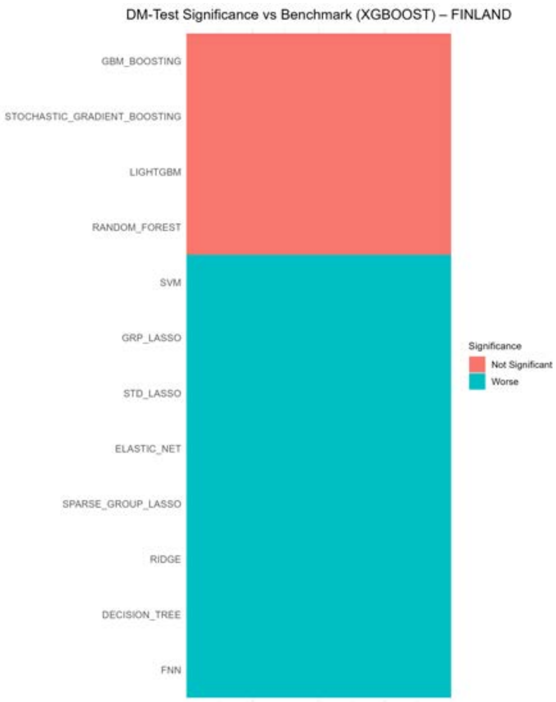


Figure 20: DM-test t-statistics vs XGBoost Finland

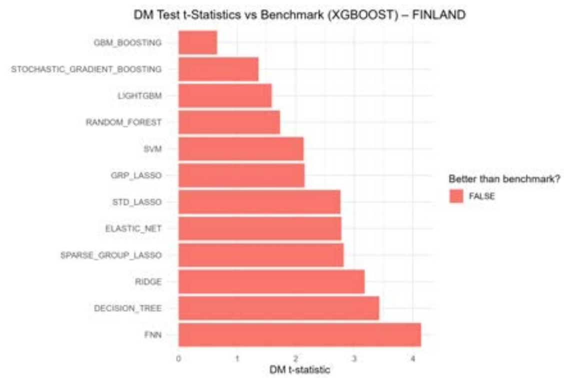


Figure 21: DM-test significance vs XGBoost France

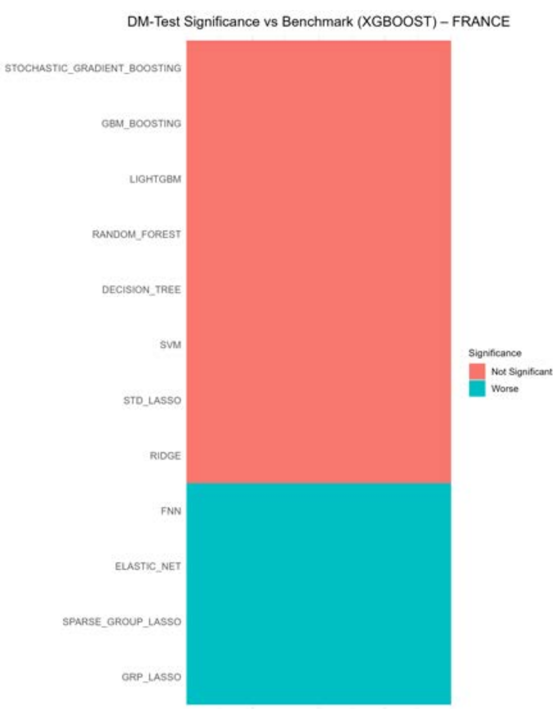


Figure 22: DM-test t-statistics vs XGBoost France

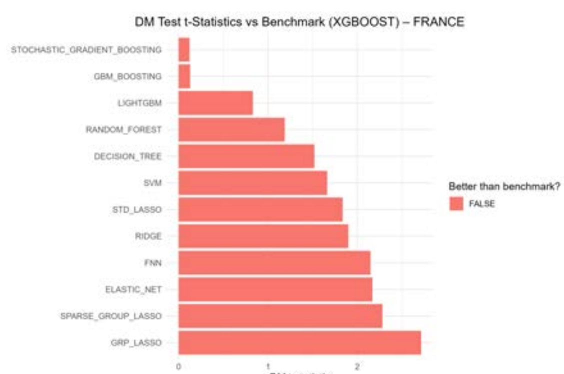


Figure 23: DM-test significance vs XGBoost Greece

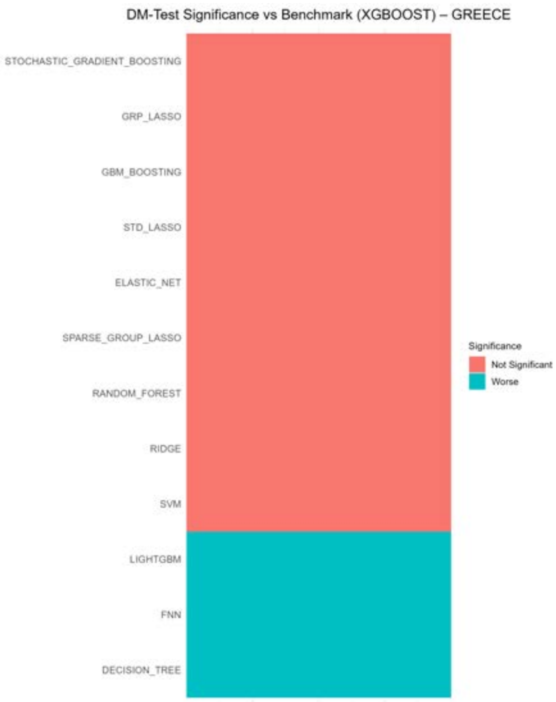


Figure 24: DM-test t-statistics vs XGBoost Greece

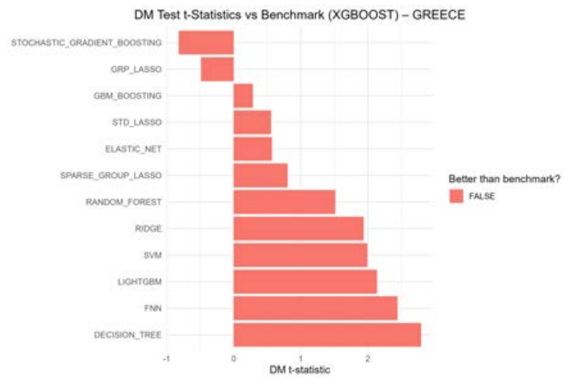


Figure 25: DM-test significance vs XGBoost Ireland

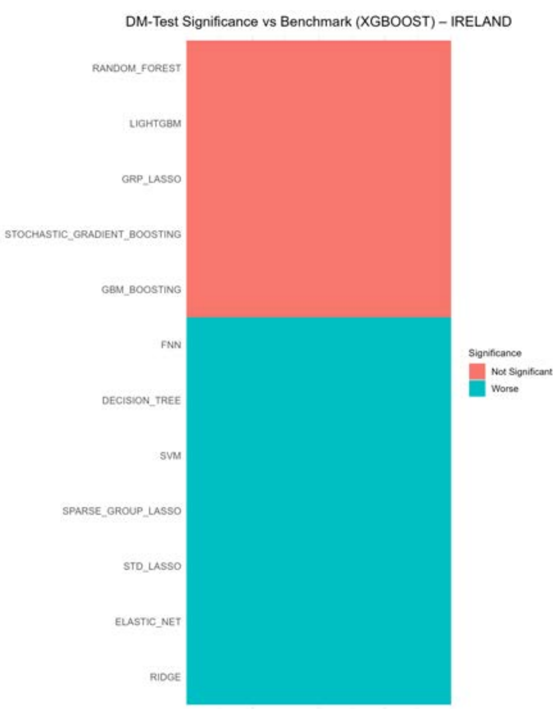


Figure 26: DM-test t-statistics vs XGBoost Ireland

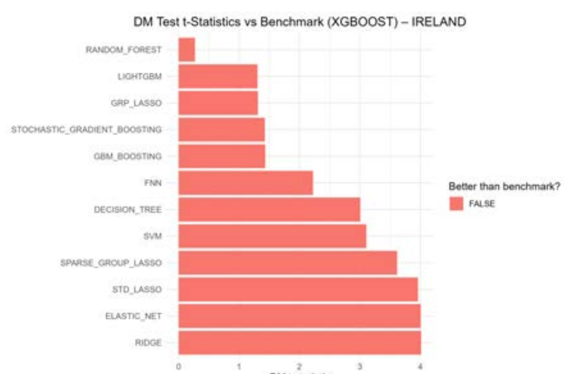


Figure 27: DM-test significance vs XGBoost
Italy

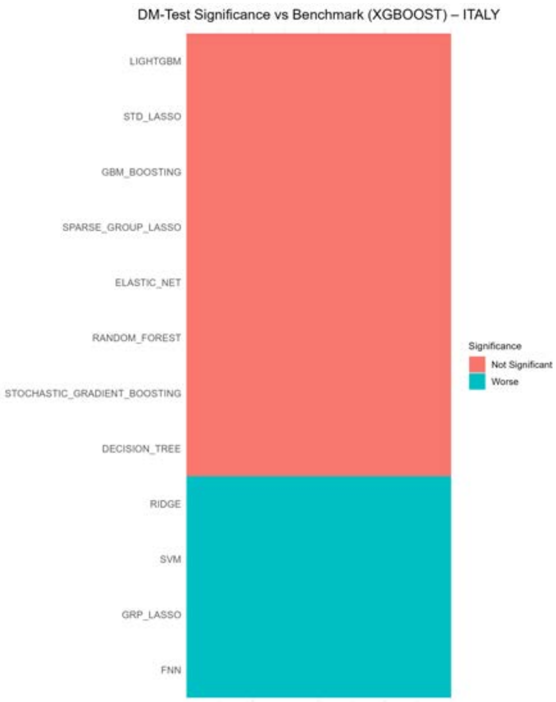


Figure 28: DM-test t-statistics vs XGBoost
Italy

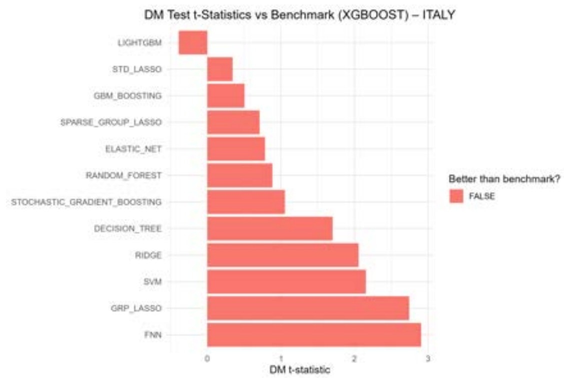


Figure 29: DM-test significance vs XGBoost
Netherlands

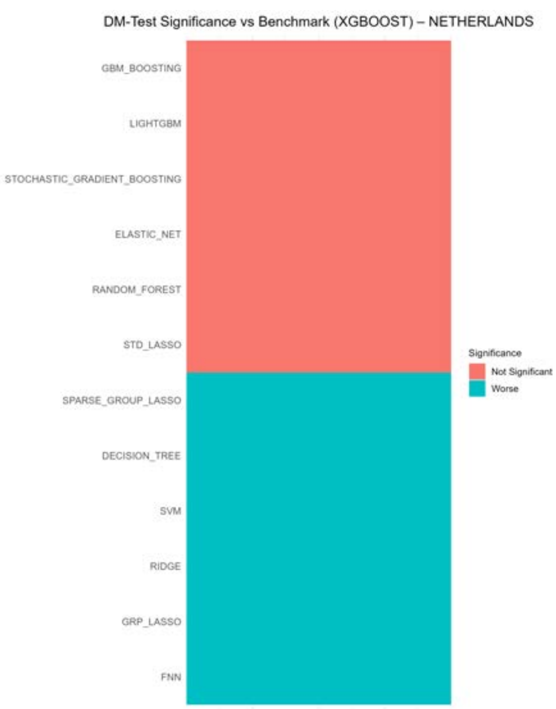


Figure 30: DM-test t-statistics vs XGBoost
Netherlands

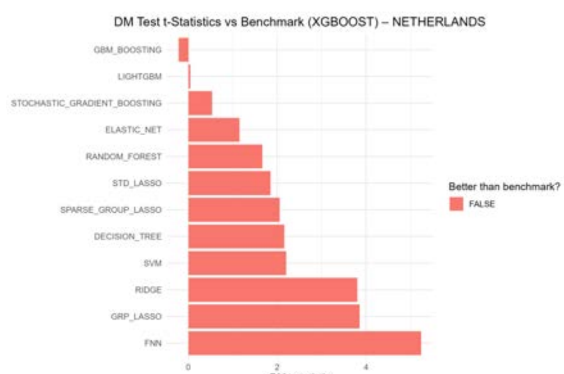


Figure 31: DM-test significance vs XGBoost Portugal

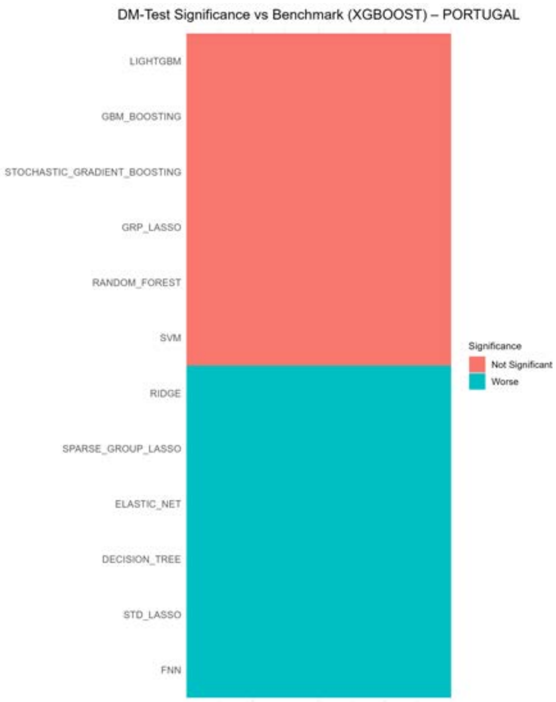


Figure 32: DM-test t-statistics vs XGBoost Portugal

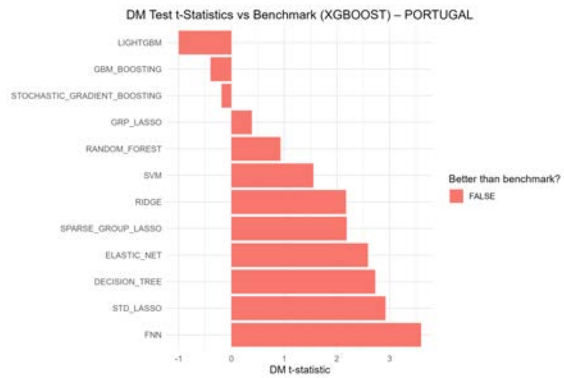


Figure 33: DM-test significance vs XGBoost Spain

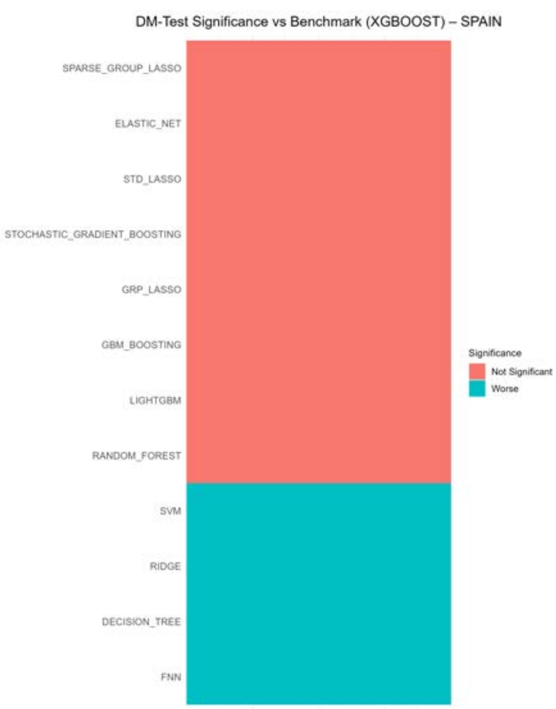


Figure 34: DM-test t-statistics vs XGBoost Spain

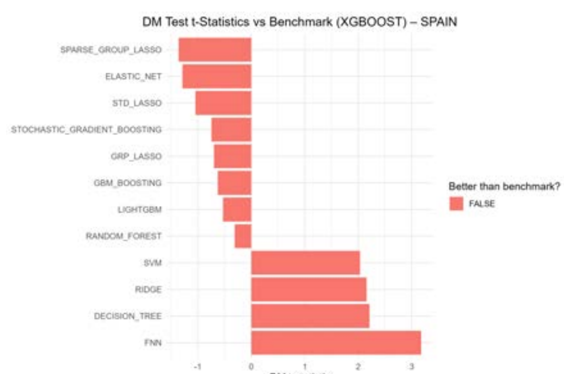


Figure 35: Top SHAP contributions – Austria

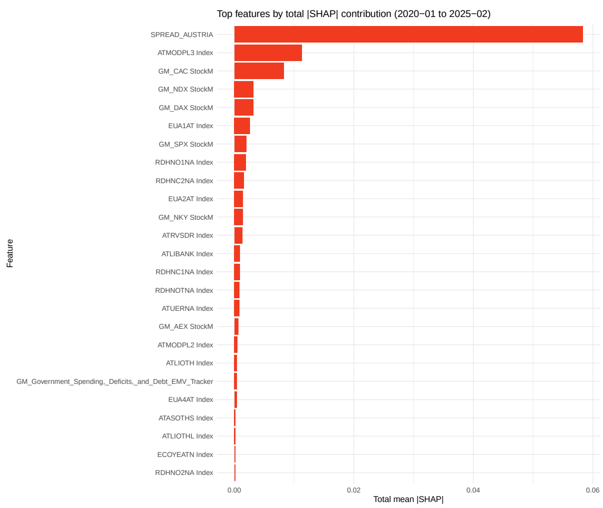


Figure 36: SHAP appearance counts – Austria

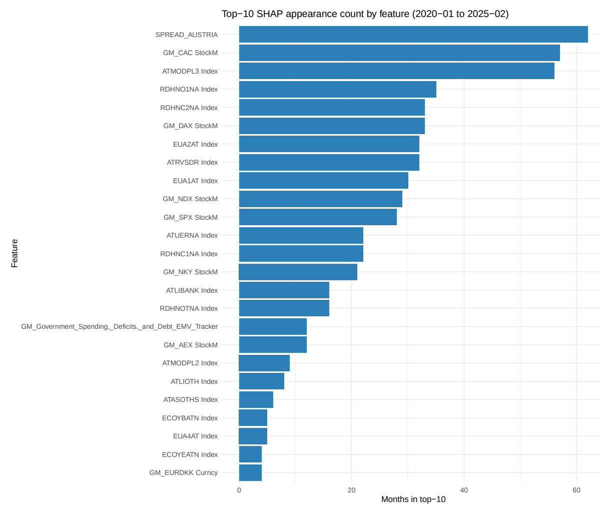


Figure 37: Top SHAP contributions – Belgium

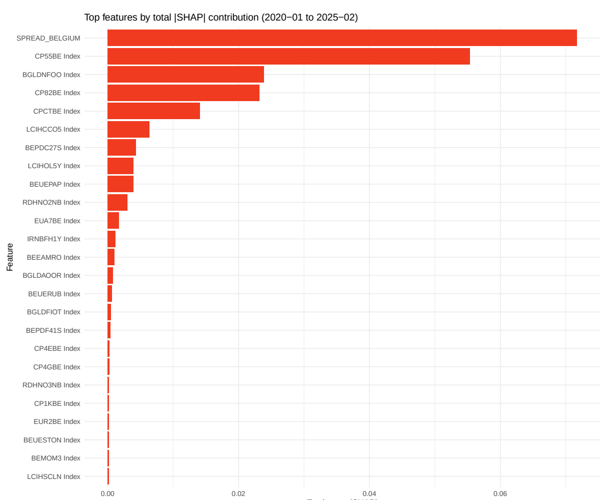


Figure 38: SHAP appearance counts – Belgium

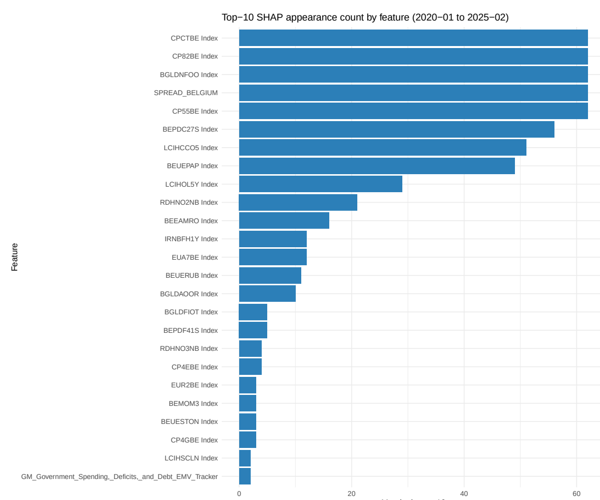


Figure 39: Top SHAP contributions – Finland

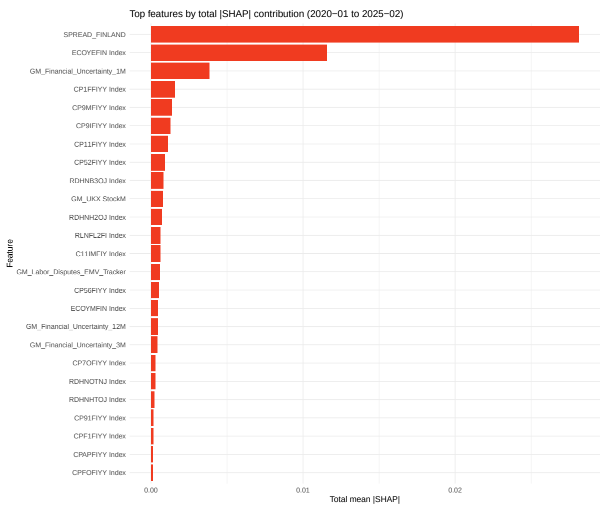


Figure 40: SHAP appearance counts – Finland

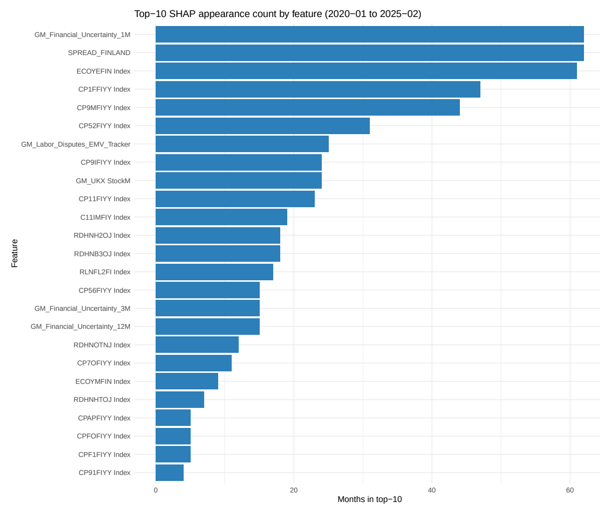


Figure 41: Top SHAP contributions – France

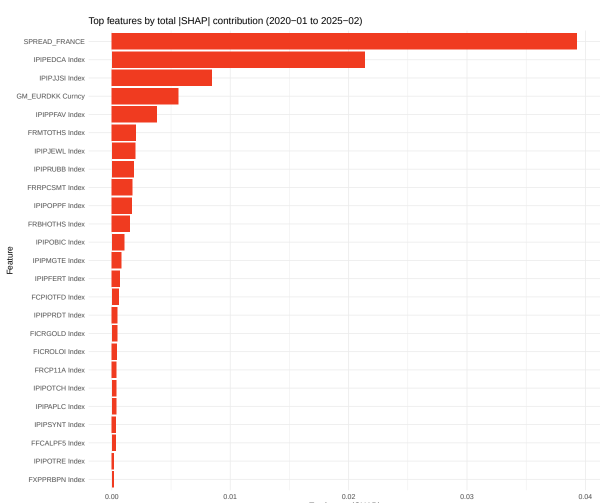


Figure 42: SHAP appearance counts – France

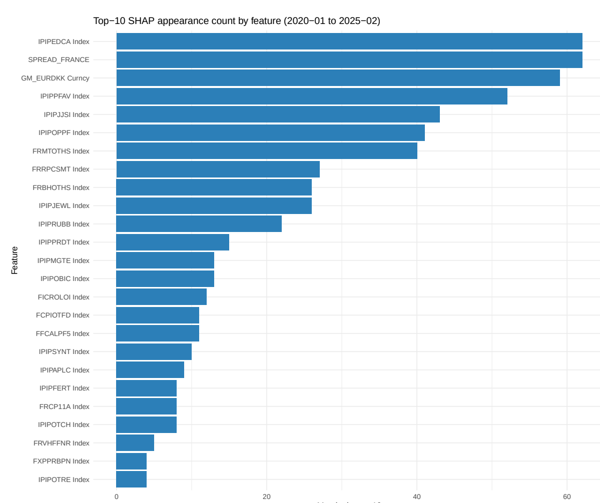


Figure 43: Top SHAP contributions – Greece

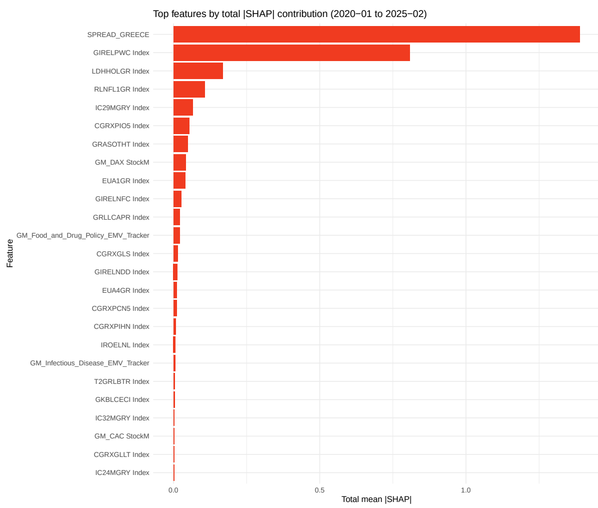


Figure 44: SHAP appearance counts – Greece

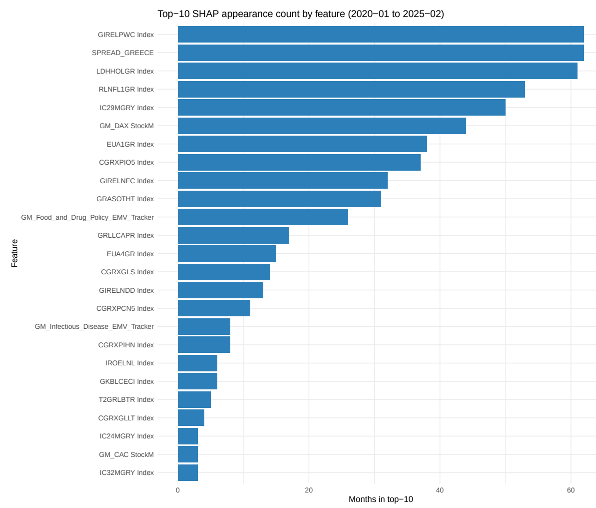


Figure 45: Top SHAP contributions – Ireland

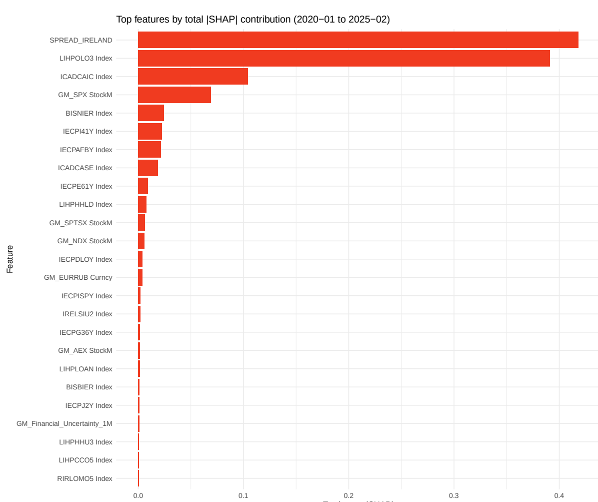


Figure 46: SHAP appearance counts – Ireland

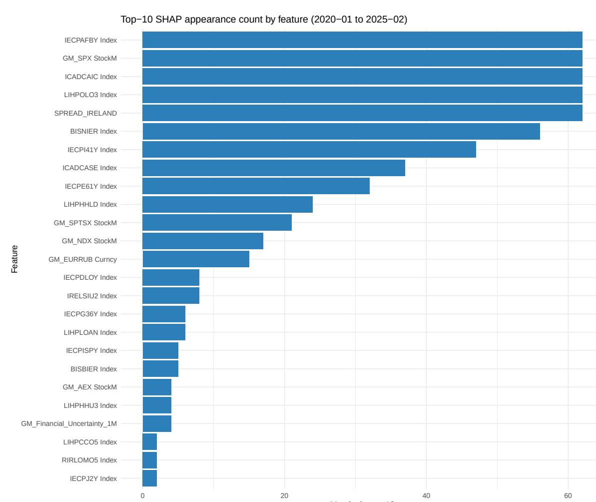


Figure 47: Top SHAP contributions – Italy

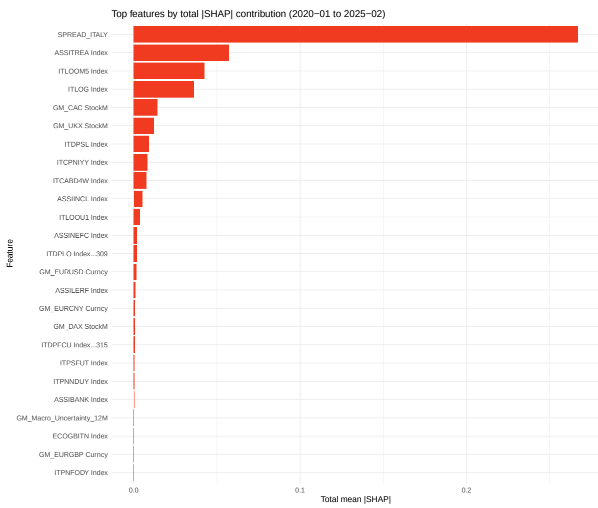


Figure 48: SHAP appearance counts – Italy

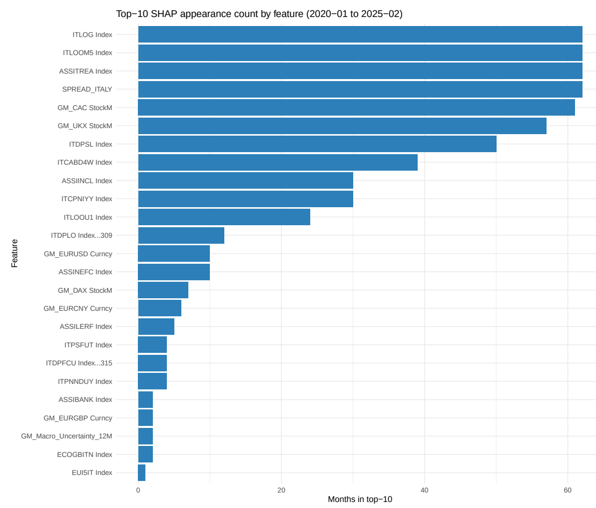


Figure 49: Top SHAP contributions – Netherlands

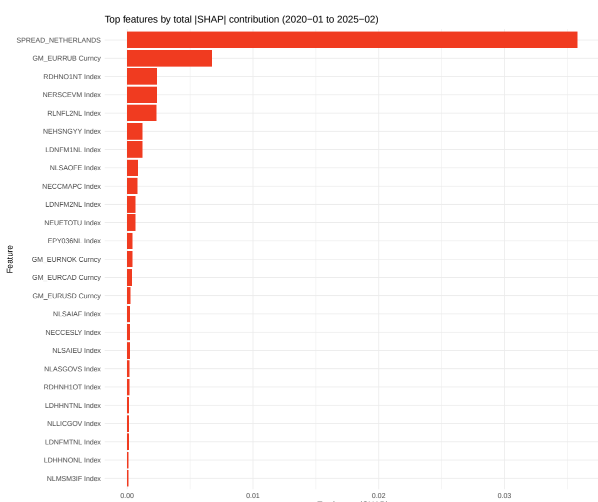


Figure 50: SHAP appearance counts – Netherlands

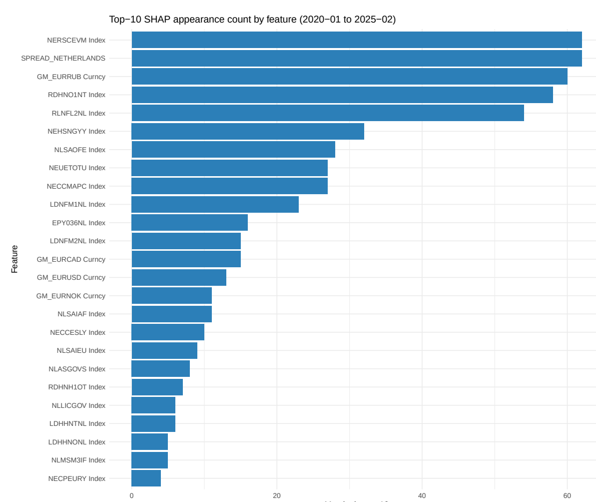


Figure 51: Top SHAP contributions – Portugal

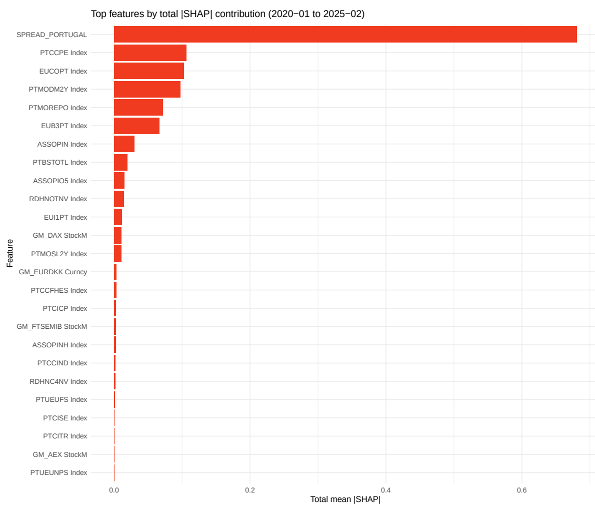


Figure 52: SHAP appearance counts – Portugal

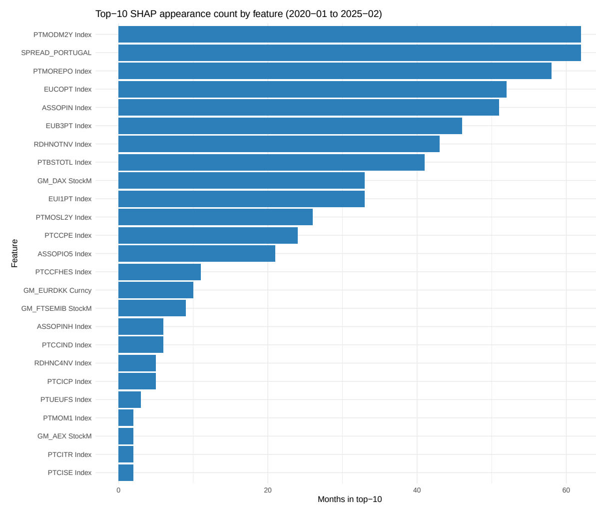


Figure 53: Top SHAP contributions – Spain

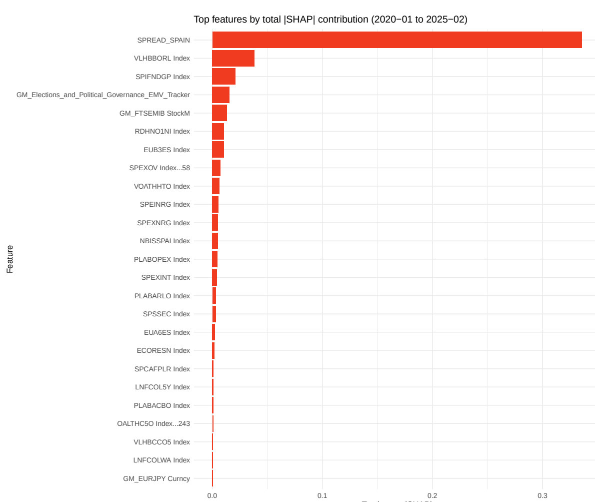


Figure 54: SHAP appearance counts – Spain